DATA SCIENCE AND BUSINESS INTELLIGENCE 179

DATA WAREHOUSE FOR GLOBAL AIR TRANSPORT DEVELOPMENT 180
Nenadović Ana

AUTOMATIC CREATION OF CLINICAL PATHWAYS – A CASE STUDY 188
Kirchner Kathrin, Marković Petar, Delias Pavlos

ASSOCIATION RULE MINING: MODERN BUSINESS APPROACH 195
Istrat Višnja, Matović Dajana, Palibrk Milko

PREDICTING RISK OF BUYING LOW QUALITY CAR WITH RAPID MINER 201
Mitic Nemanja, Jankovic Ana, Vukicevic Milan

PREDICTIVE MODELLING FOR CLAIMS PROCESSING – CASE STUDY OF BNP PARIBAS 207
Sreckovic Stefan, Suknovic Milija, Vukicevic Milan

LOGICAL CLASSIFICATION METHOD FOR BANKRUPTCY PREDICTION 213
Marković Dragana, Živančević Sanja, Poledica Ana

BUILDING A DATA-ENRICHED DEX-BASED DECISION SUPPORT SYSTEM FOR EARLY WARNING ON INCREASED RISK OF SKIING INJURY OCCURRENCE – FIRST RESULTS 221
Marković Petar, Delibašić Boris, Suknović Milija

EUROPE 2020 STRATEGY – A MULTIVARIATE APPROACH 228
Milenković Nemanja, Đoković Aleksandar, Vukmirović Aleksandra

A FUZZY LOGIC-BASED SYSTEM FOR ENHANCING SCRUM METHODOLOGY 234
Stupar Mihailo, Milošević Pavle, Petrović Bratislav

APPLYING THE ANALYTIC HIERARCHY PROCESS TO RANK CITY-BRANCHES 241
Cvetkoska Violeta, Begicevic Redzep Nina

DATA SCIENCE AND BUSINESS INTELLIGENCE

DATA WAREHOUSE FOR GLOBAL AIR TRANSPORT DEVELOPMENT

Ana Nenadović
Faculty of Organizational Sciences, University of Belgrade
e-mail: ananenadovica@gmail.com

Abstract: *The paper will seek to familiarize the fundamental conceptual framework for understanding modern systems for decision support such is data warehousing, and present this concept through its application in the airline industry, certainly one of the most challenging business areas. The additional aim is to emphasize once again the usefulness of data warehouse as the basis for business intelligence solutions in competitive global environment. This presentation includes all necessary steps taken in the design and implementation of data warehouse. By relying research on the one of the major public data sources, the idea is to generate an intuitive system designed for wider audience to serve as a business intelligence tool – for making decisions regarding the choice of the optimal time, airport or airline for travelling. In the final part of the paper, the results of the data warehouse usage will be delivered in the form of presentations created in ''Tableau Desktop'' data visualization software.*

Keywords: *data warehouse, business intelligence, decision support, airline, on-time performance*

1. INTRODUCTION

Today is often referred to as the age of the information crisis which is related to the over-saturation of data, followed by the simultaneous lack of information in spheres of business and science *(Jennex, 2013).* This phenomenon reflects a situation in which organizations have a vast amount of data in their possession while at the same time the management have difficulty accessing the information they need for decision making. As for the individual companies, there are two main factors that cause the occurrence of this information gap. Due to the piecemeal systems implementation, data are often heterogeneous and inconsistent (*Hoffer, 2007).* In addition, these systems are designed primarily with the aim to satisfy the operational objectives. Data warehousing systems have announced a revolution in the way business organizations implement business analytics and strategic decision making in almost all industries. Data warehouse is the database designed for data analysis and as such still an indispensable and significant platform for developing business intelligence solutions *(Hawking, Paul, Sellitto, Carmine, 2010).*
As the phenomenon of large amounts of data and their immediate uselessness particularly affects industries that are global by its very nature, the problem of modern airline industry is therefore particularly striking. In times when the quality of customer services is in the focus of competitive business model, the impression is the general dissatisfaction of passengers who have an increasing freedom to choose from and immediately punish for failures by deciding to fly with another company *(Vasigh, Fleming & Tacker, 2013)*. While traditional companies continue with confusing customers with various limitations, delays and with generally poor service, passengers are increasingly choosing new model of low cost airlines which offer transparent policy and affordable rates.
US Bureau of Transportation Statistics (BTS) collects data about domestic and international air transportation transactions. Thus, BTS is in possession of vast quantities of data that are publicly available for analysis and are significant source of knowledge about the global airline market. However, its datasets are difficult to manipulate by informal users which are assumed to be airline managers, the press, or ordinary passengers. The aim of this paper is to build an intuitive analytical system in order to ease working with data, so that users can gain the useful information faster and more easily. This would remarkably help pattern identification and process of creating influential reports *(Larsen, 2013).* As a result of that, gaining insights into traffic flows, flights statistics, most common delay reasons and the best time for trip would be instantly available. The limitations of reporting from BTS public data sets as well as the key advantages achieved by the implementation of data warehousing system will also be highlighted.

- Data are stored in huge datasets in CSV file format, and there is no easy and systematic way to obtain and compile the needed data. Accessing the data requires intensive and time consuming manual processing.
- There is no ability to perform custom analysis and drill down capabilities.
- There is no central place to view the data required for analysis and reporting.
- There is no automated way to get reports and quick predictions.

On the other hand, the data in the data warehouse are organized in a form that provides users with access to needed information through a minimal amount of effort and time. For this purpose is built a dimensional model that is largely intuitive and whose key benefit is data transparency. In addition, it provides a tool that guarantees the quality, accuracy and consistency of delivered data. These benefits generate the opportunity of better decision-making on the basis of theinformation not previously avaliable to the ordinary passenger.

Data warehouse for air transport data analysis ought to organize data retrievied from the web site Statistical Computing on page ASA Data Expo in a way which would provide an overview and insight into the on time arrival performance of passenger flights. The emphasis is thus on the key factors that influence the image of airline quality services such as reasons for delayed, diverted or cancelled flights.

2. DESIGNING THE DATA WAREHOUSE

The process of data warehouse development includes taking data from various sources and integrating them during which all inconsistencies are eliminated and data is transformed and archived in a way that best suits the needs of the end user.

2.1. Selecting and filtering data for analysis

On Time Performance Database contains on-time arrival data for non-stop domestic flights by major air carriers, and provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, cancelled or diverted flights, taxi-out and taxi-in times, air time, and non-stop distance.
Data are collected by BTS and consist of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008. There are nearly 120 million records in total, which take up 12 gigabytes when uncompressed. For obvious technical inability to process the amount of data, two annual files have been provided for the purposes of this work , more specifically, the subset of data covering the 2007 and 2008 calendar year.
When it comes to additional files supplied on the website, such as codebooks for airports, airlines and aircrafts, many irregularities have been observed: The table that provides data about the aircrafts and classifies them according to the model, age and the manufacturer has too many missing values, so it is estimated that this table is not useful at the moment, but is interesting provided that additional efforts are made for gathering data on aircraft registration codes and their origins. Afterwards, it was discovered that Airline codebook which ought to identify airline companies according to their IATA codes is not valid, i.e., contains information about fictive airlines, and is even not compliant with IATA coding system of two-letter codes. Similar is the case with the Airport codebook table, noting that the so-called IATA codes are actually mislabelled ICAO codes.
It was hence decided that new codebooks will be created according to the original data available on the IATA's website. Two tables have been generated in Microsoft Excel, for both the airline and the airport. The information these tables provide are details on two categories, custom facts about the flights dating from the period between 1998 and 2008. Therefore, the newly created table Airline based on the specific airline IATA code retained the information about its full name and USDOT (United States Department of Transportation) number based on its specific IATA code. Based on the particular airport IATA code, the Airport table returns the full name of the airport, latitude and longitude, the city and the country of airport.

As the data gathered and loaded into data warehouse come from different sources, the most demanding part of the entire project of data warehouse development is a process of data integration where the substantially different data from codebooks need to be associated with the transactional data in a single system. The whole set of procedures that collect and prepare data for loading in the final repository is referred to as ETL (Eng. Extraction, Transformation and Loading) processes. Prior to ETL processing the preparation of data is performed, which involves standardizing data in terms of unifying forms where the records of different backgrounds provide the necessary uniformity. On the other hand, data cleansing should be performed to remove all records that occur as a result of past mistakes. Possible cases of errors are incorrect, incomplete or inconsistent data.

2.2. Extracting data into staging database

After the process of data selection and preparation, in the first stage of the ETL process data are extracted into the staging database from which the data warehouse is being built. The packages populating the tables in the staging database are prepared in SQL Server Integration Services (SSIS) 2012. Database with transactional flights data is available in the separate annual CSV files. Each of the files contains approximately seven million rows. Using SSIS Platform for ETL processing, the available operator Union All combines multiple inputs into a single output data. Thus are the large annual files forwarded for further

processing. One of the problems that can occur when importing data from CSV to SQL Server is the issue of having to convert data types from Unicode data types to types of SQL Server. Through SSIS it is possible to enter the different types of data into the server. However, SSIS treats strings in CSV file as Unicode, while tables in SQL Server, are defined as non-Unicode. Microsoft services for data integration provide operators for conversion that are used in this step. The following operators have been applied: Data Conversion operator, whose role is to carry out the transformation of data types of corresponding columns and Derived Column that generates a new column or replaces the old one, by performing the transformation based on the given formula for the input column or more of them.
The package with which the annual files for 2007 and 2008 were extracted and loaded into the staging database Flights is shown in the following figure:

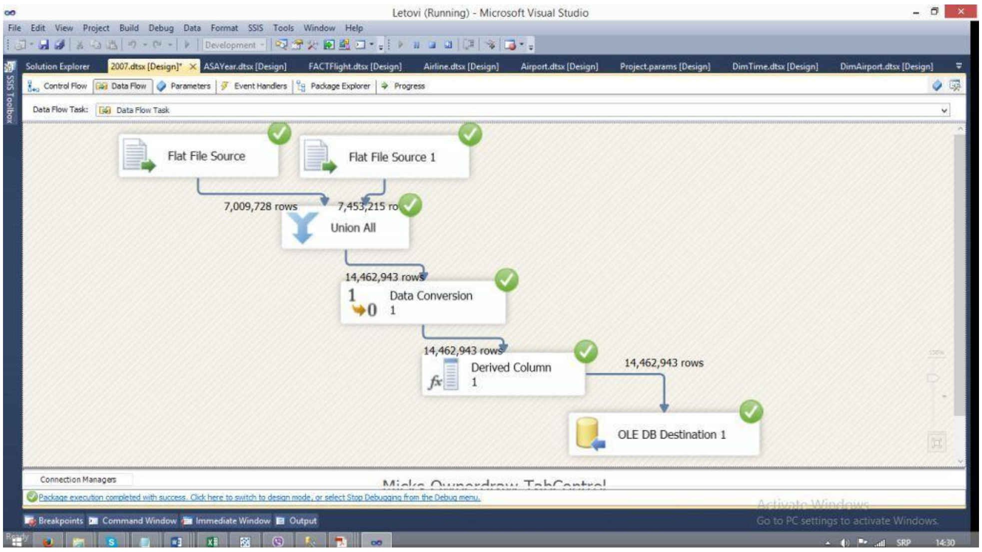

Figure 1: Extracting data into staging database

2.3 Dimensional modelling

At dimensional modelling this stage of a data warehouse it is necessary to determine how to logically group extracted data. The first step after selecting the grain of data warehouse is to choose the fact table, then the dimensions and schema of future data warehouse. When data is transferred from a source into the dimensional model, this means that we have one central fact table and a group of dimensional tables that are directly related to the fact table *(Kimball, Ross, 2013).* Facts are generally associated with numerical properties called measures. Dimensions are actually features that the client wants to analyze. The additional type of data stored in the fact table are codes, i.e. key attributes of dimensional tables by which these supporting tables are associated with the central table. The simplest and most often implemented schema is the Star schema *(Kimball, Ross, Thornthwaite, Mundy, Becker, 2010).* The Star schema is the core of Kimball methodology and because of its simplicity is closest to achieving a paradigm of business intelligence based on which its systems should be as intuitive as possible with the purpose of providing the end-user with the independent usage *(Adamson, 2010).*
The information we have in staging database Flights are records of individual flights in 29 columns concerning their timely performance or schedule deviations. More detailed information on the origin and destination of the flight as well as the airline that performed the flight are available in Airport and Airline tables.
The original dataset is extended by generating the table to which the role of time dimension will be assigned. Time dimension is one of data warehouse key components, as it provides its historicity. Firstly, columns Year, Month, and DayofMonth from the Flights table are unified into column that makes a specific date. Based on the date, Time dimension ID is generated to correspond with the same column in the fact table, with the role of key.
Selection of fact table in this case is unambiguous. Database Flights is the only one that records the transactions, and by its structure and composition is perfectly suited for the role of the fact table. In addition to a numerous measures and time determined facts it contains aggregated attribute and ID codes for airlines, airports and aircrafts. One of the most relevant measures is column ArrDelay that records the delay in minutes. Attributes that cover its special cases i.e. different reasons for the flight delay are also interesting. Database furthermore contains distance between the two airports, time spent in the air and other attributes that measure delay on the separate phases of flight. FactFlight contains 1448980 records and covers two calendar years starting from January 2007 and ending December 2008.

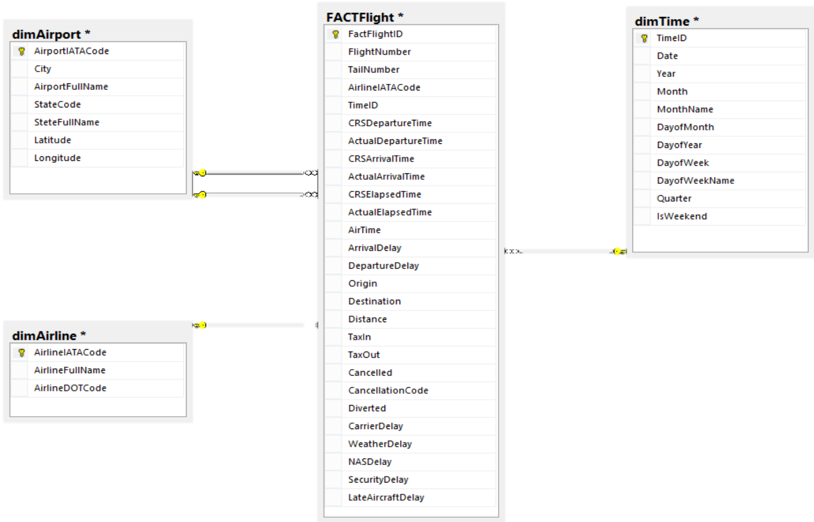

Figure 2: Data warehouse model diagram

Based on the information available in supporting tables, dimensional tables dim.Time, dim.Airport and dim.Airline are created. As IATA code uniquely identify a specific carrier and airport, there was no need for the implementation of the surrogate key on these dimensions, while the dim.Time ID is implemented in the form of a primary key that is similar to the timestamp, just as in the format YYYYMMDD as the most common practice (if there was necessity for greater level of granularity, it would contain hours, minutes, etc.).

Practically each of the dimensions contains a data hierarchy. The hierarchies are highly desirable because they allow users to analyze data aggregations in a very straightforward manner using the OLAP or data visualization software functions. Dim.Time appears to be self-explanatory and its hierarchy goes from year to the day of the month level of detail. Due to a particular nature of transactions, the columns quartile and weekend are also included because it is expected that the percentage of flights over the weekend as well as in particular seasons of the year will increase, rather than in others. Dim.Airport hierarchy simply allows individual airport data to be rolled up from IATA code, the full name of airport and the cities of origin/designation, to the names of the country.

2.4. Data warehouse implementation

The implementation of data warehouses marks the phase of physical design of the data warehouse and involves the whole process of data migration from operational or staging database into a central repository, including all necessary operations to be executed on the data, thereby customizing them for the future structure and purpose *(Sarka, Lah, Jerkič,* 2012). When the logical model is made, creating the database and its objects commences. Microsoft's SSIS tools for data integration are essential ETL tool and a platform for loading the traditional data warehouses.

Most of the data is loaded directly into the database tables, as is the case with the dim.Time which has already been prepared as being self-generated for the purposes of a specific data warehouse. The remaining dimensions, dim.Airline and dim.Airport are populated with necessary data in a similar manner. The column names are adapted semantically for the intuitive usage without a detailed introduction to metadata. Nevertheless, some transformations were necessary.

Similar problem that occurs when importing data from a CSV file in SQL server happens with transfering data from Excel tables. In SSIS some necessary conversions were made in order to store data in SQL Server tables. An additional problem occurs when transferring data from Microsoft Excel files. Because of the way that strings are stored in Excel cells, it is necessary to coverse data in WSTR form into SSIS. While it is possible to change the length of the string SSIS, Excel still keeps it in the internal form as if it were larger. As the SSIS string has the length of 50, one was to make a compromise and leave in the SSIS the column populated with the length of 50, all of which had no consequence on the query level. Same as before, to change the data type in SSIS package data flow pipeline, Data Conversion operator’s functions has been used.

FactFlight is implemented according to the presented database model diagram in such a manner that the FlightNumber column was taken as the primary key , an attribute that uniquely identifies a particular flight. As mentioned earlier, the dimTime ID is made in the form of a date, and it was necessary to adjust the current time attributes in the fact table to correspond the column in the time dimension as a foreign key. Columns for the year, month and day of the month have been integrated in the appropriate column that corresponds to the date in the dim.Time. For this purpose, the SSIS toolbox operator Derived Column Transformation creates new column or replaces the old one according to given formula.

Among the ArrDelay attribute records the extensive presence of NULL and 'NA' values was noticed where attribute values are less than or equal to zero. These are the cells that should not contain the values because it is the record of the flight for which the delay has not been recorded. The values in columns that indicate the reasons for the delay (WeatherDelay, SecurityDelay, LateAircraftDelay, CarrierDelay, NASDelay) for the flight codes which recorded t no delays, have been recognized in some places as a NULL, and in others are populated with 'NA'. It was necessary to make changes to the data, in order to make the string 'NA' recognized as a NULL in SSIS. By using the Derived Column Transformation operator, these particular values ought to be replaced. Within the Derived Column Transformation editor window it is possible to select the needed functions from the NULL functions folder, and in this case, the function that converts the DT_STR string values into NULL is chosen.

3. DATA DELIVERY

The work below illustrates data warehouse usage for the business intelligence solution development with the purpose of delivering useful information in the area of air transportation. For the purpose of this study and the particular data warehouse, the data visualization software Tableau is chosen.
1. What time of the year and days of the week are best to travel in order to minimize the likelihood of flight delay? – Presentation of flight delays by days of the week, months and seasons.

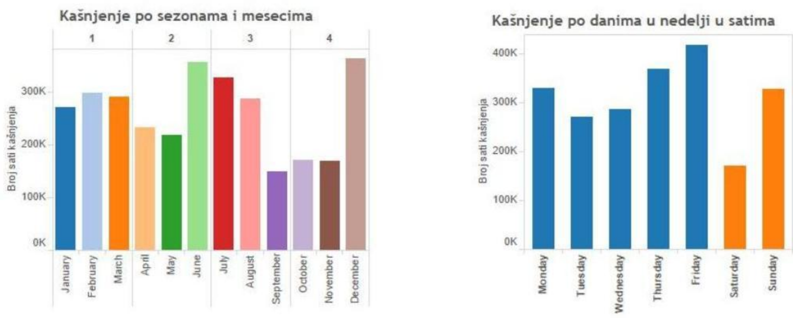

Figure 3: Flight delays per months, seasons and days of the week

Figure 3 represents the arrival delay measured in hours querying by time dimension. The first graph represents the total delays in hours per quarters and months. Filtered by months, each is marked with a different colour. Using the data hierarchy in Tableau is achieved by simply dragging the columns and is visible at the level of the data source.

In the second graph, the total delay per day of the week is presented. Bar colours indicate if it is weekend or working day. The vertical axis shows arrival delay measure values. As ArrivalDelay column records arrival delay in minutes, there are lots of negative values and delayed flights do not include the flights whose delay is less than 15 minutes, and the attribute is filtered to include only those values that are greater than or equal to 15. Similar data filtering could be done at the source level as Tableau allows connecting to the custom SQL query results.

2. What are the most common reasons for flight delay? - Presentation per months and in percentage.

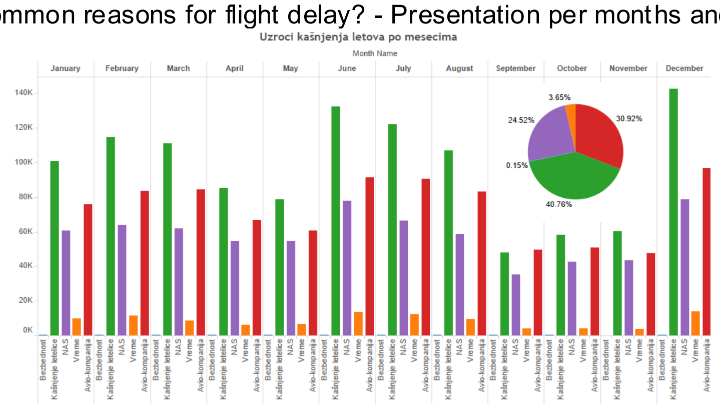

Figure 4: Presentation of specific delay reasons

As in the previous examples, when one wants to compare multiple categorical values, the best solution for graphical representation is a bar chart diagram. In the following diagram, the time axis is partitioned by months where for each month the most influential factors on flights delay are comparatively observed. Different colours mark individual delay reasons. The measure expressed by a total number of delay in hours records the values in the range from 189 to 142.433.

Pie chart graph shows the delay reasons in percentage. Comparable distinctive measures of the same category analysis in the Tableau is solved by filtering the field Measure Value. When connection to the data source is made, Tableau automatically generates fields Measure Values and Measure Names which are the parameter of all available measures, and inside of which the selection of any required measure is performed.

3. How does the weather affect the flight delays? – Presentation showing the correlation of total arrival delay and delays caused by weather conditions.

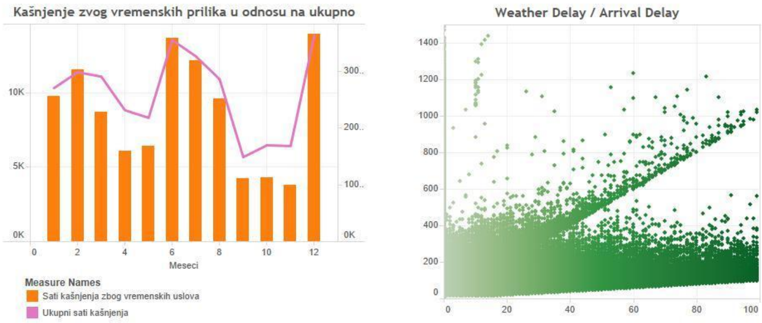

Figure 5: Weather delay and total arrival delay correlation

Although the previous analysis has shown that weather conditions have a small percentage in the total amount of flight delays, drawing the diagram in which it is possible to examine comparable trends of two dimensions, indicates strong correlation between these two measures. The impact of weather on total delay is shown by scatter plot graph where the weather is placed on the horizontal axis, whereas the vertical axis represents the total delay in hours. Pie chart graph shows the percentage of flight cancellation caused by weather in the total amount of cancelled flights. Database records are filtered keeping the cancelled flights, and this value is further broken into subcategories where each of the separate reasons is presented by different colour.

4. Which airports and airlines are more prone to delays than others? – Presentation of airports and airlines with the highest delay rate.

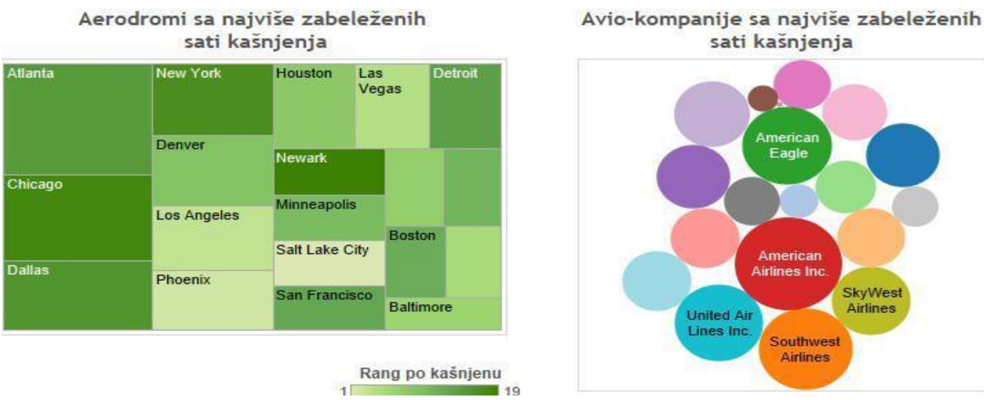

Figure 6: Airport and airlines

Airports are at first examined through the tabelar presentation where the calculated field that counts the amount of delay by the number of reported flight for an individual airport is generated. Thus are the airports that are most prone to flight delays discovered. These are then ranked and shown in a more intuitive way by drawing heat map chart where you can best see thescale and they were ranked according to arrival delay frequency. In the first graph in Figure 6. shows the airports using their full names, where the intensity of the color displays the values from 1 to 19, in a manner that the largest number is assigned to the airport with the highest delay level expressed in hours by number of records for the particular airport.

When it comes to airlines, the following graph in Figure 6. presents airlines with full names, marked with different color, and the size of the airline displays the total amount of delay for the airline expressed in hours. Thediagram does not show the airline with a negative value of the total delay. Further analysis has shown that it is Aloha Airlines.

5. What air lines include the highest traffic density? – Presentation of city pairs that represent most frequent routes.

Tableau allows for the possibility to analyse paths between two geographical points. To visualize these patterns, special data structure is required. Table Route is generated from the data warehouse as a SQL query result, and contains information about all routes realized between any two airports. As Tableau for this kind of analysis requires a specific data structure, Route table was not necessary to link with others in the data warehouse schema. For every data pair the unique key is made that identifies them as a pair. Each path takes two rows in the table, one for both of the directions, so a column was created in the table to define the path order according to latitude and longitude of the origin airport.

Table 1: Route Table

Column Name	Description
Origin	Origin airport IATA code
Destination	Destination airport IATA code
OriginAirportfullName	Origin airport full name
DestinationAirportfullName	Destination airport full name
Latitude	Origin airport latitude
Longitude	Origin airport longitude
PathOrder	Path order number (1 or 2)
RouteName	Route full name

By connecting Tableau to the Route table a map is generated showing a network of all itineraries shown in Figure 7. Different color is assigned to each route. The thickness of the lines represents the number of flights realized between the two airports of the particular route. Ability to filter itineraries is made possible by generating the parameter, according to the number of realized flights based on which they are ranked. It is possible to locate the intersection nodes of two airlines on the map.By filtering the map of the 30 most frequent routes, their names will appear.

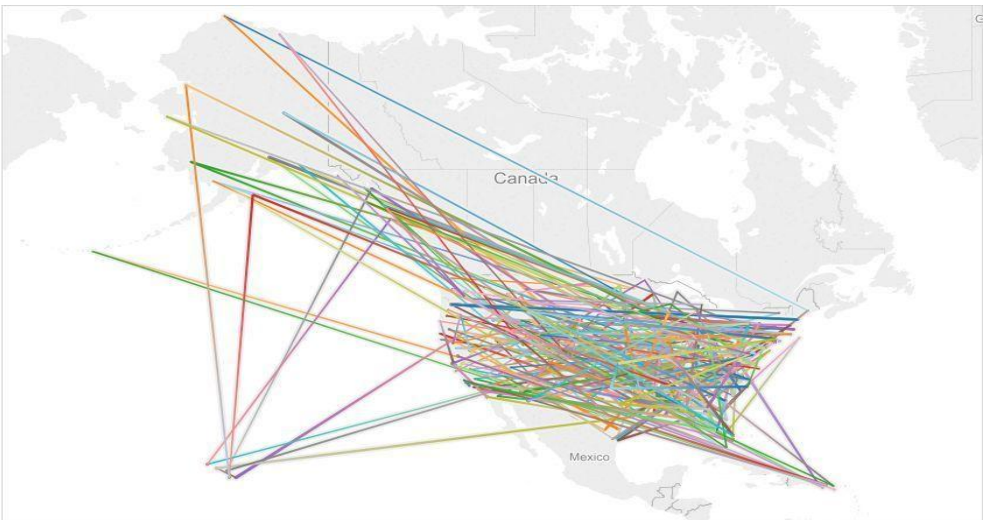

Figure 7: Route network

4. CONCLUSION

Business intelligence has evolved because the grow of data altering how organizations and individuals use information (Larson 2016).The subject of the paper was designing and implementation of the data warehouse for the air transportation data analysis. In the times when the informed customer is on the very first place of a succesful business model, the lack of quality information is visible when it comes to air transport services quality*(Davenport, 2014)*. Although there are some sources of open data in the area of flight performance, they are useless in a straightforward manner and require time consuming and manual processing effort in order to get and compile data. There was no automated and systematic way for the simplest analysis to be done. A data warehousing system is a platform for business intelligence solutions development, and enables what was not possible before – to organize data from various sources in a thematic and intuitive way, therefore providing the users with an integrated tool for performing the most demanding analytical operations on data in a fast and easy way.

The aim of the study was the use of publicly available data for creating business intelligence solution that would provide assistance to passengers in the choice of time, place and the airline for trip, in order to avoid long waits, jams and other air traffic inconveniences caused by delays. As the schedule deviation is one of the key reasons that influence the airline quality valuation by the passengers, it is necessary to emphasize the importance of the availability of data as the basis for making decisions of this kind, which would be grounded upon the reliable knowledge.

REFERENCES

Adamson C., (2010). *Star Schema: The Complete reference*, The McGraw-Hill.

Davenport T. H. (2014). *Big data @ work: dispelling the myths, uncovering the opportunities*. Harvard Business Review Press.

Hawking, Paul and Sellitto, Carmine, "Business Intelligence (BI) Critical Success Factors" (2010). ACIS 2010 Proceedings. Paper 4.

Hoffer J, Marry B. Prescott & Fred R. McFadden,(2007).“Modern Data Management”, Eight Edition, Pearson Prentice Hall, New Jersey.

Jennex E., (2013). *Knowledge discovery, transfer, and management in the information age*, IGI Global

Kimball R., Ross. M,(2013).*The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley.

Kimball R., Ross M.,Thornthwaite W., Mundy J., Becker B., (2010).*The Kimball Group Reader - Relentlessly Practical Tools for Data Warehousing and Business Intelligence*, Wiley.

Larsen T., (2013).Cross-Platform Aviation Analytics Using Big-Data Integration Methods, *Integrated Communications Navigationand Surveillance (ICNS) Conference.*

Larson D., and Chang V., (2016) A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science. International Journal of Information Management, 36, (5), 700-710.

Sarka D, Lah M, Jerkič K,(2012). *Training Kit (Exam 70-463):Implemmenting a Data Warehouse with Microsoft SQL Server 2012*, Microsoft Press Training Kit, 1st edition.

Vasigh B., Fleming, K.,& Tacker, T.(2013). Introduction to Air Transport Economics: From Theory to Applications. Ashgate Publishing.

AUTOMATIC CREATION OF CLINICAL PATHWAYS – A CASE STUDY

Kathrin Kirchner*[1], Petar Marković[2], Pavlos Delias[3]
[1]Berlin School of Economics and Law, Germany
[2]University of Belgrade, Faculty of Organizational Sciences, Serbia
[3]Eastern Macedonia and Thrace Institute of Technology, Kavala, Greece
*Corresponding author, e-mail: kathrin.kirchner@hwr-berlin.de

Abstract: *In hospital environments, treatment processes, respectively clinical pathways, are adopted based on the health state of a patient. Modeling of pathways is time consuming and due to the involvement of many participants, the introduction of clinical pathways is cost-intensive. A possibility for automatic or semi-automatic creation of clinical pathways is process mining. But such algorithms usually have problems to discover such processes from log data that are recorded during process execution in information systems. In this paper, we discuss the application of a spectral clustering algorithm to cluster process flows that can provide clearer process maps. We apply this method on an anonymized real world clinical dataset and discuss challenges and first results.*

Keywords*: clinical pathway, process mining, spectral clustering, flexible process, case study*

1. INTRODUCTION

A clinical pathway is defined as a structured, multidisciplinary care plan which defines the steps of patient care for a certain disease in a specific hospital (Rotter et al., 2010). They can improve the efficiency and transparency of patients" treatments: Length of stay in the hospital as well as cost can be reduced, patient safety is increased and new medical personnel can learn more quickly how a certain treatment process is executed.

A clinical pathway is usually developed manually by the medical personnel which is costly and time consuming. A once modeled pathway has to be updated regularly so that it is always following new regulations. An automatic support of clinical pathway creation would therefore be helpful (Rebuge & Ferreira, 2012).

A possibility for automatic or semi-automatic creating clinical pathways is process mining. Process mining is an emerging field that connects business process management and data mining (van der Aalst, 2012). It has three general purposes: as-is business process discovery from data, conformance checking of detected processes with pre-designed process models, and enhancement of the process model. In this paper we concentrate on the first topic. Process mining is used in industrial and administrative processes. Mans et al. (2012) applied process mining algorithms in a dentistry case, and found that it was difficult to handle flexibilities in the pathways. Poelmans et al. (2010) used process mining on breast cancer data and faced problems with data quality.

In process mining the structuredness of processes can vary between the so-called: “Lasagna processes” - simple in structure, consist of a relatively small number of activities and have a consistent flow, and the other side of the spectrum, which corresponds to “spaghetti processes” - with a very diverse and inconsistent process flow, and a large number of activities. Thus, the first are easier to analyze using standard process mining algorithms and techniques, whilst the second are more challenging.

One problem for process mining of clinical pathway data is the amount of flexibility in patient treatment which leads to spaghetti processes. An approach proposed by Delias et al. (2015) suggests applying clustering algorithms to group similar traces before applying a process mining algorithm, and this way reduce the high amount of different traces.

In this paper, we apply this approach by Delias et al (2015) on data collected from a clinical information system (Kirchner, 2015). We investigate whether spectral clustering can solve the problems with process flexibility in our case study. Challenges are discussed, and ideas of further improvement are developed.

Thus, our paper is structured as follows: In section 2, we describe the medical data that we analyze in section 3. Section 4 discusses our findings. In section 5, we summarize our results and give an outlook on future work.

2. DATA AND METHODOLOGY

During the research project PIGE (Kirchner et al., 2013), a clinical pathway for living liver donors was modeled in BPMN together with the medical personnel. This process was modeled by hand within a team of physicians and process modeling experts. Afterwards, data from a clinical information system was added to the process steps. This data contained timestamps and the name of the treatment procedure for a certain patient.

The process can be roughly described as follows: A healthy person can donate a part of her/his liver to a near relative. Before one becomes a living donor, she or he must undergo testing to ensure that the individual is physically fit. Sometimes computer tomography (CT) scans or magnetic resonance tomography (MRT) are done to image the liver. The pre-examinations are predetermined, but can change in the sequence depending on the availability of necessary resources. During and after operation, complications can occur that lead to additional interventions or even an additional operation. The data set was extracted from a clinical information system. All patient data which were marked as living liver donors in a time period of 3 years were selected. The resulting data set contained 50 living liver donors with 331 events. Not all patients went through all process steps. If the pre-examination found the person not suitable for donating the liver, an operation is not done. Therefore, the number of process steps for patients was different. Patients that were already in a later process step in the considered time period were also in the data set. Thus, not all pathways had the same start- and endpoint. Furthermore, the timestamp for all events were only dates, and several events can be done on a day.

Data was analyzed in a first step using Fuzzy miner algorithm in the version implemented in Disco software (www.fluxicon.com). Fuzzy miner uses correlation metrics to simplify the process model at a certain level of abstraction. To solve problems with flexible pathway execution, it can ignore less important activities or cluster them (Günther & van der Aalst, 2007). Figure 1 shows our first obtained process map that we obtained as a spaghetti process. The map comprises 45 process variants for the 50 patients. The shortest pathway contained only one event, the longest one 26 events. Filtering out seldom used pathways, as it is possible with Fuzzy miner, would delete also pathways with interesting characteristics (e.g., complications after operation).

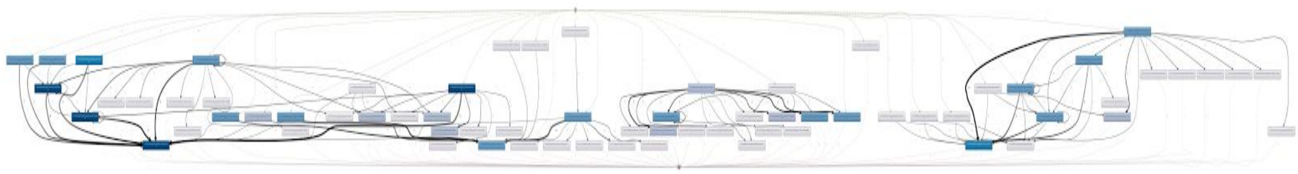

Figure 1: Original Process Map for living liver donors discovered by Disco-Fluxicon

For successfully analyzing such spaghetti processes, some form of data preprocessing is necessary. The most common approaches are filtering of activities and/or traces (cases) or clustering of activities and/or traces (cases). Some of the widely used clustering approaches can be found in (Veiga & Ferreira, 2010), (Song, Günther, & van der Aalst, 2009), (Jung, Bae, & Liu, 2009), (Bose & van der Aalst, 2009), (Luengo & Sepúlveda, 2012). Besides this, there is a group of robust algorithms such as (Delias et al, 2015), that efficiently resolve spaghetti processes in the domain of healthcare without any filtering, and which we chose to apply in this paper.

Our methodology applied in this paper is based on the methodology proposed by Delias et al. (2015). Thus, the following five steps were applied:

1. Creation of Event log from the hospital information system
2. Traces generation from the Event log
3. Calculation of traces" similarity using cosine similarity and robust similarity concept
4. Spectral clustering
5. Visualization of obtained clusters

3. DATA ANALYSIS

The first step in the implementation of our methodology is the creation of an event log, which is a starting point in any process mining analysis. Data were created from the hospital information system. For that purpose, a query was defined on the information system, selecting all patient treatments for patients marked as possible living liver donor. In order to follow the data security requirements, no personal information was extracted and patient identification numbers and dates were anonymized. The structure of the resulting event log used is shown in figure 2. Mandatory fields for further analysis are: Patient ID, Activity (Treatment Procedure) and Timestamp. The rows in the event log represent occurrence of a single event in the system (e.g., patient is sent to do the „CT of Abdomen" or the patient is sent to the „Operation room").

Patient_ID	OPS Code	Treatment Procedure	Day of treatment	Admission day	Discharge day
12345678	3-225	CT: Abdomen	10.10.2014	10.10.2014	12.10.2014
12345678	3-226	CT: Pelvis	10.10.2014	10.10.2014	12.10.2014
....					
23456789	3-225	CT: Abdomen	08.02.2015	08.02.2015	10.02.2015
....					

Figure 2: Event log structure

In order to reduce the complexity of the data, we analyzed which activities are often executed together on the same day. We calculated the co-occurrence matrix (Fig. 3). It consists of 32 activities, whilst the values of cells represent the average number of appearance of activities A and B in days where there were at least two activities per case. The matrix is symmetrical, as ordering of activities is not important (e.g. it is irrelevant if A follows B, or vice versa). The most frequent co-occurring activities in the heatmap are represented with red color, medium co-occurrences with black, whilst the least frequent combinations are colored with green. From the matrix, we derived that six activities, namely CTs and MRTs of different parts of the body (activity numbers 5, 6, 7, 19, 20 and 21), are often (at least in 20% of all cases) done on the same day. Therefore, we grouped them into a new activity named "CT/MRT diverse". Selection of merging candidates was done by reading the heatmap values, but the domain knowledge was the main criterion used for final decision. Therefore, e.g., activity number 9 (partial resection of liver) as part of operation of patient was not merged with pre-operational steps like activity number 5 (radiopaque CT abdomen). The steps that are described in the next section were applied on this modified data set.

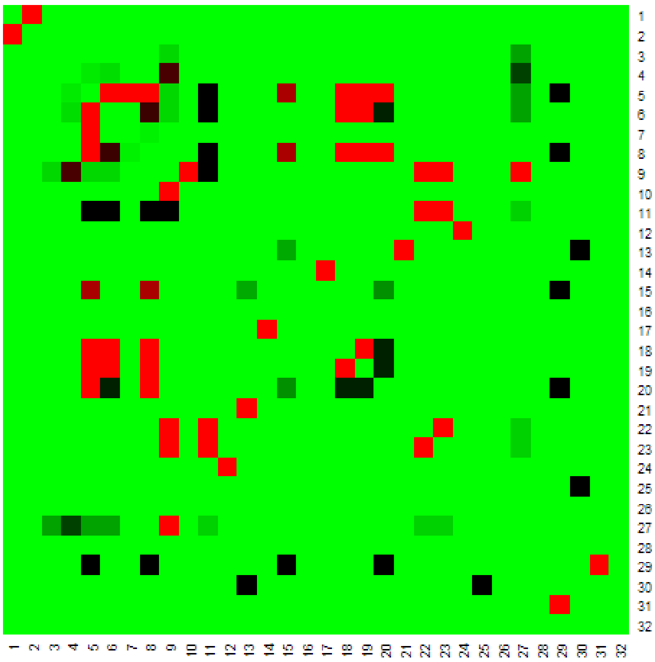

Figure 3: Co-occurrence matrix for the data set

When the structure of the event log is known, traces are generated from it. Traces represent process paths of each recorded case (patient). In other words, traces can be observed as chronologically ordered vectors of patients" activities. They are obtained from the event log, when activity fields, from all recorded events in the event log, are ordered chronologically and grouped by case ID. The example of a single trace is given in the figure 4.

Patient ID 12345678: <CT: Abdomen, CT: Pelvis>

Figure 4: A single trace (example)

After all traces are created, it is necessary to calculate their similarity in order to find the groups of patients" process paths that have high intra-group similarity, but are very different compared to the groups. In terms of process mining, two popular criteria for process similarity are: activity similarity and transition similarity both represented with corresponding vectors and defined with formulas 1 and 2, respectively.

$$SIMtransitions(T_i,T_j) = \frac{a(i)\cdot a(j)}{|a(i)||a(j)|}\frac{\sum_k t_k(i) \times t_k(j)}{\sqrt{\sum_k t_k(i)^2 \ \times \sum_k t_k(j)^2}} \quad (1)$$

$$SIMactivities(T_i,T_j) = \frac{a(i)\cdot a(j)}{|a(i)||a(j)|}\frac{\sum_k a_k(i) \times a_k(j)}{\sqrt{\ {}_k\, a_k(i)^2 \ \times \ {}_k\, a_k(j)^2}} \quad (2)$$

The first formula (1) presumes that two cases (patients) are similar if they undergo the same medical procedures, whilst the latter presumes that two cases are similar if they undergo those procedures (activities) in the same order. These two types of similarity are combined into overall similarity, represented with their convex linear combination - formula 3.

$$s(T_i,T_j) = S_{ij} = W_a \cdot SIMactivities(T_i,T_j) + W_t \cdot SIMtransitions(T_i,T_j) \quad (3)$$

Weights of the overall similarity components are determined based on expert judgement. Due to data quality issues we preferred to give more weight to transition similarity. All similarities are cosine similarities of corresponding vectors.
Lastly, the robust similarity is calculated using local densities concept (Chang & Yeung, 2008). The reason for using this approach is to enable low-frequent, outlier traces, to be spotted and prevent their influence on the clustering process. The idea behind the robust similarity concept is that an object, surrounded with more objects, should have higher chances to be grouped together with his neighbors. The measure used for estimating local density is given in formula 4 as $s'_{ij}\, l_i$ represents local density of the object.

$$l_i = \underset{j \in N_i}{\quad} S_{ij} \ where\ N_i\ is\ neighborhood\ of\ an\ object\ i \quad (4)$$

$$s'_{ij} = s_{ij}\, l_i l_j$$

This overall robust similarity is represented in the form of similarity matrix and represents the entry point for the spectral clustering step. Visualization of similarity matrix is given in the Figure 5 in the form of a heatmap, where the most similar cases/traces are colored with red, cases with medium similarity colored in black, and the least similar cases colored in green.

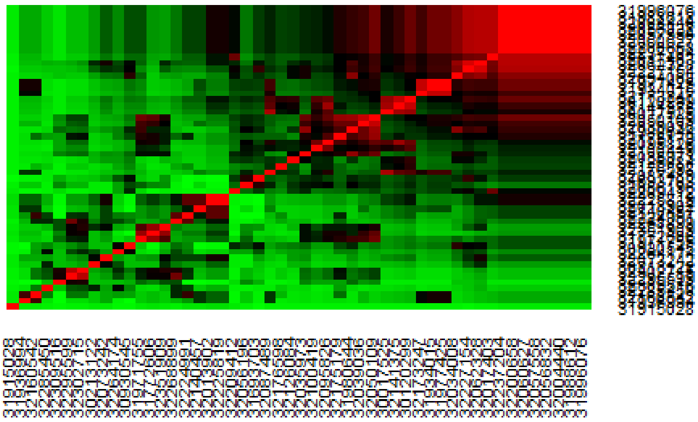

Figure 5: Heatmap of traces" (cases/patients) similarity

Spectral clustering (von Luxburg, 2007) is the following step, and consists of two substeps. First, the Laplacian matrix, which is derived from similarity matrix obtained in the previous step, is analyzed using its eigenvectors and their corresponding eigenvalues in search for the optimal lower-level subspace representation of the starting matrix. This is achieved through selection of first k eigenvectors of the Laplacian matrix which capture the highest variability in the data. They can be observed in the figure 6 as the eigenvectors after which exists a significant drop in the eigenvalues. Afterwards, the number of first n eigenvectors from the previous substep is used as a parameter k (number of clusters) in the K-means clustering algorithm. For our experiment, we selected k=2 as the number of clusters. Final results are discovered clusters consisting of similar traces. Each case‟s corresponding cluster number is added into the event log.

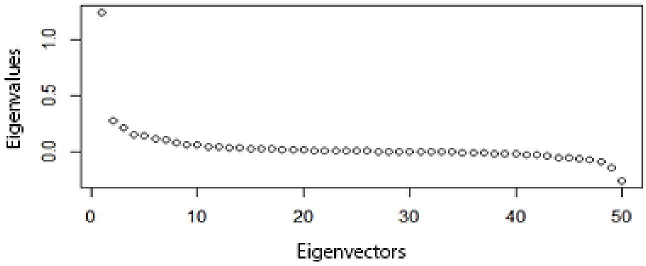

Figure 6: Eigenvectors and eigenvalues for the similarity matrix

The final step consists of visualizing obtained process maps using Fluxicon Disco process mining tool, and it is described in detail in the following chapter.

4. DISCUSSION OF RESULTS

After a careful interpretation of the heatmaps and the eigenvalues, we tried several numbers of clusters as well as different weights for the activities and transitions similarities. We then analyzed the resulting clusters and tried to identify the cluster solution that fits the best from the application point of view.

We achieved the best results with activity similarity of 0.1 and transition similarity of 0.9. Two clusters were identified, one of them (cluster 1) being illustrated in figure 5 contains 28% of cases and 30% of events. The longest path has 8 events, the shortest only three events (Figure 7). In principle, three types of clinical pathways can be derived here:

1. Patients that undergo the usual treatment: A pre-evaluation phase to check whether they can become a living liver donor and the operation phase.
2. Patients that undergo the usual treatment, but have complications after the operation
3. Patients who undergo just the pre-evaluation phase, but are not considered as living liver donors, so the treatment process finishes earlier.
4. Furthermore, there were also 2 patients that were already in the operation phase in the time period that was considered for the data collection. Therefore this pathway starts immediately with the operation.

The second resulting cluster was similar to the first one, but consisted of 72% of the cases. Here, cases having several process steps processed on one day are included. Because of data issues (date format) the process discovery algorithm cannot have a clear picture of what the sequence of events on a specific day was, therefore, the process execution seems to be very flexible. We could reduce this flow variability through a preprocessing step that merges the frequently co-occurring process steps (i.e., steps occurring frequently on the same day). More than 80% of cases are unique, and half of the cases have less than 5 events. Therefore, it is impractical to discover a process map that summarizes the behavior of the second cluster. Other techniques that could describe marginal aspects of the process (e.g., association rules) could add some value in process comprehension, yet they are out of the scope of this paper.

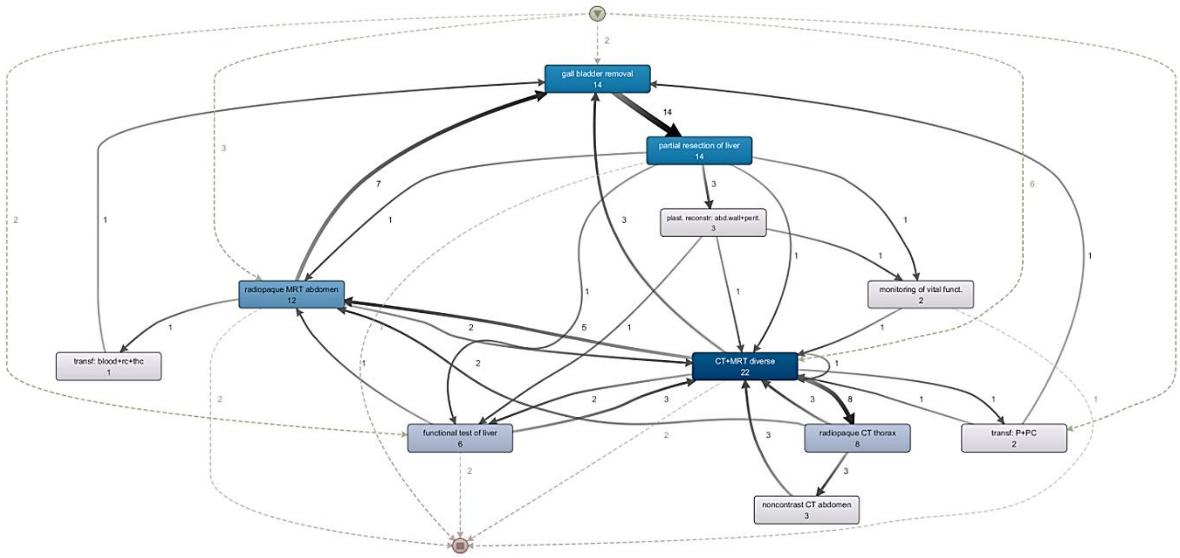

Figure 7: Resulting process map for cluster 1 of the data set (Visualization created with Disco-Fluxicon, 100% activities - 100% for transitions)

5. CONCLUSION AND OUTLOOK

In this paper we tried to discover patterns of process behavior for a particularly flexible environment: the living liver donors‟ treatment process. Since variability of flow is commonly expected in such a healthcare setting, we applied a trace clustering approach to summarize the flows.

The case study posed some additional challenges, because of the coarse timestamps, the small size of sample, and the large deviation of the pathway sizes. However, we were able to reach two significant results: First, there is a cluster of cases that exhibits some strong patterns, and we were able to plot them with a process map. Medical personnel could benefit from such a map by gaining a quick yet comprehensive understanding of the process. Second, trace clustering approaches are inappropriate to deal with the remaining population.

It is within our future plans to elaborate on the more idiosyncratic part of the population by applying different approaches (e.g., declarative/hybrid models, association rules, using patients‟ characteristics to correlate them with flows, etc.), to mine for any possible hidden underlying patterns.

REFERENCES

Bose, R. P., & van der Aalst, W. M. (2009). Context Aware Trace Clustering: Towards Improving Process Mining Results. *Proceedings of the SIAM International Conference on Data Mining, SDM*, 401-412

Chang, H., & Yeung, D.-Y. (2008). Robust path-based spectral clustering. *Pattern Recognition, 41(1)*, 191-203.

Delias, P., Doumpos, M., Grigoroudis, E., Manolitzas, P. & Matsatsinis, N., (2015). Supporting healthcare management decisions via robust clustering of event logs. *Knowledge-Based Systems*, 84, pp.203-213.

Günther, C. W. & van der Aalst, W.M.P. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics, *Proc. of BPM 2007*, 328–343.

Jung, J.-Y., Bae, J., & Liu, L. (2009). Hierarchical clustering of business process models. *International Journal of Innovative Computing, Information and Control 5(12)*, 1349–4198.

Kirchner, K. (2015). Toward Automatic Creation of Clinical Pathways. *Proc. of 1st EWG - DSS Int. Conf. on Decision Support System Technology*. Belgrade, Serbia, 32, ISBN: 978-86-7680-313-2.

Kirchner, K., Herzberg, N., Rogge-Solti, A. & Weske, M. (2013). Embedding Conformance Checking in a Process Intelligence System in Hospital Environments, *BPM 2012 Joint Workshop ProHealth 2012/KR4HC 2012*, Tallinn, Estonia, 126-139.

Luengo, D., & & Sepúlveda, M. (2012). Applying Clustering in Process Mining to Find Different Versions of a Business Process That Changes over Time. *Business Process Management Workshops vol. 99*, 153-158.
Mans, R., Reijers, H., van Genuchten, M., & Wismeijer, D. (2012, January). Mining processes in dentistry. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 379-388. ACM.
Poelmans, J., Dedene, G., Verheyden, G., Van Der Mussele, H., Viaene, S., & Peters, E. (2010). Combining business process and data discovery techniques for analyzing and improving integrated care pathways. In *Advances in Data Mining. Applications and Theoretical Aspects*, 505-517. Springer Berlin Heidelberg.
Rebuge, Á., & Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, *37*(2), 99-116.
Rotter, T., Kinsman, L., James, E., Machotta, A., Gothe, H., Willis, J., et al. (2010). Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs. *Cochrane Database Syst Rev*, 3(3).
Song, M., Günther, C., & Aalst, W. (2009). Trace Clustering in Process Mining. In D. Ardagna, M. Mecella, & J. Yang (Eds.),. Business Process Management Workshops vol.17, 109-120.
Van der Aalst, W. (2012). Process Mining Manifesto. Business Process Management Workshops, 69-194. Berlin: Springer.
Veiga, G., & Ferreira, D. (2010). Understanding Spaghetti Models with Sequence Clustering for ProM. Business Process Management Workshops vol. 43, 92-103
Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, *17*(4), 395-416.

ASSOCIATION RULE MINING: MODERN BUSINESS APPROACH

Višnja Istrat*[1], Dajana Matović[1], Milko Palibrk[2]
[1]Faculty of organizational sciences, University of Belgrade
[2] Administration for Joint Services of the Republic Bodies
*Corresponding author, e-mail: visnja.istrat@gmail.com

Abstract: *The business problem has been elaborated by analyzing data through methods and techniques of business intelligence, with emphasis on association rules. In this paper attributes in the air traffic field have been analyzed in order to determine the conditions for landing of the aircraft. Based on the applied data model it could be determined if an aircraft can land on predefined conditions, or not. The business problem has been elaborated in real-world setting. Evaluations of landing conditions and prediction through association rules have big relevance because the safety of the landing of the aircraft with crew could be significantly improved. Research in this paper shows the potential and significance of business intelligence application in air traffic management.*

Key words: *business intelligence, association rules, sustainable business development*

1. INTRODUCTION

Suknović (2001) defines ˝Decision making is the selection of one from the group of available alternatives" (p. 14). It is a very complex process that should result in getting appropriate managers' decisions. Therefore, it is necessary to continuously research and improve methods and techniques of modern decision making. Perfect solution for the defined problem does not exist. The highlight should be on the most acceptable solution. Managers should know well the Theory of decision making, as well as how to use practical experience to get the business decision that will result with maximization of profit.

Suknović (2001) discusses the three dimensions that determine the complete development of this discipline. These are: qualitative, quantitative and information-communication aspects (p. 13). These three aspects implicate all concepts of development of modern decision making, on both theoretical and practical level. ˝Quantitative approach in modern decision making defines the basic formalism of general decision making problem˝ Suknović and Delibašić (2010) point out. Suknović (2001) defines ˝the decision making problem is a five item problem $(\mathrm{A}, \mathrm{X}, \mathbb{F}, \Theta, \succ)$ where (p. 13):

A: represents a definite set of available alternatives (actions) ranked by a session participant in order to select the most acceptable one;
X : represents the set of possible outcomes as a consequence of selecting an alternative;
Θ: represents a set of world states, and depends on the unknown state $\theta \in \Theta$ because the consequences of selecting alternative $\alpha \in \mathrm{A}$ may differ;

$$\mathbb{F} : \mathrm{A} \times \theta \rightarrow \times$$

(1)
for each world state ϖ, and for each alternative α, determines the resulting consequence $\mathrm{X} = \mathbb{F}(\alpha, \varpi)$

$\succ$: weak order relation on X, i.e. a binary relation that satisfy the two following criteria:

i) Completeness: either $\mathrm{X} \succ \gamma$ or $\gamma \succ \mathrm{X}, \forall\, \mathrm{X}, \gamma \in \mathrm{X}$;

ii) Transitivity: if $\mathrm{X} \succ \gamma$ and $\gamma \succ \mathrm{Z}$ then $\mathrm{X} \succ \mathrm{Z}$, $\forall \mathrm{X}, \gamma, \mathrm{Z} \in \mathrm{X}$.

Relation $\succ$ features the decision maker and is called a preference relation Two other important relations can be derived from the preference relation. The first is the strict preference relation where $\mathrm{X} \succ \gamma$ if and only if both $\mathrm{X} \geq \gamma$ and not $\gamma \geq \mathrm{X}$. The second is the indifference relation where $\mathrm{X} \sim \gamma$ if and only if $\mathrm{X} \geq \gamma$ and $\gamma \geq \mathrm{X}$. The most often way of solving decision making problems is the transformation of weak order $\succ$ to X into normal order $\geq$ over the field of real numbers by the means of utility functions. *Authors Anderson, Sweeney, Williams, Camm and Cochran (2015) describe quantitative methods of decision making (p. 154).*

Recommendation is that manager posess relevant experience, knowledge and wisdom in using the techniques and methods of modern decision making. *Behaviour of managers has been described by authors Bazerman and Moore (2012) on (p. 103)*. Intuition is also important in the decision making process, as well as the proactive approach of solving problems. Intersection of different scientific disciplines points out to interdisciplinary framework of modern decision making.

Recommendation is that strategic positioning of business entity is oriented towards life-long learning and application of methods and techniques of modern decision making. The goal is achieving business excellence and leadership position in business environment. *Organizational decision making has been described by author Pettigrew (2014) in book (p. 229). Authors Power, Sharda and Burstein (2015) describe the latest tendencies in business decision support system (p. 20). Efficient strategies for mining association rules have been described by the authors Nguyen, Vo and Le (2014) in their research (p. 4716). Authors Howson and Hammond (2014) discuss the benchmarking of companies from the field of business intelligence with latest business solutions (p. 57). Latest tendencies of decision support system have been described by author Sauter (2014) on (p. 35).*

˝Association rules are useful tool for extracting new information from raw data described in comprehensive manner for decision makers" is defined by authors Ruiz, Gómez-Romero, Martin-Bautista, Sánchez, Vila and Delgado (2015) in their research (p. 247). Concept of association rules and datasets has been discussed by authors Simovici and Djeraba (2014) in their research (p. 647). Mining of association rules through new statistical method has been described by authors Zhang and Shi (2016) on (p. 10). *˝Analysis of relations between different technologies could result with maximization of profit. Association rules have been suggested for determination of dependencies between technologies", Altuntas, Dereli and Kusiak (2015) on (p. 249). Ding, He and Liang (2015) show the comprehensive system of evaluation of association rules (p. 304).*

*Humphrey (2005) describes SWOT analysis (**S**trengths, **W**eaknesses, **O**pportunities & **T**hreats Analysis) as technique of strategic management where strategic choices that connect and evaluate internal strengths and weaknesses of companies with opportunities and threats of external environment˝, (p. 7). SWOT analysis application was described by authors Ayub, Razzaq, Aslam and Igtekhar (2013) in their research (p. 92).* Table 1 shows the overview of significance of SWOT analysis of business decision making for business entity. Internal strengths and weaknesses of the business entity have been analyzed that relate to the process of business decision making. Based on the shown SWOT analysis, it is concluded that strengths and opportunities of the concept of business decision making are much better and more prosperous than the weaknesses and threats.

Table 1: SWOT analysis of significance of business decision making for business entity at the market

Strengths	**Weaknesses**
Higher effectiveness of managers' decisions. Strengthening of process of decision making. More effective management team. Educated human resources with knowledge and experience in decision making. Systematic positioning of concept of business intelligence in decision making process.	Continual costs of training and education of managers. Waiting time to see the effects of decisions. Uncompetitive management. Unavailable high-quality literature of topic of business decision making.
Opportunities	**Threats**
Better positioning at the market. Maximization of profit. Growth and development of business subject as consequence of taken suitable manager's decision. Taking leadership position at the market. Improved communication. Transparency of business.	Poor business of company as result of wrong manager's decision. Bad choice of strategic partnerships that leads towards bankruptcy of the company. Wrong choice of information from external sources that leads towards wrong decision making. Resistance of the environment towards changes.

Source: Author

2. RESEARCH FRAMEWORK

This paper describes the process of data mining concept implementation by creating model of association rules. CRISP-DM methodology and software Orange have been used in the research. *More details about* *phases of CRISP-DM methodology can be seen in research of Suknović and Delibašić (2010)*. Business problem consists of analyzing data of an air traffic field by methods and techniques of business intelligence, with highlight to association rules. Attributes and conditions for landing of an aircraft have been analyzed. Based on the applied data models it will be evaluated if an aircraft could land or not, taking into account weather conditions.

Evaluation of landing conditions and predictions through association rules have big significance because based on this the safety of an aircraft with crew has been evaluated. Datasets from real-term business settings have been used. There were 253 data iterations processed. Attributes that are being researched are: stability, consistency, magnitude, wind, altitude and visibility. Numerical values have been shown in database with *tab (tab delimited file) format, which is suitable for processing in software Orange.

After software evaluation, it was concluded that data in this format were prepared for modeling phase. Modeling is the central and the shortest phase of the process and it is done by numerous software solutions for data mining through algorithms. In further research of database of an air traffic, Apriori algorithm has been used. *More about Apriori algorithm can be seen in book by Adamo (2012) on (p. 33).* Solution will be effectively applied in real-term business settings. Association rules have been generated. (p. 47).

Figure 1 shows desktop of an open-code software Orange, which was used during the research. On the left-handed side there is menu with options for data processing and tools of business intelligence. On the right-handed side there is desktop space where models of business intelligence are being created. Icons are being connected on the certain pattern and therefore models of different purposes have been created. *Authors Rouhani, Ashrafi, Ravasan and Afshari (2016) describe the influence of model of business intelligence to decision making and organizational benefits (p. 19).* Figure 1 shows the model of association rules application that has been used during the research.

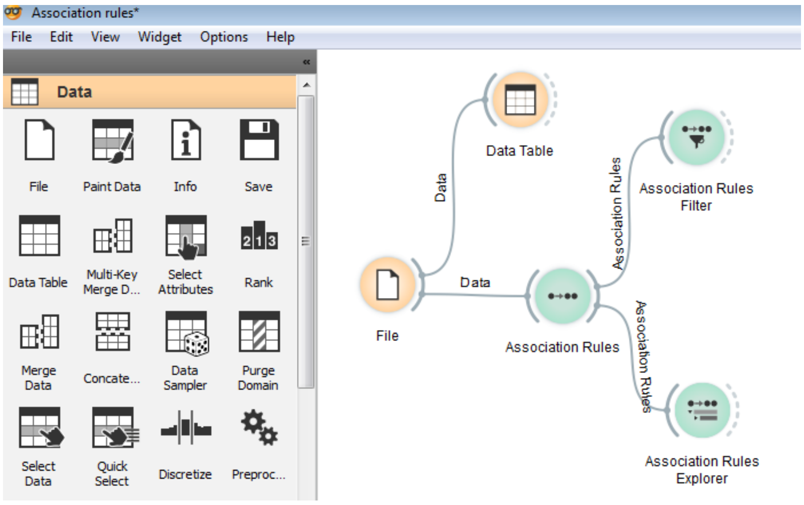

Figure 1: Desktop of software Orange

Further analysis provides overview of data without missing values. The iterations of data are clearly shown, as well as the visibility and data value of the collected data for the research (Figure 2). *Authors Fan and Bifet (2013) describe mining of big databases, as well as future tendencies in this field (p. 3).* After data preparation and immediately before modeling, these characteristics have been shown:

- Database on Figure 2 contains 253 data iterations. Taking into account that data are from real business settings and that it is very difficult to get data from the real business system of air traffic, it is concluded that database is on satisfactory level for the analysis.

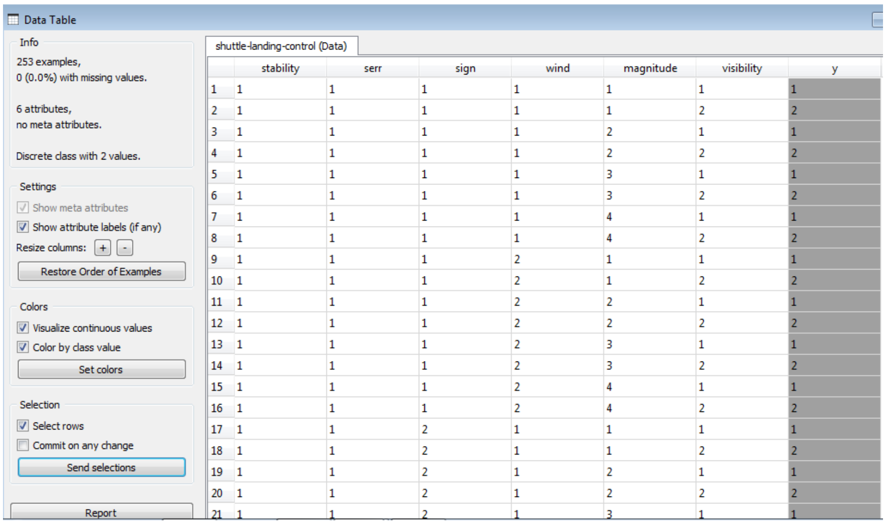

Figure 2: Overview of project data

- After an extensive research of domestic search engines and databases, it is evident that there is no scientific paper that is dealing with the application of association rules as part of business intelligence applied on the real data in air traffic field. Therefore, there is need and interest for research of this field.
- It rarely happens that when preparing data from database there is no missing values. However, when it is case, then further prepare data before processing is needed by special algorithms and processes.. It is concluded that existing database is very suitable for further research.
- Six attributes present the important items that analyze parameters necessary for evaluation of landing the aircraft.

Figure 3 shows list of five association rules by using generic model.

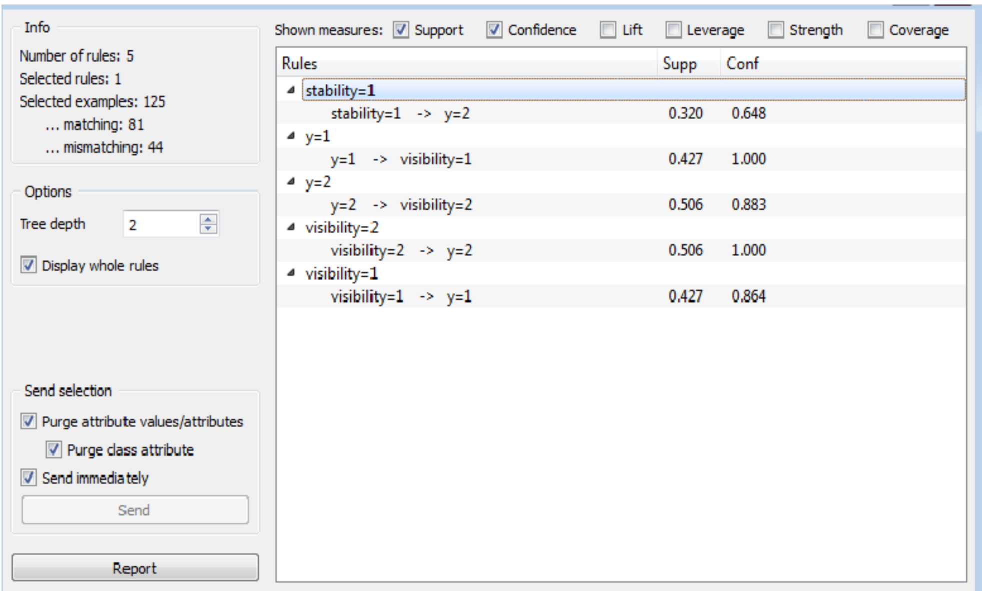

Figure 3: Overview of association rules

For each association rule there are quality measures shown: support and confidence. The first association rule has been highlighted. It can be seen that from 125 selected examples, 81 data examples do not match with first association rule. Further it can be seen that 44 examples do not match with gained association rule.

3. DISCUSSION

There are five association rules:

1. If attribute stability has numerical value 1, than it follows that iteration belongs to class Y 2, which is marked with landing confirmation, with support of 0.320 and confidence of 0.684.
2. If an iteration belongs to class Y 1 (landing rejection), than visibility has been marked with numerical value 1, with support of 0.427 and confidence of 1000.
3. If an iteration belongs to class Y 2 (landing confirmation), than it follows that visibility belongs to numerical value 2, with support of 0.506 and confidence of 0.883.
4. If visibility is marked with numerical value 2, than it follows that iteration belongs to Y 2, than it is recommended to land (confirmation), with support of 0.506 and confidence of 1000.
5. If visibility is marked with numerical value 1, than iteration belongs to class Y 1, as rejection for landing the aircraft, with support of 0.427 and confidence of 0.854.

Analyses of gained association rules have shown that rules number 2 and number 4 have biggest values of quality measures of confidence. It is determined that these two rules have biggest possibility to belong to certain class (for rule number 2 it is class Y 1, and for rule number 4 it is class Y 2). Analyses of support have shown that rules number 3 and number 4 have biggest possibility to fulfill conditions from the IF –part of the rule. Biggest possibility is that attribute belongs to class that is confirmed for landing (rule number 3) and that visibility has been marked with numerical value 2 (association rule number 4).

Based on the practical application of association rules it is determined that there is potential and confidence in calculations with parameters for more precise determination if landing conditions are suitable or not. Significance of gained results is obvious in the possibility of systematic overview of conditions in air traffic and based on that getting appropriate decisions on landing that were helped by this tool of business intelligence.

4. CONCLUSION

Presented model of decision making is helped by the methods of business intelligence of tool of an open-code software Orange has shown application in high effectiveness and accuracy in real-term settings of air traffic. It is recommended to further research and apply the model in the field of air traffic because this field has big potential and it is still relatively unresearched area. Model can be tested and developed in other similar fields of interest. In industry accent is on practical application of sustainable business development. For instance: evaluation of working of cars and conditions for driving; evaluation of various machines in factories; education of managers. In Academia the application of model can be used for the improvement at different levels of studies; use of case studies in the laboratories; students' education and the use of their knowledge and skills for different kinds of application and creation of new model. In public institutions the use of this model can be in Civil Aviation Directorate, also in different agencies for market analysis, etc. Civil society can use the model for personal non-commercial purposes for the research and application of model, in order to gain knowledge and education.

REFERENCES

Adamo, J. M. (2012). *Data mining for association rules and sequential patterns: sequential and parallel algorithms.* Springer Science & Business Media.

Altuntas, S., Dereli, T., & Kusiak, A. (2015). Analysis of patent documents with weighted association rules. *Technological Forecasting and Social Change, 92*, 249-262.

Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2015). *An introduction to management science: quantitative approaches to decision making.* Cengage Learning.

Ayub, A., Razzaq, A., Aslam, M. S., & Iftekhar, H. (2013). A conceptual framework on evaluating SWOT analysis as the mediator in strategic marketing planning through marketing intelligence. *European Journal of Business and Social Sciences, 2(1)*, 91-98.

Bazerman, M., & Moore, D. A. (2012). *Judgment in managerial decision making*.

Ding, S., He, X., & Liang, H. (2015, August). A Comprehensive Evaluation System of Association Rules Based on Multi-index. In 2015 *14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)* , 304-307. IEEE.

Fan, W., & Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM sIGKDD Explorations Newsletter, 14(2),* 1-5.

Howson, C., & Hammond, M. (2014). *Successful Business Intelligence: Unlock the Value of BI & Big Data*. McGraw-Hill Education.

Humphrey, A. S. (2005). *SWOT analysis*.

Hwang, H., & Yoo, K. (2013). Mining Association Rules from the Web Access Log of an Online News website. *Journal of the Korea Industrial Information System Society, 18(2),* 47-57.

Nguyen, D., Vo, B., & Le, B. (2014). Efficient strategies for parallel mining class association rules. *Expert Systems with Applications, 41(10)*, 4716-4729.

Pettigrew, A. M. (2014). *The politics of organizational decision-making*. Routledge.

Power, D. J., Sharda, R., & Burstein, F. (2015). *Decision support systems*. John Wiley & Sons, Ltd.

Rouhani, S., Ashrafi, A., Zare, A., & Afshari, S. (2016). The impact model of business intelligence on decision support and organizational benefits. *Journal of Enterprise Information Management, 29(1).*

Ruiz, M. D., Gómez-Romero, J., Martin-Bautista, M. J., Sánchez, D., Vila, M. A., & Delgado, M. (2015). *Fuzzy Meta-Association Rules.*

Sauter, V. L. (2014). *Decision support systems for business intelligence*. John Wiley & Sons.

Simovici, D. A., & Djeraba, C. (2014). Frequent Item Sets and Association Rules. *In Mathematical Tools for Data Mining,* 647-668. Springer London.

Suknović, M., & Delibašić, B. (2010). *Poslovna inteligencija i sistemi za podršku odlučivanju*. FON, Beograd. In Serbian.

Suknović, M. (2001). *Razvoj metodologije podrške grupnom odlučivanju*. Doktorska disertacija, Fakultet organizacionih nauka, Beograd. In Serbian

Zhang, A., Shi, W., & Webb, G. I. (2016). Mining significant association rules from uncertain data*. Data Mining and Knowledge Discovery*, 1-36.

PREDICTING RISK OF BUYING LOW QUALITY CAR WITH RAPID MINER

Nemanja Mitic[1], Ana Jankovic[1], Milan Vukicevic[1]
[1]Faculty of Organizational Sciences University of Belgrade, Belgrade, Serbia
Corresponding author, Milan Vukicevic e-mail:vukicevicm@fon.bg.ac.rs

Abstract:*In this research, we build and evaluate several models for prediction of risk of buying alow quality used car. Highly accurate predictive model can serve as a decision support for car resellers and private users. Data is provided from the Kaggle.com challenge "Don't Get Kicked" hosted by Carvana.We followed CRISP-DM methodology for building and evaluation of predictive models. Special attention was given to automatic feature selection and algorithm parameter optimization.Complete experimental evaluation is conducted in Rapid Miner – one of the most popular open source Data Science environments.*

Keywords*:Car buying, Predictive modeling, Rapid Miner, CRISP-DM, Feature selection, Parameter optimization*

1. INTRODUCTION

One of the biggest challenges today is to find right product with good quality that fits all our needs. This problem becomes greater when aproduct that we want to buy is already used. In nowadays car market of used automobiles is very rich. And when we go to theauto dealership with the intent to buy a car, we want to make a good selection and to be able to trust the condition of the car that we want to buy. The output of this buying is agreat risk that the vehicle might have serious issues like mechanical issues, tampered odometers etc. The auto community calls these unfortunate purchases "kicks".

In this paper, we are building models that can predict and quantify the risk of buying a car that has low quality. We based our work on CRISP-DM methodology. CRISP-DM methodology is Cross Industry Standard that addressed parts of problems by defining a process model which provides a framework for carrying out data mining projects which are independent of both the industry sector and the technology used.(Wirth andHipp, 2000). Using this methodologywe have followed next steps:

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation
- Deployment

Every phase of this process is important but crucial steps are business and data understanding and preparation of data which took most of the time in our process. We used Rapid Miner for creating our model. Rapid Miner is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics.(Hofmann andKlinkenberg, 2013).RapidMinerwas used for building several processes for automatic building and evaluation of the models, but also feature selection and algorithm parameter optimization. Our main goal was to investigate theperformance of single and ensemble predictive models and to try to further improve them with feature and sample selection techniques. Winner of this competition focused his efforts on feature construction and building of different ensemble models, that are mostly based on LogitBoost and Decision stump models. These algorithms are evaluated in this research. However, we couldn't make a direct comparison with the winner of Kaggle competition, since the exact experimental setting is not published. Some details of winning model can be found in following the blog: http://blog.kaggle.com/2012/02/06/kicking-goals-an-interview-with-marcin-pionnier-of-the-winning-team-in-dont-get-kicked/.

2. DATA UNDERSTANDING AND PREPARATION

Our first major challenge was the preprocessing of data. We obtained our data set from the Kaggle.com challenge "Don't Get Kicked" hosted by Carvana(https://www.kaggle.com/c/DontGetKicked/data). The dataset contained 32 unique features with 72,983 samples along with a labeling of 0 for good car purchases and 1 for "kicks".Some of themain attributes include odometer readings, selling prices, vehicle age and vehicle model.

We observed that output class was highly imbalanced8976 (12%) cases were labeled with 1 (badbuys) while 64007 (88%) cases were labeled with 0 (good buys). Figure 1 shows a number of good and bad buys (y-axis) over different car ages (x-axis). It can be seen that the best selling cars are aged between 2 and 6 years. Additionally, we can see that larger percentage of bad buys occur for older vehicles.

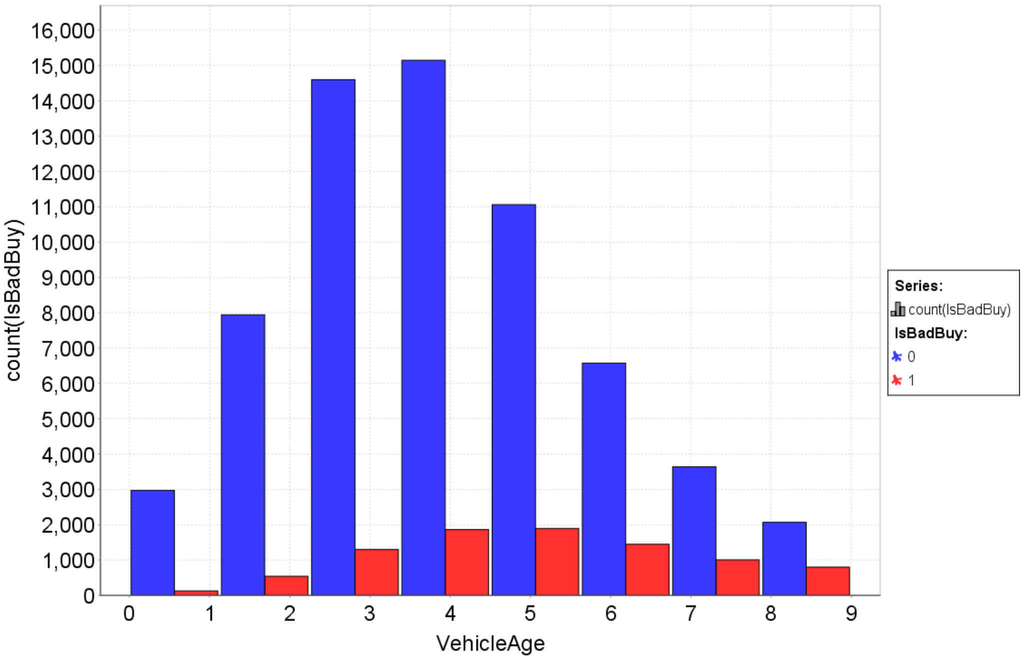

Figure 1: Histogram of good and bad buys over vehicle age

Further, we wanted to investigate theimportance of input features for predicting theoutput.For that purpose, we used three feature weighting techniques available in RapidMiner: Weight by Information Gain, Weight by Information Gain Ratio and Weight by Gini index. Detailed inspection of the results showed that all three feature weighting techniques had a consensus about 5 least important features (“Purch date”, “IsOnlineSale”, “Transmission” , “Nationality” , “Auction” , “VNZIP 1” and “Colour”). This is expected result since these features should not have relation with the car quality. Training data contain 2,674 missing values. Something we have observed immediately is that "good cars" are the most common in this set of data and make 87,7% of the examples and thus we replaced missing with average values. However, we acknowledge that some other type of imputation could lead to better predictive results.

3. MODELING

In this section, we briefly describe predictive models and feature selection techniques used in this research.

3.1. Naïve Bayes

The Naive Bayesian learning uses Bayes theorem with “Naive” assumption of independence between predictors (Russell and Norvig, 2003). Examples are classified based on the posterior probability that an example should be assigned to class.Even though independence assumption is violated in most real world applications, Naive Bayes often demonstrated satisfactory performance in practice (Rish, 2001), and was classified as one of the top 10 algorithms in data mining (Wolfson et al., 2014). Additionally, Naive Bayes models are easy to construct, without any setting or adjusting of complex parameters, computational and time effectiveness. Naive Bayes have the ability to work with large datasets (big data) and provide good interpretability (Hand and Yu, 2001).

3.2. Decision trees

Decision trees (DT) are predictive algorithms based on “greedy”, top-down recursive partitioning of data. DT algorithms perform an exhaustive search over all possible splits in every recursive step. The attribute (predictor) demonstrating the best split by some evaluation measure is selected for branching the tree. Regularly used are information theoretic measures (e.g. Information Gain, Gain Ratio, Gini etc.) or statistical tests quantifying the significance of the association between predictors and class. The procedure is recursively iterated until a stop criterion is met. Greedy strategy for DT building is often criticized, but the ability to build straight forward and the highly interpretable rules on massive data led to many successful applications in medical applications (Chao et al., 2014; Ting et al., 2014). They are also attractive classifiers due to their high execution speed. However, trees derived with traditional methods often cannot be grown to arbitrary complexity for the possible loss of generalization accuracy on unseen data. The limitation on complexity usually means suboptimal accuracy on training data. The essence of this method is to build multiple trees in randomly selected subspaces realize their classification in complimentary ways, and their combined classification can be monotonically improved (Ho,1995). This finding was used for efficient ensemble algorithms that will be described in the further text.

3.3. Ensemble algorithms

Ensemble algorithms combine multiple models in order to provide more accurate or more stable predictions. Ensemble models can aggregate the same model that is built on different subsamples of data, different models built on the same sample or a combination of the previous two techniques. Ensemble methods are often used to improve the individual performance of algorithms that constitute ensembles (Kuncheva and Whitaker, 2003) by exploiting the diversity among the models produced.

3.4.Random Forest

Random Forest (RF) is an ensemble classifier that evaluates multiple decision trees and aggregates their results, by majority voting, in order to classify an example (Breiman, 2001). There is a two level randomization in building these models. First, each tree is trained on a bootstrap sample of the training data and second, in each recursive iteration of building a DT (splitting data based on information potential of features), a subset of features for evaluation is randomly selected. This strategy allows efficient model building, and despite its random nature often provides highly accurate predictive models. In this research, we grew and evaluated Random Forest (RF) with 10 trees (with the default parameters of Weka’s implementation of Random Forest).

Random forestsis a notion of the general technique of random decision forests that are an ensemble learning method for classification, regression,and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfittingto their training set.(Hastie et al., 2008)

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges as to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them (Breiman, 2001).

3.5. Boosting

Boosting is an ensemble technique developed in order to improve supervised learning performance of weak learners (models whose predictive performance is only slightly better than random guessing). Boosting algorithms are built on a principle that subsequent classifiers are adapted to improve predictive performance on those instances that are misclassified by previous classifiers. In this research, we use Logit Boost (Friedman et al., 2000) technique and decision stump (weak learner) as suggested by thewinner of this competition.

3.6. Forward selection and backward elimination

The Forward Selection technique starts with an empty selection of attributes and, in each round, it adds each unused attribute of the given dataset. For each added attribute, the performance is estimated using the internal validation (usually cross-validation). Only the attribute giving the highest increase of performance is added to the selection. Then a new round is started with the modified selection. Backward elimination is using opposed strategy: it starts with all features and removes one by one until stop criteria are met.

4. EXPERIMENTAL EVALUATION

In our experiments, we wanted to compare the performance of several algorithms, and further to try to improve benchmark performance with algorithm parameter optimization, feature selection, and sample selection. RapidMiner process/sub-process structure and highly flexible process control, allowed us to define a process for automatic evaluation ofa number of algorithmsin one execution. Figure 2 shows one part ofthe automatic process. The upper part of the figure shows inner operators of “Validation” operator. “Validation” splits the data on training and test set (in our case 70% and 30% of the initial dataset). Further, In training part of “Validation” we used “Select sub-process” that allows iterative execution of a number of different algorithms (lower part of the figure). Finally, upper-right part of the figure shows part of testing, where models produced by the algorithms are applied and evaluated on the test set. This process also contained a procedure for decision tree optimization. We used grid optimization strategy that is available through “Optimize Parameters” operator. Decision Tree algorithm is optimized based on splitting criterion (gain ratio, information gain, gini index, accuracy) and stop criterion (maximal depth with possible values of 5,10,15,20,25, respectively).

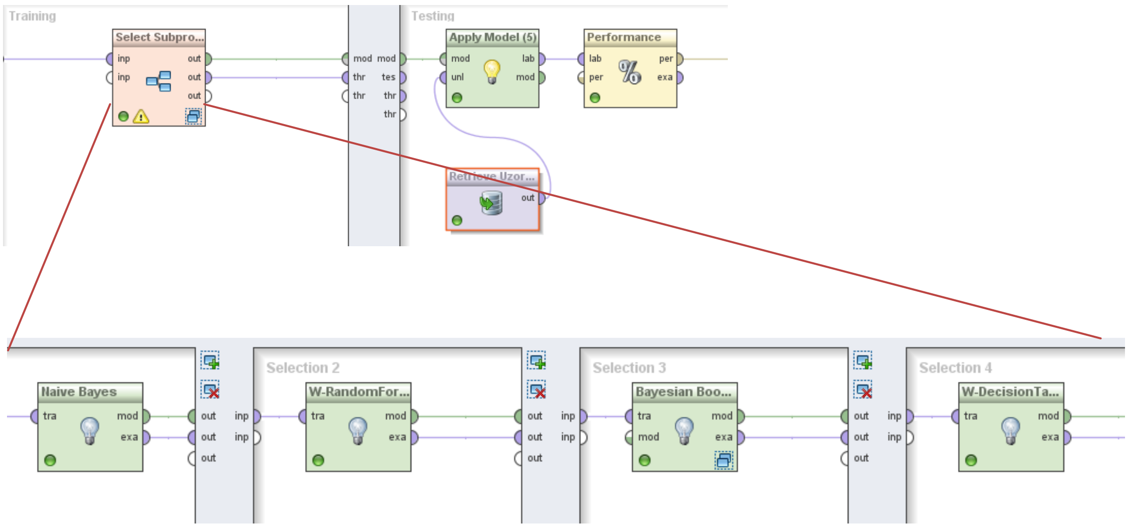

Figure 2: Process for automatic algorithm evaluation

4.1. Evaluation

All experiments are evaluated with three performance measures: Accuracy, F-measure, and Area Under Curve (AUC). Accuracy represents the percentage of testing set examples thatare correctly classified by the model. Even though this measure is frequently used in data mining research, it often results in misleading conclusions if it is not used along with some other evaluation measures. Namely, in the case of imbalanced output data, F-measure (also called F-score) is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score: p is the number of correct positive results divided by the number of all positive results, and r is the number of correct positive results divided by the number of positive results that should have been returned. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst at 0.AUC is defined as the Area Under the Curve of the Receiver Operating Characteristics (ROC) and allows comparison of the general performance of binary classifiers, regardless of threshold selected for predictions. This is why it is often better for model evaluation than F-measure or Accuracy. Additionally, it does not give misleading results for imbalanced data.

4.2. Results

In the first part of the experiment, we evaluated benchmark models that are described in the previous section. Additionally, we also evaluated models that are built on balanced data. We undersampled majority class (leaving 6000 of positive samples and 12000 of negative samples) since this technique often leads to improvement of predictive performance of the models in cases of imbalanced data (He and Garcia, 2009). The results are showed in Table 1. On the left side of the table, we showed model performances for original and on the right side for balanced data. It is important to note that data balancing was conducted only on the training set. However, results show that performance by all evaluation measures and for all algorithms is decreased. This can be attributed to much smaller sample that was available with balanced data. t can be seen that best performance by all measures was achieved by Logit Boost (in both balanced and unbalanced cases). Further, decision tree (with and without optimization) showed the worst results. This is expected since it is widely known that these models have poor performance on imbalanced data.

Table 1: Performances of algorithm

Algorithm performance - training on unbalanced data				Algorithm performance - training on balanced data		
Algorithm	**AUC**	**Acc**	**f measure**	**AUC**	**Acc**	**f measure**
Naïve Bayes	0.682	0.751	0.848	0.680	0.699	0.807
Random Forest	0.696	0.874	0.931	0.663	0.785	0.874
Logit Boost	0.740	0.899	0.943	0.740	0.893	0.941
Decision Table	0.713	0.897	0.944	0.715	0.891	0.940
Decision tree	0.501	0.877	0.934	0.501	0.877	0.934
Decision tree (opt)	0.501	0.877	0.934	0.501	0.877	0.934

Additionally, it can be concluded from the results, that decision trees predicted only negative class since they had accuracy exactly the same as the percentage of the negative class in thetest set (and initial set). Since balancing the data showed adecrease in performance for all models, we excluded this dataset from further experiments.We additionally tried to improve thepredictive performance of our benchmark models by forward selection and backward elimination (described in theprevious section). The results are presented in Table 2.

Table 2: Results of algorithms using optimize selection

Algorithm performance –forward selection				Algorithm performance – backward elimination		
Algorithm	**AUC**	**Acc**	**f measure**	**AUC**	**Acc**	**f measure**
Naïve Bayes	0.658	0.885	0.937	0.751	0.711	0.808
Random Forest	0.658	0.885	0.937	0.751	0.792	0.895
Logit Boost	0.658	0.893	0.941	0.755	0.899	0.971
Decision Table	0.658	0.885	0.937	0.748	0.876	0.945
Decision tree	0.501	0.877	0.934	0.501	0.877	0.934
Decision tree (opt)	0.501	0.877	0.934	0.501	0.877	0.934

It can be seen that forward selection slightly decreased performance and backward elimination slightly increased theperformance of all models except decision tree.

5. CONCLUSION AND FUTURE RESEARCH

There has been a rapid growth of data mining in different application areas like business, healthcare etc.In this research, we exploited RapidMiner's possibilities for making flexible and automated process. We evaluated several single and ensemble predictive models and tried to improve their performance with feature selection, class balancing, and parameter optimization. In general, ensemble algorithms gave the best performance. Predictive performance on unseen cases is additionally improved by backward elimination procedure for feature selection. Undersampling of majority class did not give satisfactory results, most likely because of arelatively small sample of the data. In general, results obtained from presented experiments, show that they could be useful as a decision support for car buying. However, there is still a lot of room for further investigation and model improvement.

As a part of our future work, we plan to pay more attention to data pre-processing. We will try different sampling techniques (such as cluster sampling or systematic sampling), feature construction and selection techniques (i.e. PCA). Additionally, we will pay more attention to handling missing data. Finally, we will try to exploit more available ensemble algorithms as well as Support Vector Machines, Neural Networks etc.

REFERENCES

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32

Chao CM, Yu YW, Cheng BW, &Kuo YL (2014). Construction the model on the breast cancer survival analysis use support vector machine, logistic regression and decision tree. *J Med Syst.* 38(10): 1–7.

Friedman, J., Hastie, T., &Tibshirani, R. (2000).Additive logistic regression: a statistical view of boosting (The annals of statistics, 28(2), 337-407.

Gupta, R andPathak, C. (2014), A Machine Learning Framework for Predicting Purchase by Online Customers based on Dynamic Pricing, Elsevier Science, 599-605.

Hand DJ, Yu K. (2001) Idiot's Bayes—not so stupid after all?*International Statistical Review*.69 (3): 385–398.

Ting H, Mai YT, Hsu HC, Wu HC, & Tseng MH (2014) Decision tree based diagnostic system for moderate to severe obstructive sleep apnea. *J Med Syst*. 38(9):1–10.

Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005).The elements of statistical learning: data mining, inference, and prediction.The Mathematical Intelligencer, 27(2), 83-85.

He, H., & Garcia, E. A. (2009).Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on*, *21*(9), 1263-1284.

Ho, A., Romano, R., & Wu, X. A. (2012).Don't Get Kicked-Machine Learning Predictions for Car Buying

Zhang H. (2004), The optimality of naive Bayes, AA, 1(2), 3

Hornick, M. F., Marcadé, E., &Venkayala, S. (2010). Java data mining: strategy, standard, and practice: a practical guide for architecture, design, and implementation. Morgan Kaufmann.

Hofmann, M., &Klinkenberg, R. (Eds.). (2013). RapidMiner: Data mining use cases and business analytics applications. CRC Press.

Kuncheva LI, Whitaker CJ (2003).Measures of diversity in classifier ensembles.*Machine Learning*.51: 181–207.

Rish I.An empirical study of the naive bayes classifier (2001)In: IJCAI 2001 Workshop on empirical methods in artificial intelligence; 41–66.Available: http://www.cc.gatech.edu/~isbell/reading/papers/Rish.pdf.

Russell S, Norvig P (2013). Artificial intelligence: a modern approach 2nd ed. Prentice Hall Series.

SameerchandPudaruth (2014), Predicting the Price of Used Cars using Machine Learning Techniques, International Journal of Information & Computation Technology, 4(7), 753-764.

Vipin Kumar (2014),Rapid Miner, Data mining use cases and business analytics applications,University of Minnesota, Department of Computer Science and Engineering Minneapolis, Minnesota, U.S.A.

Wah, Y. B., Ismail, N. H., & Fong, S. (2011). Predicting car purchase intent using data mining approach.In *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2011 Eighth International Conference on (Vol. 3, pp. 1994-1999).IEEE.

Wolfson J, Bandyopadhyay S, Elidrisi M, Vazquez-Benitez G, Musgrove D, Adomavicius G, et al. (2014) A Naive Bayes machine learning approach to risk prediction using censored, time-to-event data. ARXIV [Internet]. 2014; Apr 8. Available: http://arxiv.org/abs/1404.2124.

PREDICTIVE MODELLING FOR CLAIMS PROCESSING – CASE STUDY OF BNP PARIBAS

Stefan Sreckovic[*1], Milija Suknovic[1], Milan Vukicevic[1]
[1]University of Belgrade, Faculty of Organizational Sciences
[*]Corresponding author, e-mail: vukicevicm@fon.bg.ac.rs

Abstract: *With the rising amount of business operations, claims management became an issue in insurance companies in order to process them accurately and fast at the same time. In this paper, a model is proposed for classifying new claim requests in BNP Paribas Cardif to ones that meet requirements for faster payments, and others that need some additional information before further processing. Training data set contained over 5000 instances, which represent claim requests, and more than 130 attributes. The data set was provided by www.kaggle.com as a part of 'BNP Paribas Cardif Claims Management' data mining contest. Using different algorithms, accuracy of 78.47% and AUC of 69.93% was achieved.*

Keywords*: predictive modeling, classification, insurance, claims management, ETL*

1. INTRODUCTION

BNP Paribas Cardif is a subsidiary of BNP Paribas Group based in France, which specializes in life and property insurance. It was founded more than 40 years ago and today operates in 36 countries while employing about 9000 people. Over 90 million people trusted BNP Paribas Cardif with insurance of their life and property, mainly in Europe, Asia and South America.

With the growing volume of business and rising amount of claims received, there is an issue of documentation processing speed required for the payment of insurance premiums. It is necessary to accurately process a lot of papers in a short period of time. With this in mind, people from BNP Paribas Cardif decided to set up their business problem to Kaggle platform in the hope of finding a model that will most accurately predict whether the claim contains a complete documentation so it can be processed faster or customers may need additional information before insurance company approves it.

The goal was to develop a model which precise predict the new claims requests, whether they are complete or not, depending on various attribute parameters. That model would speed up the claims management processes in an insurance company, reduce costs and would also allow the company to serve better customer experience to the customers and thus increase their satisfaction.

To develop a prediction model, software called RapidMiner was used, which is one of the leading freemium solutions on the market for predictive analytics. It supports the entire ETL (extract, transform, and load) process, pre-processing and visualization of data, statistical data processing, development of predictive models and much more. It supports various add-in extensions including Weka, R, Python and other which even more expand its possibilities.

2. STATE-OF-THE-ART

(Smith et al., 2000) were one of the first to publish a case study that involves data mining in the insurance industry. Their paper presents a case study involving two such problems and solves them using a variety of techniques within the methodology of data mining. The first of these problems was the understanding of customer retention patterns by classifying policyholders as likely to renew or terminate their policies. The second was better understanding claim patterns, and identifying types of policy holders who are more at risk. Each of these problems impacts on the decisions relating to premium pricing, which directly affects profitability. A data mining methodology was used which views the knowledge discovery process within a holistic framework utilizing hypothesis testing, statistics, clustering, decision trees, and neural networks at various stages. (Derrig and Ostaszewski,1995) have also provided an overview of fuzzy pattern recognition techniques in their paper and have used them in clustering for risk and claims classification

Some authors (Della Rocca, Johnson, 2010) were pointing out that increasing consistency in the claim-handling process not only improves policyholder satisfaction but also delivers measurable results to enhance a carrier's bottom line. Advanced analytic technology and data-collection techniques can help an insurer settle similar claims in similar ways, leading to an adequate claim reserves, improved loss cost management, and more equitable settlements.

(Lentz, 2013) was pointing that some publications and vendors appear to promote predictive modeling as a way to automate decisions and allow a claim department to do more with fewer people. However, the principal goal of carriers using this technology is a better allocation of time and talent of claim professionals. Its purpose is to assist the claim professional through the course of a claim and provide supplemental information to be used in the decision-making process. While predictive modeling may enhance good claim handling, it cannot rectify poor claim handling. It is not a substitute for proper claim management methods and does not take the place of individual claim handling analysis.

(Rose, 2013) proposed that there are six areas where analytics can make a difference in claims management: fraud, subrogation, settlement, activity and litigation. He stated that adding analytics to the claims life cycle can deliver a measurable ROI with cost savings and that just a 1 percent improvement in the loss ratio for a $1 billion insurer is worth more than $7 million on the bottom line. (Derrig and Ostaszewski, 1995) have provided an overview of fuzzy pattern recognition techniques in their paper and have used them in clustering for risk and claims classification

Regarding fraud in claims management, it was discussed in many papers. (Phua et al., 2010) made a comprehensive survey that has explored almost all published fraud detection studies at the time, including claims. It was found that (Von Altrock, 1996) was able to achieve better claims fraud detection performances with his algorithms than the experienced auditors, while (NetMap, 2004) reported that visual analysis of insurance claims by the user helped discover the fraudster. (Huang, 1997) was using K-means algorithm for clustering large categorical data sets in his study and states that same technique can be used for detecting under-represented concepts like a fraud in a very large number of insurance claims (Ngai et al., 2011) made an exhaustive review of application of data mining techniques in financial fraud detection. (Sundarkumar and Ravi, 2015) proposed a novel technique for predictive modeling on imbalanced datasets in banking and insurance. This technique is based on under-sampling of training data with majority class. (Shi et al., 2015) proposed a series of multivariate negative binomial models for insurance claim counts.

(Frees and Valdez, 2008) were using negative binomial regression model for assessing claims frequency in automobile insurance records. Besides their second model was a multinomial logit to predict the event of a claim and severity component as a third one. When taken together, the integrated model was allowing more efficient prediction of automobile claims compared with than traditional methods. Using simulation, they demonstrated that by developing predictive distributions and calculating premiums under alternative coverage limitations.

A Towers Watson survey of Property and Casualty claim officers (Towers Watson, 2012) found that 63% of respondents have started to explore the use of predictive modeling in their claim operations. The survey further stated that 7% have already been using claim analytics for more than three years. (Lentz, 2013) proposed that the use of predictive modeling tools in claims is not meant to supplant the claim handler’s own thought process; its purpose is to provide decision support for his or her individual analysis. As with other forecasting tools, predictive models indicate probabilities, not certainties.

At BNP Paribas Cardif Claims Management competition, training data set given to data analysts to create predictive classification models. (Lentz, 2013) was pointing that claim records with missing, incomplete, or erroneous data have a negative impact on the accuracy of the predictive model. As such, efforts to ensure clean and accurate data must be undertaken. This process may be long and costly for some insurers and more so for those with multiple legacy systems.

3. DATA

The data sets for this paper have been taken from the online predictive modeling and analytics competitions platform called Kaggle (https://www.kaggle.com/) and the theme was called “BNP Paribas Claims Management”. Data sets are divided into training and test set. Training data had over 114.000 instances and 131 attributes. Besides them, there are two special attributes, ID or row number and Target, which represents output attribute or label with the values 0 and 1, where 0 means that claim documentation is not complete and thus needs further verification before payment, while 1 means that the claim documentation is complete and can be processed faster. There was a slight class imbalance, considering that there were 87 000 instances of class 1 and only 27.000 cases of class 0. In Figure 1 class distribution is shown.

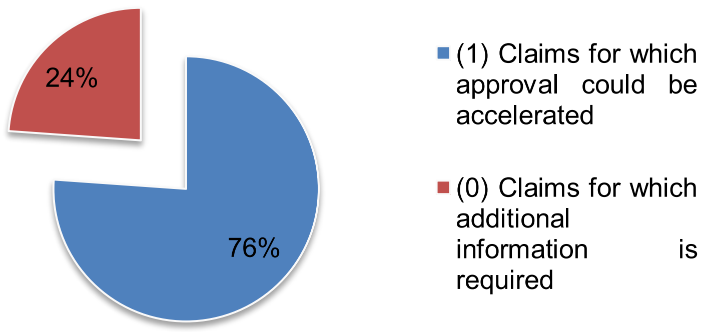

Figure 1: Class distribution

The attributes are of type integer, real or polynomial, which represent the vast majority. Also, attributes are unlabeled and marked with “v1, v2… v131” so I was denied of the meaning of each of them. On one hand, it prevents us to become more familiar with the data and eventually gather some insight of the problem, while on the other hand, compels us to focus on the very construction of the model predictions.

After loading the data, the first thing noticed was that there are a lot of missing values. Exactly 102 attributes have more than 40 percentage of missing values. As a result, almost 50.000 instances, or more than 40 percent of all user cases have around 70 percent of missing values. As the database contained over 100.000 instances and over 130 attributes, due to hardware limitation, experiments are conducted on 5.000 randomly sampled.

4. EXPERIMENTAL SETTING

In order to develop a highly accurate model for claim requests classification, we used the following predictive algorithms: Neural Net, Decision Tree, K-NN, W-Random Forest, Logistic Regression and Bayesian Boosting. Before the initial commissioning test, all the missing values were replaced with median values of attributes. It is important to note that other types of imputation techniques could be used for missing value imputation (i.e. model-based imputation) and possibly improve the final predictive performance of the models). Since we aimed to compare the performance of different predictive models and different feature selection techniques on a specific problem, detailed analyses of models for missing value imputation was out of the scope of this paper.

Polynomial attributes were converted to binominal given that Logistic Regression and Neural Net do not work with categorical values. Bearing this in mind, the attribute v22 was removed from further testing due to the fact that there are more than 17.000 categorical values and trying to convert its type to numerical or even binominal, each time the computer ran out of memory, the process froze and the operation could not be executed. Because of complexity and memory requirements of the algorithms (and optimization procedures) used in this research, original training set that contained 100.000 impressions, was reduced to 5.000 instances. In order to preserve distribution of the output class in the training set, stratified sampling technique was used.

In order to prevent overfitting that could be caused by extensive optimization of algorithm parameters and feature sets, inner validation has been used. Each model was trained on 70 percent of training data and validated on 30 percent. Finally, estimation of the future performance of optimized models was conducted on a complete test set that contained over 110.000 instances.

To measure the quality of selected models, I have decided to use accuracy and AUC. Accuracy as a measure represents the proportion of truly predicted cases over the entire data set. In this case, it is calculated as the sum of the number of predicted insurance claim requests that contain complete documentation which they truly do and the number of claims that do not have complete documentation which they truly don’t, in relation to the number of the sum of all correct and incorrect predictions of a model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The area under the curve, known as AUC, presents a measure of the discriminatory power of classifiers. It is very good for measuring performances of binary classifiers and it takes values from zero to one. When the value is above 0.5 it is a positive class prediction, and when is between 0.7 and 0.8 it is a good classifier, between 0.8 and 0.9 very good, and above 0.9, it is excellent.

5. RESULTS

A comparative review of the performance of the algorithms is shown in Table 1. Results are obtained through testing the model of the initial set and then adding weighting to the attributes with operators "Weight by Gini" and "Weight by SVM ".

Regarding the initial model, only necessary changes of obtained training data set were carried out, so that all algorithms work properly. All categorical attributes were converted into numerical because Neural Net and Logistic Regression do not work with categorical ones, the missing values have been replaced with average values of specific attributes and already mentioned attribute v22 was removed from the training set.

Best performance was achieved with Bayesian Boosting model. It is based on Bayesian theorem and it is used to improve the performance of other algorithms, in this case, W-Decision Table. The accuracy obtained with this model was 78.00%; however, the Neural Net and W-Random Forest were not far behind with 76.13% of accurate claims predictions.

Table 1: Initial performances of algorithms

	Initial model		Weight by Gini		Weight by SVM	
Algorithm	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**
Neural Net	68.76%	76.13%	71.08%	76.13%	66.43%	76.67%
W-Random Forest	61.01%	76.13%	65.49%	76.47%	63.86%	76.13%
Bayesian Boosting - W Decision Table	69.91%	78.00%	69.93%	78.47%	68.93%	78.00%
K-NN Classification	50.00%	67.47%	50.00%	68.07%	50.00%	65.80%
Logistic Regression	69.63%	75.40%	68.17%	64.47%	69.90%	66.80%
Decision Tree	50.00%	75.73%	50.00%	76.00%	50.00%	75.93%

Additionally, we compared model performances in synergy with feature weighting and selection scheme. Weighting factors have been assigned to attributes via operator "Weight by Gini". After assigning weights, by the operator "Select by weight" I have influenced the selection of attributes which will be used in modeling algorithms.

The best performances were achieved by filtering the attributes that had normalized weight value greater than 0.26. Again, Bayesian Boosting algorithm gave the best results of all algorithms, and its performances were slightly increased compared to initial model: accuracy at 78.47% and AUC 69.93%. Other algorithms have also made a small progress in their performances, as it is shown in Table 1. The exception is Logistic regression which showed significantly poorer performance after this process.

By assigning weighting coefficients to the attributes using operator "Weight by SVM" and afterward filtering them, only Neural Net model showed improved accuracy (76.67%). However, AUC was only 66.43% comparing to satisfying 71.08% when the attributes were selected after weighting them using "Weight by Gini".

Further, parameter optimization is conducted. Starting from K-NN algorithm, k parameter and distance measure were optimized using validation so that model would not over fit. With values k=9 and using "Mixed measures" as distance measures to nearest neighbors, accuracy was 77.20% and Area under the curve was 66.07%, which is far better than the performance that I have got in previous interactions.

Optimizing Logistic regression and its parameter C, I have also been able to get better performance of the algorithm. In this case, accuracy was 77.20% while AUC was at 67.88%.

When it comes to Decision tree algorithm, by optimizing depth of the tree at 17, minimal gain at 0.03 with Gini index, criterion and turned off pruning option, results became much better. The accuracy of the model was 78.33% while the AUC was 67.54%, which is significantly better than the original 50.00% gained with the initial model parameter settings.

Table 2: Final performances of algorithms

Algorithm	AUC	Accuracy	Obtained by
Neural Net	71.08%	76.13%	Weight by SVM
W-Random Forest	65.49%	76.47%	Weight by Gini
Bayesian Boosting - W Decision Table	69.93%	78.47%	Weight by Gini
K-NN Classification	66.07%	77.20%	Optimize parameters
Logistic Regression	67.88%	77.20%	Optimize parameters
Decision Tree	67.54%	78.33%	Optimize parameters

Best performances by algorithms that have been obtained in this paper are shown in Table 2. By optimizing parameters of the algorithms, we were able to improve results compared to the initial model settings, especially with the Decision tree. Although all algorithms have approximately same performance, the Bayesian Boosting ensemble stands as the best algorithm outs with an accuracy of predicting new claim requests at BNP Paribas Cardif at 78.47%.

Predictive models for claims management that were obtained in this paper could be used by insurance organizations employees that deal with claims on daily basis in order to classify new requests; whether the claims payout process could be speed up, leading to higher customer satisfaction and cutting the costs of employees efforts analyzing requests, or the claims documentation should be revised. It could help employees make faster and more accurate decisions, leading to lower costs and higher efficiency.

6. CONCLUSION

In this research, we compared the predictive performance of several predictive models for classification of new claim requests in BNP Paribas Cardif. Development of accurate predictive model could serve as decision support for claims approval: weather approval could be accelerated, leading to faster payments, or it requires additional information before payout. Development of highly accurate predictive model could help the BNP Paribas Cardif insurance company to faster cultivate claims, reduce costs and increase the satisfaction of its clients. In order to improve predictive performance of the models, we conducted a number of experiments that include optimization of model parameters as well as feature sets. The results that were obtained in this research were satisfying, but there is certainly space for improvement. In our future work, we plan to conduct experiments on the larger training set, to investigate the performance of different methods for missing value imputation and to use advanced predictive methods like Support Vector Machines and Deep Learning based on Neural Networks.

REFERENCES

Della Rocca, R. & Johnson J. (2010). Advanced Claim Analytics and Data Mining. Retrieved April, 2016. from http://www.propertycasualty360.com/

Derrig, R. A., & Ostaszewski, K. M. (1995). Fuzzy techniques of pattern recognition in risk and claim classification. *Journal of Risk and Insurance*, 447-482.

Frees, E. W., & Valdez, E. A. (2008). Hierarchical insurance claims to model. *Journal of the American Statistical Association*, *103*(484), 1457-1469.

Huang, Z. (1997, May). A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. In *DMKD* (p. 0).

Lentz, W. (2013). Predictive Modeling - An Overview of Analytics in Claims Management. Retrieved April 2016. from http:// http://media.genre.com/documents/ii1311-en.pdf

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559-569.

Netmap. (2004). Fraud and Crime Example Brochure.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

Rose, S. (2013). Predictive Claims Processing: Transforming the Insurance Claims Life Cycle Using Analytics. Retrieved April, 2016. from http://www.sas.com

Shi, Peng, and Emiliano A. Valdez. "Multivariate negative binomial models for insurance claim counts." *Insurance: Mathematics and Economics* 55 (2014): 18-29.

Smith, K. A., Willis, R. J., & Brooks, M.. (2000). An Analysis of Customer Retention and Insurance Claim Patterns Using Data Mining: A Case Study. *The Journal of the Operational Research Society*, *51*(5), 532–541. http://doi.org/10.2307/254184

Sundarkumar, G. G., & Ravi, V. (2015). A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance.*Engineering Applications of Artificial Intelligence*, *37*, 368-377.

Towers Watson (2012). Insights: Property & Casualty Claim Officer Survey.

Von Altrock, C. (1996). *Fuzzy logic and neurofuzzy applications in business and finance*. Prentice-Hall, Inc..

LOGICAL CLASSIFICATION METHOD FOR BANKRUPTCY PREDICTION

Ana Poledica*[1], Dragana Marković[1], Sanja Živančević[1]
[1]University of Belgrade, Faculty of Organizational Sciences,
*Corresponding author, e-mail: ana.poledica@fon.bg.ac.rs

Abstract: *The purpose of this paper is to introduce a novel logical classification (LC) method and further evaluate its application for bankruptcy prediction. Logical classification is based on nearest prototype classifier that employs a novel similarity measure and logical aggregation based on interpolative Boolean algebra (IBA). In this research logic-based classifiers are compared to Altman Z score as a benchmark model for bankruptcy prediction. Their classification performance is evaluated on Serbian middle-sized companies. The analysis shows that LC models are reasonable good classifiers and that they outperform Altman Z score.*

Keywords: *logical classification, similarity, IBA similarity measure, bankruptcy, company*

1. INTRODUCTION

Bankruptcy prediction is one of the main issues threatening many companies and governments. Bankruptcy is the inability of debt repayment to creditors. According to Chaudhuri (2013) financial distress begins when an organization is unable to meet its scheduled payments or when the projection of future cash flows points to an inability to meet the payments in near future. Chaudhuri (2013) also describes bankruptcy prediction as a complex process that consists of numerous inseparable factors. The causes leading to business failure and subsequent bankruptcy can be divided into economic, financial, neglect, fraud, disaster and others (Chaudhuri, 2013). With more accurate bankruptcy detection techniques, companies could take some preventive measures in order to minimize the risk of falling to bankruptcy.

Developing different methods for bankruptcy prediction is a popular subject among different authors who addressed the problem of finding the more accurate technique for this problem. There are two dominant approaches when it comes to predicting bankruptcy. The first one is based on statistical methods and was pioneered by Beaver (1966) and Altman (1968) who used multi-discriminant analysis. Beaver (1966) used univariate approach and found that net income to total debt had highest predictive ability. Olson (1980) and Zmijewski (1984) developed stochastic models such as logit and probit. As parametric models dominated the research focus at beginning, studies using non parametric models started to develop in late 1980s. Besides classical methodology many researchers have started introducing numerous alternative methods to predict bankruptcy. In the past decade, three important models namely Bayesian, Hazard and Mixed logit have been applied to bankruptcy prediction (Chaudhuri, 2013). According to Chaudhuri (2013) these models have subtle theoretical advantages over previous ones. He also points out that prior knowledge, practical estimates and subjective preferences are easy to incorporate into the model, either simultaneously or separately. The advantage of Hazard model over logit is that the former explicitly models bankruptcy not as process that happens at a point in time, an assumption made by all previous models, but as process that lasts for period of firm’s life (Chaudhuri, 2013). Hazard model is preferable as it incorporates time varying covariates. The improvement of mixed logit model over logit is that it takes into account both observed information and unobserved information. Under this setting, there are two means to model unobserved information. Second approach utilizes artificial intelligence (AI) and adapts it for predicting bankruptcy. Min and Jeong (2009) described AI methods which include a decision tree, fuzzy set theory, case-based reasoning, genetic algorithm, support vector machine, data envelopment analysis, rough sets theory, and several kinds of neural networks such as BPNN (back propagation trained neural network), PNN (probabilistic neural networks), SOM (self-organizing map), Cascor (cascade correlation neural network). Kumar and Ravi (2007) has given an extensive review of bankruptcy prediction in banks and firms via statistical and intelligent techniques studies. Comparative analysis of data mining methods for bankruptcy prediction is given in (Olson, Delen, & Meng, 2012).

In many applications a good classifier should not only offer superior class predictions but also interpretation and understandability of the classification results. For instance, a manager wants to understand the characteristics of the situation that led to the decision proposed by the model (Kriegel, & Schubert, 2006). In this paper, we propose a novel similarity-based classifier - logical classification (LC). LC algorithm is based on the nearest prototype (centroid) classifier with a difference that it employs a similarity measure based on interpolative Boolean algebra (IBA) (Radojević, 2000) and pseudo-logical aggregation (Radojević, 2008) for measuring similarities. The proposed classifier has strong mathematical background and is computationally efficient technique. It can also provide semantic information about classification results which is another important aspect of designing a classification algorithm. Further in the paper, the proposed logic-based classification approach is evaluated in bankruptcy prediction of Serbian middle-sized companies.

The paper is structured as follows. First of all, in Section 2 we present briefly theoretical concepts and methods used in this paper – Altman Z-Score model, IBA similarity measure and logical aggregation. In section 3 a novel logical classification method is presented. Further, Section 4 describes our company dataset and how the experiment is conducted. Here we also present and discuss results. In the final section we conclude the paper.

2. THEORETICAL BACKGROUND

2.1. Altman's Z-Score model

The first multivariate model for bankruptcy prediction of companies was Z-score (publicly traded firms) by Altman (1968). This model has proved to be the most efficient and has been in use for more than four decades. In determining the Z-score model it is necessary to divide sample into two groups, one group for bankrupt and the other for the non-bankrupt companies. For his research Altman divided 66 companies into two groups and used five out of 22 financial indicators for multivariate discriminant analysis. Altman (2000) found that the following five ratios provide the best results of Z-score model:

$$Z = 1.2X1 + 1.4X2 + 3.3X3 + 0.6X4 + 0.999X5 \quad (1)$$

Where:
X1 = Working Capital / Total Assets;
X2 = Retained Earnings / Total Assets;
X3 = EBIT (Earnings before Interest and Taxes) / Total assets
X4 = Market Value of Equity / Book Value of Total Liabilities
X5 = Sales / Total Assets

1) The Working capital/Total assets ratio (X1) is a measure of the net liquid assets of the companies relative to the total capitalization.
2) The Retained Earnings/Total assets ratio (X2) refers to the earned surplus of a company over its entire life.
3) The Earnings before interest and taxes/Total assets ratio (X3) is a measure of the true productivity or profitability of the assets of a firm. It is not affected by any tax or leverage factors. It reflects the earning power of the assets that determines the value of assets.
4) The Market value equity/Book value of total liabilities ratio (X4) shows how much the assets of a firm can decline in value (measured by market value of equity plus debt) before the liabilities exceed the assets and the firm becomes insolvent.
5) The Sales/Total Assets ratio is the standard capital-turnover ratio illustrating the sales generating ability of the assets of a firm. It refers to the capability of management in dealing with competitive conditions. This ratio was dropped in the Z"-Score model.

Altman's Z-score lower than 1.81 indicates that the company is heading for bankruptcy. Companies with scores above 2.99 are unlikely to enter bankruptcy. Scores between 1.81 and 2.99 mean that the company is in a gray area.

2.2. IBA similarity measure

Similarity between two objects is perceived differently when diverse similarity measures are applied (Poledica et al., 2015). From the standpoint of logic, similarity may be perceived by means of logical relations e.g. bi-implication and equivalence. In (Poledica et al., 2015; Poledica et al., 2013) IBA equivalence is introduced as a logic-based similarity measure where similarity is measured using [0,1]-valued realization of equivalence relation in interpolative Boolean algebra.

Interpolative Boolean algebra (Radojević, 2000) is [0,1]-valued realization of Boolean algebra (BA). It is a consistent framework in the sense it preserves all Boolean axioms (e.g. excluded middle, contradiction, etc.). In IBA logical Boolean functions are uniquely mapped to corresponding generalized Boolean polynomials

(GBP). Introducing intensity of relation, i.e. gradation of Boolean variables and functions, IBA is suitable for many real problems thanks to the descriptiveness of gradation.

When it comes to modeling similarity, logical relation of equivalence in its original form treats only two values: 0 and 1. However, human perception of similarity is rather graded. To calculate the degree of similarity between objects on the whole [0,1] interval, the GBP for logical expression of equivalence is first determined. Finally, its corresponding realization on the valued level is:

$$S_{IBA}\left(a,b\right)=1-b-a+2\cdot a\otimes b=1-b-a+2\cdot\min\left(a,b\right) \tag{2}$$

Radojević (2000) defined the mapping procedure of Boolean functions into GBPs, and the software support is provided in (Milošević et al., 2014).
In (Poledica et al., 2015; Poledica et al., 2013) it is also shown that the previous expression satisfies all properties for similarity measure: non-negativity, symmetry and limited range. In addition to strong mathematical background, IBA similarity measure is valuable from the aspect of interpretation. When comparing two objects it measures the degree to which a certain quality is present but also includes the degree of the absence of that quality as equally important. The graphical interpretation of similarity based on IBA equivalence is also provided in (Poledica et al., 2015). In Figure 1 $\min\left(a,b\right)$ describes how much two objects are the same with respect to a certain feature, and $1-a-b+\min\left(a,b\right)$ defines how much two objects are the same regarding the absence of that feature.

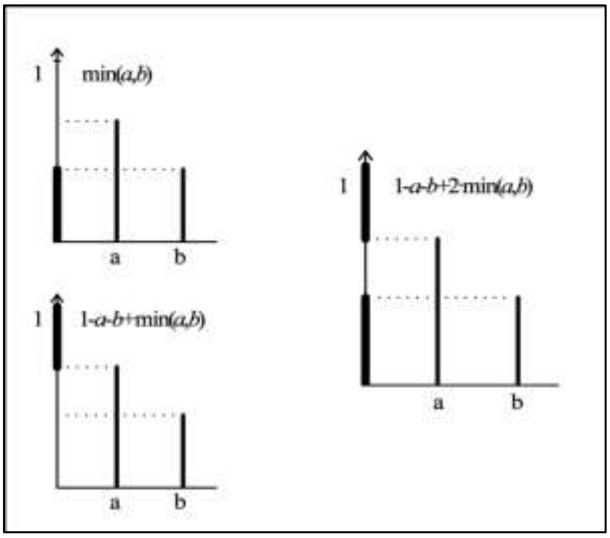

Figure 1: Graphical interpretation of IBA similarity measure

Additionally, similarities based on IBA equivalence may be further aggregated with logical aggregation (Radojević, 2008) which is also based on IBA.

2.2. Logical aggregation

Logical aggregation (LA) is a consistent and transparent procedure based on IBA for aggregating factors (Radojević, 2008). The main idea is to create aggregation functions in form of logical expressions which are used to aggregate primary/combined attributes into a single resulting value. It has two steps:

- Normalization of attributes' values:

$$||\bullet||:\Omega\rightarrow\left[0,1\right] \tag{3}$$

- Aggregation of normalized attributes' values into one resulting value by logical or pseudo-logical function used as a LA operator:

$$Aggr\left[0,1\right]^{n}\rightarrow\left[0,1\right] \tag{4}$$

Pseudo-logical function, called pseudo GBP, is a linear convex combination of generalized Boolean polynomials. Logical aggregation depends on choosing: the measure of aggregation (structural function of pseudo-logical function) and operator of generalized product. As a special case, LA operator can be expressed in form of other well-know aggregation operators such as weighted sum, arithmetic mean, min or max function, discrete and generalized Choquet integral, etc. Logical dependence between aggregating factors may also be taken into account (Radojević, 2008). Being an expressive aggregation tool, LA is able to build various aggregation functions that depend on the nature of a problem (e.g. Milošević et al., 2013; Poledica et al., 2010).

3. LOGICAL CLASSIFICATION METHOD

In this paper, we propose a novel similarity-based classifier - logical classification (LC) method. It is based on nearest prototype (centroid) classifier in sense that it assigns to an instance the class of training set whose prototype is the closest to the instance. The proposed method is 'logical' as it employs logical relation of equivalence for measuring similarity. More precisely, multi-valued [0,1] realization of equivalence - IBA equivalence and LA tool are used to calculate similarities in the classification model. Thus, it is possible to build classifiers that provide semantic information about classification results.

Logical classification is a straightforward procedure and has the following steps:

- Scaling between [0,1] of training and test dataset
- Calculation of prototypes for every class
- Calculation of similarities between test set and prototypes
- Assigning class to test set

The problem is to classify dataset X into n different classes $C_1, C_2, ..., C_n$. Every data instance is described by different m attributes $a_1, a_2, ..., a_m$ that can be expressed or normalized in values between 0 and 1. This is a mandatory condition in IBA framework that is for IBA similarity and logical aggregation. Prototype vectors p_i for each class are represented by aggregated attributes for all instances in the training set of size dim.

$$p_i\left(a_1, a_2, ..., a_m\right) = \left(Agg_{j=1}^{\dim}\left(a_{1j}\right), Agg_{j=1}^{\dim}\left(a_{2j}\right), ..., Agg_{j=1}^{\dim}\left(a_{mj}\right)\right), \quad i = 1, ..., n \tag{5}$$

In this paper, to obtain attributes of prototype vectors for each class we used simple mean aggregation, as a special case of LA aggregation. Further, we calculate similarities of corresponding attributes for all instances of test data and prototype vectors by means of IBA similarity measure:

$$S_{IBA}\left(x, y\right) = Agg_{k=1}^{m}\left(1 - x\left(a_k\right) - y\left(a_k\right) + 2 * \min\left(x\left(a_k\right), y\left(a_k\right)\right)\right), \quad x, y \in \left[0,1\right]^m \tag{6}$$

In the next step, based on maximum similarity between test instances and prototypes, test instances are assigned to appropriate class. Formally, if test instance $x \in C_n$ then:

$$S_{IBA}\left(x, p_n\right) = \max_{i=1...n} S_{IBA}\left(x, p_i\right) \tag{7}$$

In the algorithmic form LC classifier is as follows:

```
Algorithm: Logical classification
      input: trainData and testData with i=1…n classes,
           size of testData d,
      output: class
begin
   scale trainData between [0,1]
   scale testData between [0,1]

   for i=1 to n do
     p[i]=computePrototype(trainData[i])
     similarity[i]=S_IBA^(j=1…d)(p[i],testData[j])
   end

   class=argmax_i similarity[i]
end
```

In this experiment we used different logical aggregation functions that are further presented:

- LC_1 classifier uses simple average as a special case of pseudo LA operator:

$$S_{IBA}(x,y)=\sum_{k=1}^{m}\left(1-x(a_k)-y(a_k)+2*\min\left(x(a_k),y(a_k)\right)\right) \tag{8}$$

- LC_2 classifier aims to include interaction in the data and uses common product for operator of generalized product in LA aggregation between all attributes:

$$S_{IBA}(x,y)=\prod_{k=1}^{m}\left(1-x(a_k)-y(a_k)+2*\min\left(x(a_k),y(a_k)\right)\right) \tag{9}$$

- LC_3 classifier is similar to LC_2 but attempts to treat similar nature in the data:

$$S_{IBA}(x,y)=S_{IBA}(x,y)_{1,2}\cdot\prod_{k=3}^{m}\left(1-x(a_k)-y(a_k)+2*\min\left(x(a_k),y(a_k)\right)\right) \tag{10}$$

where $S_{IBA}(x,y)_{1,2}$ is used to calculate similarities only for the first two (highly correlated) attributes in our data set. Minimum is applied for the generalized product in LA because these attributes are similar by nature.

$$S_{IBA}(x,y)_{1,2}=\min\left(1-x(a_1)-y(a_1)+2*\min\left(x(a_1),y(a_1)\right),1-x(a_2)-y(a_2)+2*\min\left(x(a_2),y(a_2)\right)\right) \tag{11}$$

4. EXPERIMENT AND RESULTS

4.1. Data and experiment

In the experiment we used a database provided from “CUBE Risk Management Solutions” a company that operates in the field of business information and credit risk services. The database contains financial data on 1020 middle-sized companies from Serbia. The companies are selected according to Accounting and Auditing Law ("Off. Gazette of RS", no. 46/2006, 111/2009 and 99/2011) which treats classification of legal entities in Serbia. The financial dataset include figures from annual financial statements balance sheet, income statement and statistical annex for 2014 and 2015. There are 984 companies which are active and 36 in the process of bankruptcy or liquidation. For each company we calculated five financial ratios used in Altman Z score model described in the second chapter (Eq. 1). The financial ratios for year 2014 are used as inputs in logical classification model. The predicted output is company status in year 2015: active or bankrupt.

In the experiment we compared Altman Z score as benchmark model with three LC models described in previous section: LC_1 (Eq. 8), LC_2 (Eq. 9) and LC_3 (Eq. 10, Eq. 11). The idea was to use the same ratios from linear Altman model as inputs in LC classifiers to test whether we can improve classification performance. The experiment was conducted on a non-balanced dataset.

To enable fair comparison of the models, it is important to consider various factors in the dataset that may have an effect on classification performance. In case of Altman's Z score model these factors are year of bankruptcy, company size, company age, industry and country of origin (Altman et al., 2014).

For his model Altman suggested utilizing data as near to the present as possible in order to develop more accurate bankruptcy prediction model. Also, significant impact on the performance of the model has size of companies. To eliminate this weakness, our analysis is focused only on middle-sized companies. Altman noticed that the age of the company may have an influence on bankruptcy risk. In particular, very young firms typically show a very high risk of failure. However we used middle-sized companies that are considered to be long-term and stable. Another important factor is country of origin. Different economic and culture factors, financial markets, legislation and accounting practices affect the financial behavior of companies. The Altman model is not adapted for Serbian economy.

In order to test the influence of data normalization to LC performance we scaled data on [0,1] interval using three standard normalization functions:

$$N_1=\frac{x+\min(X)}{\max\left(X+\min(X)\right)} \tag{12}$$

$$N_2=\frac{x}{\max(X)} \tag{13}$$

$$N_3 = \frac{x - \min(X)}{\max(X) - \min(X)} \qquad (14)$$

where $x \in X$, and $\min(X)$ and $\max(X)$ represent minimum and maximum value of X , respectively.

In this paper, cross validation is used to evaluate the performance of the LC models. The cross validation was run 100 times. Each time data set was randomly portioned in two parts for training and testing 50-50 or 70-30. Slightly better results were achieved when we used 70% of data for training and 30% for testing.

4.2. Results and analysis

In this section we compare classification performance for Altman Z score and three LC models. For the dataset of Serbian middle-sized companies, Altman Z score achieved **73.24%** accuracy. In the following table (Table 1) classification results for LC models are presented, where all three LC models shows better accuracy than Altman benchmark model.

Table 1: Bankruptcy prediction results for LC classifiers

Normalization	LC_1		LC_2		LC_3	
	Accuracy	**Variance**	**Accuracy**	**Variance**	**Accuracy**	**Variance**
N_1	0.6999	0.0017	0.7126	0.0013	0.7239	0.0015
N_2	**0.7436**	0.0009	**0.7567**	0.0010	**0.7800**	0.0009
N_3	0.6959	0.0018	0.7002	0.0021	0.7226	0.0016

In Table 1 we can see that the best classification accuracy is achieved with normalization function N_2 (Eq. 13) for all three LC classifiers. Variance is very low for all three models and all normalization functions so we can conclude that LC models give quite stable results.

As expected the first LC1 model has the lowest accuracy, because it employs simple and linear classifier function (Eq. 8). LC2 model is developed to capture interaction in data by means of nonlinear aggregation function (Eq. 9). And the last LC3 model (Eq. 10, Eq. 11) aims to additionally include the same or similar nature of data, so that it shows the best accuracy.

The classification performance is also presented by confusion matrices in Figure 2 below. It is important to note that the results in confusion matrices are average values for 100 iterations. Confusion matrix consists of the accuracy (green fields) and error rates (red fields):

- TP true positive defines active (non-bankrupt) firms that are classified as active.
- TN true negative indicates bankrupt firms that are classified as bankrupt.
- FP false positive indicates bankrupt firms that are classified as active.
- FN false negative defines active firms that are classified as bankrupt.

LC1 classifier

Actual \ Predicted	**Active**	**Bankrupt**
Active	220.96 (71.97%)	75.04 (24.44%)
Bankrupt	3.66 (1.19%)	7.34 (2.39%)

LC2 classifier

Actual \ Predicted	**Active**	**Bankrupt**
Active	225.24 (73.37%)	70.76 (23.25%)
Bankrupt	3.94 (1.28%)	7.06 (2.29%)

LC3 classifier

Actual \ Predicted	**Active**	**Bankrupt**
Active	232.53 (75.74%)	63.47 (20.67%)
Bankrupt	4.08 (1.33%)	6.92 (2.25%)

Altman Z score

Actual \ Predicted	**Active**	**Bankrupt**
Active	719 (70.49%)	265 (25.98%)
Bankrupt	8 (0.78%)	28 (2.75%)

Figure 2: Confusion matrices for LC models and Altman Z score

When type I and type II errors are considered, LC classifiers have better error rates than Altman Z score for active firms that are classified as bankrupt but they provide somewhat worse error rates for bankrupt firms classified as active. Even though that difference is rather small, it can be indicator for further analysis and LC models improvement.

5. CONCLUSION

The aim of this paper is to propose new logical classifiers and evaluate their classification performance in bankruptcy prediction. Three LC models are tested whether they are able to improve Altman Z score as a benchmark model for bankruptcy prediction. In case of Serbian middle-sized companies LC classifiers outperformed Altman Z score. Even though LC method shows more accurate results, Altman model behaves slightly better for grey-zone companies, where it is important to reduce number of bankrupt firms identified as active. However there are still room left for improvement of LC model.

For further research, we suggest adding another class in the output that will identify these particular companies in pre-default state that are critical to fail in near future but still active. We also intend to consider additional variables that may help boost the classification performance to a higher level. One of the ideas is to analyze LC models on small-sized companies and try to identify other appropriate logical classifiers.

In a nutshell, the analysis in this study shows that LC model is reasonable good classifier for predicting company default. It is a rather stable classification model with low variance. In particular, LC method makes possible to understand the reasoning behind classifier function, which is important for decision making that is based on the classification models.

REFERENCES

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4), 589-609. doi: 10.2307/2978933

Altman, E. I. (2000). Predicting financial distress of companies: revisiting the Z-score and ZETA models. *The journal of finance,* 9-12. doi: 10.4337/9780857936097.00027

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2014). Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management & Accounting,* doi: 10.1111/jifm.12053

Accounting and Auditing Law (2016). Retrieved from http://www.nbs.rs/.

Chaudhuri, A. (2013). Bankruptcy Prediction Using Bayesian, Hazard, Mixed Logit and Rough Bayesian Models: A Comparative Analysis. *Canadian Center of Science and Education*, 6(2), 1913-8989. doi:10.5539/cis.v6n2p103.

Du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing*, 73(10), 2047-2060. doi: 10.1016/j.neucom.2009.11.034

Grice, J. S., & Ingram, R. W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model. *Journal of Business Research*, 54(1), 53-61. doi: 10.1016/S0148-2963(00)00126-0

Jackson, H.G.R., Wood A.P. (2013). The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review* 45(3), 0890-8389. doi: 10.1016/j.bar.2013.06.009.

Kasgari, A. A., Divsalar, M., Javid, M. R., & Ebrahimian, S. J. (2013). Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses. *Neural Computing and Applications*, 23(3-4), 927-936. doi: 10.1007/s00521-012-1017-z

Korol, T., & Korodi, A. (2011). An evaluation of effectiveness of fuzzy logic model in predicting the business bankruptcy. *Romanian Journal of Economic Forecasting*, 3(1), 92-107. Retrieved from http://www.ipe.ro/rjef/rjef3_11/rjef3_2011p92-107.pdf

Kriegel, H. P., & Schubert, M. (2006). Advanced Prototype Machines: Exploring Prototypes for Classification. *In SDM*, 176-187. doi: 10.1007/s10618-007-0067-9

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European journal of operational research*, 180(1), 1-28. doi: 10.1016/j.ejor.2006.08.043

Milošević, P., Nešić, I., Poledica, A., Radojević, D. G., & Petrović, B. (2013). Models for ranking students: selecting applicants for a master of science studies. *In Soft computing applications* 93-103. doi: 10.1007/978-3-642-33941-7-11

Milošević, P., Petrović, B., Radojević, D., & Kovačević, D. (2014). A software tool for uncertainty modeling using Interpolative Boolean algebra. *Knowledge-Based Systems*, 62, 1-10. doi: 10.1016/j.knosys.2014.01.019

Min, J. H., & Jeong, C. (2009). A binary classification method for bankruptcy prediction. *Expert Systems with Applications*, 36(3), 5256-5263. doi: 10.1016/j.eswa.2008.06.073

Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, 52(2), 464-473. doi: 10.1016/j.dss.2011.10.007

Patouchas, A. (2010). The Art of Bankruptcy Prediction: a Review of Developed Methods. Aalto University School of Economics Department of Accounting and Finance. Helsinki, Finland.

Poledica, A., Bogojević-Arsić, V., & Petrović, B., (2010). Logical aggregation as similarity measure in case-based reasoning, *In Computational Intelligence Foundation and Application: Proceedings of the 9th International FLINS Conference, World Scientific Publishing Co*., 585-590. doi: 10.1142/9789814324700_0088

Poledica, A., Milošević, P., Dragović, I., Radojević, D., & Petrović, B. (2013). A consensus model in group decision making based on interpolative boolean algebra. *In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13). Atlantis Press*, 32, 648-654 doi: 10.2991/eusflat.2013.98

Poledica, A., Milošević, P., Dragović, I., Petrović, B., & Radojević, D. (2015). Modeling consensus using logic-based similarity measures. *Soft Computing*, 19(11), 3209-3219. doi: 10.1007/s00500-014-1476-5

Radojević, D. (2000). (0,1)-valued logic: A natural generalization of Boolean logic. *Yugoslav Journal of Operations Research*, 10(2), 185-216. doi: 10.1016/j.eswa.2008.06.050

Radojević, D. (2008). Logical Aggregation Based on Interpolative Boolean Algebra. Mathware & Soft Computing,15(1),125-141.Retrieved from http://upcommons.upc.edu/bitstream/handle/2099/13098/22.pdf

Y., Gaunt C., Gray S. (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting and Economics*, 6(1),1815-5669. doi:10.1016/j.jcae.2010.04.002.

Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems*, 41, 16-25. doi: 10.1016/j.knosys.2012.12.007

BUILDING A DATA-ENRICHED DEX-BASED DECISION SUPPORT SYSTEM FOR EARLY WARNING ON INCREASED RISK OF SKIING INJURY OCCURRENCE – FIRST RESULTS

Petar Marković*[1], Boris Delibašić[1], Milija Suknović[1]

[1]University of Belgrade, Faculty of organizational sciences, Serbia
*Corresponding author, e-mail: markovic26110mo@fon.bg.ac.rs

Abstract: *In this paper we present first results of building a decision support system for early warning on increased risk of skiing injury occurrence. We utilize the power of recently proposed DEX-based data-enriched modeling technique, showing that, on average, there is an evident supremacy of such models compared to the ones obtained using traditional data-mining and machine learning techniques in terms of prediction accuracy, consistency and interpretability. We argue the reason for such outcome lies in fact that traditional machine learning models are unable to explain the high variability in data, as it is known that skiing injuries are very rare events (only a 2.5 per mille of skier visits gets injured) and their spatio-temporal distribution in ski resort is very variable due to difference in skiing regions nature during different time periods (different popularity and difficulty of the slopes, weather conditions, skiing style of skiers, etc.) which also leads to high class imbalance problem. Thus, machine learning algorithms do not have enough data-based evidence to create strong rules for making a stable decision. On the other hand, traditional decision support systems, which are strictly expert-made, do not incorporate data-based knowledge within themselves. However, ski resorts safety experts are able to handle uncertainty and based on their experience recognize the patterns which lead to increased risk of injury, and express them in the form of a decision rule. Therefore, we effectively combine the best from both approaches to overcome this problem. For the purpose of this paper we used data from largest Serbian ski resort at Mt. Kopaonik from representative season 2010. Dataset was obtained through integration of three different data sources: weather data, ski lifts gates entrance data and injury records from the rescue service. The results are 22 location-specific models, plus 2 global models used for benchmark purposes. We compare the experimental results obtained with 7 most commonly machine learning algorithms used, tested on the same data with 10-fold cross-validation applied.*

Keywords*: DEX, decision support system, increased risk of ski injury, early warning, machine learning, Mt. Kopaonik ski resort*

1. INTRODUCTION

Skiing injuries are rare events. On average, only 2.5‰ of ski visits get injured (Conn et al, 2003). However, consequences of these events have strong impact both to skiers and ski resorts, as injuries are usually very severe which can generate high costs of medication (Warme et al, 1995). Report of Vanat (2014) states that the number of skiers around the world is around 400 million, which means that around 10,000 people gets injured, which makes skiing accidents prevention in ski resorts the topic of the utter importance on the global level (Nerin et al, 2007).

Until present, there were only a few papers which address this problem (Bohanec and Delibašić, 2015), (Dalipi et al, 2015), (Delibašić and Obradović, 2012).

This paper is an extension of recent paper of Bohanec and Delibašić (2015) which proposes a decision support system for skiing injury prediction using data-based knowledge and DEX decision support methodology. The methodology used in this paper is based on the same idea, but with major improvements on the modeling with respect to time and location. In original paper, the ski resort was observed as a whole, and the predictions were made on the daily level, which is in contrast to this paper, where the ski locations are modeled separately, and the time frame was a half of a skiing day.

Mt. Kopaonik is the largest Serbian ski resort. In the representative skiing season 2010, resort consisted of 14 skiing locations equipped with RFID ski gate scanners on lifts (SUN – Suncana dolina, MAR – Marine vode, DUB - Duboka, GOB - Gobelja, GVO - Gvozdac, JAR - Jaram, JEZ - Jezero, KAR - Karaman, KGB – Karaman Greben, KNE – Knezeve bare, MAK – Mali karaman , MAS - Masinac, PAN – Pancicev vrh, KRS - Krst).

Observing the integrated data from three different sources (ski gate entrance data, rescue service injury records and weather data) 11 locations were found eligible for further analysis, since KRS did not have any injuries recorded, whilst JAR had a 100% rate of injuries, and GVO had too few readings (Figures 1 and 2).

On average, on Mt. Kopaonik, injury rate is 1.5‰, which is in accordance with literature. Figure 1 shows that locations DUB, KGB, MAK, MAR are especially dangerous. On the other hand, figure 2 shows that the part of the day is an important factor for predicting injuries, as not all tracks are equally dangerous in the morning and in the afternoon. For example, GVO is much safer in the afternoon, as its percent of injuries drop from 78% in the morning to 57% in the afternoon.

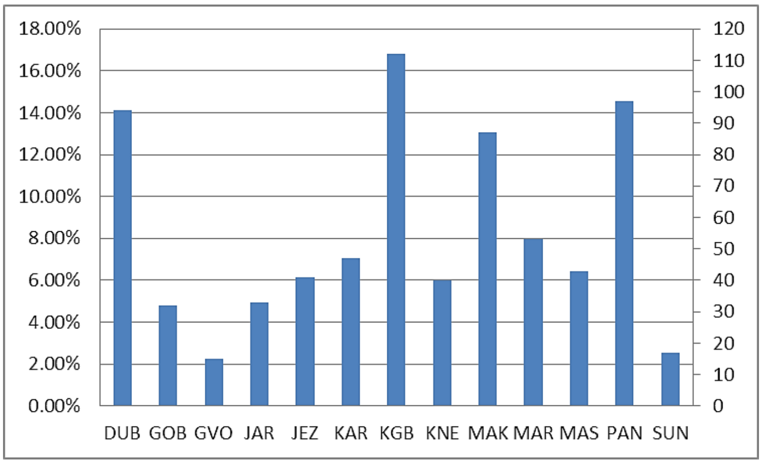

Figure 1: Distribution of skier days with at least one ski injury, per location (skiing region)

Row Labels	before noon 0	1	before noon To	afternoon 0	1	afternoon Total
SUN	33.33%	66.67%	50.00%	58.33%	41.67%	50.00%
MAR	60.61%	39.39%	50.00%	30.30%	69.70%	50.00%
DUB	71.23%	28.77%	49.66%	72.97%	27.03%	50.34%
GOB	78.26%	21.74%	54.76%	78.95%	21.05%	45.24%
GVO	21.43%	78.57%	66.67%	42.86%	57.14%	33.33%
JAR	0.00%	100.00%	68.57%	0.00%	100.00%	31.43%
JEZ	69.23%	30.77%	51.32%	45.95%	54.05%	48.68%
KAR	48.28%	51.72%	48.33%	16.13%	83.87%	51.67%
KGB	73.81%	26.19%	52.17%	74.03%	25.97%	47.83%
KNE	15.63%	84.38%	62.75%	21.05%	78.95%	37.25%
MAK	62.50%	37.50%	50.91%	62.96%	37.04%	49.09%
MAS	68.09%	31.91%	54.65%	48.72%	51.28%	45.35%
PAN	53.85%	46.15%	48.51%	65.22%	34.78%	51.49%
Grand Total	**57.82%**	**42.18%**	**52.42%**	**56.02%**	**43.98%**	**47.58%**

Figure 2: Relative distribution of skiing days with below-average (class label 0) and above-average (class label 1) number of injuries, per location and day period

2. METHODOLOGY

The core of methodology is based on previous research of Bohanec and Delibašić (2015) where data-enriched DEX-based decision support system was created for the ski resort as a whole, on the daily level. However, as our approach presumes independent location-time modeling, we adjusted the methodology so that attribute selection for each model is done separately, whilst the following steps are repeated iteratively for each of the models built. In short, the benefits of such modeling approach come from the fact that DEX models are transparent and consistent, whilst the knowledge which cannot be found in the data is complemented with the expert knowledge.

DEX models (Bohanec et al, 2013) represent a class of qualitative multi-criteria decision-making models. These models have hierarchical structure of attributes, where the attributes on the higher level of hierarchy are decomposed by set of attributes on the lower level. The attributes are defined using qualitative scales, ordered from the worst to the best values. Attributes on the lower level are aggregated to the higher level attributes through utility functions (tables of elementary decision rules) where each combination of values of input (lower level) attributes are mapped to the corresponding value of the output (upper level) attribute. Consistency in model is achieved through presumption of utility function monotonicity, which means that for two combinations of non-decreasing input attributes' combinations, the consecutive output attribute value could not be lower than the previous. Thus, DEX model can be used as a classifier, as it actually represents a hierarchically ordered set of IF-THEN rules. For in-depth explanation on DEX models please consult the original reference.

The first step in methodology is discretization and attributes selection. This step is done in Orange data-mining software (Demšar, 2013), using interaction graphs. Interaction graphs (Jakulin, 2015) show which input attributes observed together, rather than separately, bring additional information for class determination. Therefore, the attributes with positive interaction values are good candidates for grouping, whilst ones with negative values should be disregarded from the model. Also, the attributes that do not have any information gain are excluded from the model. For the discretization sub-step, according to the previous paper of Bohanec and Delibašić (2015), 3-bins (low, medium, high) equal-frequency discretization was used, as it initially gave better results than the 5-bins version.

Second step is DEX modeling phase: attribute hierarchy creation, attribute scales determination based on relative frequencies observed in data, and definition of decision rules for utility functions. It is presumed that total ski injury risk can be decomposed into risk from crowding, risk from skiing style and risk from weather conditions. Therefore, in each model, the target attribute Total risk (on the first level of hierarchy) is decomposed into previously mentioned subgroups of attributes, which consist of belonging attributes selected in the previous step. When attribute hierarchy is created, the attribute scales for the elementary attributes (attributes on the lowest level of hierarchy) are defined by observing relative frequencies of class attribute in the data; whilst for the aggregate attributes the scales are always defined manually as H-M-L (as the lowest values of risk are the best in this case). The procedure is the same when determining decision rules for the aggregate attributes. Utility functions of aggregate attributes consisting of elementary attributes are defined by reading relative frequencies distributions of class attribute from data, whilst missing rules are defined by experts. On the other hand, utility functions of aggregate attributes consisting of other aggregate attributes are completely defined by experts (e.g. high risk of skiing style & high risk of crowding & high risk of weather conditions lead to high level of total risk of injury). In-detail explanation of attributes used is given in the following section.

Third and final step is Machine learning modeling phase where the machine learning models were created for the same groups of attributes used in the step one; 10-fold cross-validation was used. Classification tree, Logistic regression, Majority, CN2 rules, k-NN, SVM, Naïve Bayes classifiers were used, as they represent the most commonly used algorithms for this purpose. Due to space restrictions we skip detailed parameters settings.

3. EXPERIMENTAL SETUP AND MODELING

Original paper (Bohanec and Delibašić, 2015) presumed that all locations behave the same way, or in other words, it observed a ski resort in whole, and did not presume that different parts of the day have significant impact on levels of ski injury risk. However, observing the data from the representative skiing season (figures 1 and 2), we noticed that different locations have different injury rates at different period of day, thus we modeled them independently. This presumption was also proved when interaction graphs from the step 1 of methodology were created, which showed that not all attributes were important for predicting the injury on specific locations. Therefore, two types of models are created: location-specific models (represented as DEX in results), and global models (represented as DEX_all in results).

Due to space restrictions, we provide only a short preview of attributes (and their corresponding groups) used in models, whilst the true hierarchical structure between them is not presented. Figure 3 shows attributes used for the before noon models, whilst figure 4 shows attributes used for the afternoon models. There are three major groups of attributes: crowding-related, skiing-style-related and weather-related attributes.

Detailed explanation of each attribute is provided in the Table 1.

	Before noon	JEZ	MAS	SUN	PAN	DUB	KGB	KNE	MAK	GVO	MAR	KAR	GOB	JAR	DEX_all	SUM (w/o DEX_all)
crowding	noSkiers									/				/		8
	noRuns									/				/		9
	utilizationNorm									/				/		9
skiing	avgElevation									/				/		8
	avgTimeRun									/				/		4
	avgTimeSki									/				/		6
	minElevation									/				/		3
	minTimeRun									/				/		1
	minTimeSki									/				/		5
	maxElevation									/				/		3
	maxTimeRun									/				/		1
	maxTimeSki									/				/		5
	avgSpeedRun									/				/		8
weather	tempAvg									/				/		9
	tempMin									/				/		5
	tempMax									/				/		8
	cloudiness									/				/		5
	windSpeed									/				/		9
	snowNewCM									/				/		9
	geoAspect									/				/		0
	geoAspect2									/				/		0
	geoAspectNonBin									/				/		0
	geoAspectNonBin2									/				/		0
	SUM	12	13	9	11	11	5	11	9	/	12	11	11	/	19	

Figure 3: Attributes used for the morning models

On average, 11 attributes were needed for modeling. From figures 3 and 4 it is observable that in the afternoon, weather-related attributes and skiing-style-related attributes are more significant, compared to the before noon, where crowding-related attributes are more important. Both in the afternoon, and before noon, there are locations that can be modeled quite well with small number of attributes (such as KGB and DUB).

The number of attributes used potentially could be even smaller, if the data was more precise, as from the previous figures it is observable that some of the attributes in the same group are significantly less frequently used than the others, and probably are highly correlated to the most frequently used ones (e.g. minTimeRun, maxTimeRun, etc.).

	afternoon	JEZ	MAS	SUN	PAN	DUB	KGB	KNE	MAK	GVO	MAR	KAR	GOB	JAR	dex_all	sum (w/o dex_all)
crowding	noSkiers									/				/		8
	noRuns									/				/		5
	utilizationNorm									/				/		9
skiing	avgElevation									/				/		7
	avgTimeRun									/				/		5
	avgTimeSki									/				/		2
	minElevation									/				/		0
	minTimeRun									/				/		8
	minTimeSki									/				/		9
	maxElevation									/				/		4
	maxTimeRun									/				/		4
	maxTimeSki									/				/		5
	avgSpeedRun									/				/		7
weather	tempAvg									/				/		9
	tempMin									/				/		9
	tempMax									/				/		8
	cloudiness									/				/		9
	windSpeed									/				/		9
	snowNewCM									/				/		7
	geoAspect									/				/		0
	geoAspect2									/				/		0
	geoAspectNonBin									/				/		0
	geoAspectNonBin2									/				/		0
	sum	12	13	11	11	8	15	12	11	/	9	12	10	/	19	

Figure 4: Attributes used for the afternoon models

In table 1 we provide the detailed information on all attributes used. The attributes in italic are used for the modeling purpose.

Table 1: Explanation of attributes used

ID	Attribute group	Attribute	Description
1.	geo-spatial	date	
2.		region	
3.		dayPeriod	
4.	*crowding*	*noSkiers*	*Number of skiers observed*
5.		*noRuns*	*Number of ski lift transportations observed*
6.		*utilizationNorm*	*Normalized ski lift capacity utilization*
7.	*skiing*	*avgElevation*	*Average height difference during skiing*
8.		*minElevation*	*Minimum height difference during skiing*
9.		*maxElevation*	*Maximum height difference during skiing*
10.		*avgTimeRun*	*Average time on slope (including lift transportation time)*
11.		*minTimeRun*	*Minimum time on slope (including lift transportation time)*
12.		*maxTimeRun*	*Maximum time on slope (including lift transportation time)*
13.		*avgTimeSki*	*Average time of skiing (without lift transportation time)*
14.		*minTimeSki*	*Minimum time of skiing (without lift transportation time)*
15.		*maxTimeSki*	*Maximum time of skiing (without lift transportation time)*
16.		*avgSpeedRun*	*Average skiing speed*
17.	*weather*	*tempAvg*	*Average daily temperature in ski resort*
18.		*tempMin*	*Minimum daily temperature in ski resort*
19.		*tempMax*	*Maximum daily temperature in ski resort*
20.		*cloudiness*	*Daily cloudiness level*
21.		*windspeed*	*Wind speed*
22.		*snowNewCM*	*Height of new snow in CM*
23.		*geoAspect*	*Impact of geographical aspect of slope, three categories*
24.		*geoAspect2*	*Impact of geographical aspect of slope (alternative), three categories*
25.		*geoAspectNonBin*	*Impact of geographical aspect of slope, two categories*
26.		*geoAspectNonBin2*	*Impact of geographical aspect of slope, two categories*
27.	*injury*	injury	Absolute number of ski injuries per ski lift transportation
28.	*injury*	*manyInjuries*	*If there is an above-average level of injuries takes value 1, otherwise 0*

4. RESULTS

The classification accuracy was used as a main criterion of a model quality. The overall results are presented with two violin-plots in figures 5 and 6, respectively. Mean is marked with red dot, whilst the median is represented with white dot.

Figure 5 shows distribution of classification accuracy for the morning models. It is noticeable that mean values of location-specific model are much higher than other algorithms (Classification tree, Logistic regression, Majority, CN2 rules, k-NN, SVM, Naïve Bayes) and a-priori class probability. However, due to skewness of distributions it seems that median values are not so high, but still, for some of the locations such as KGB and DUB where it significantly dominated other models, and on the locations GOB and SUN was equal with the first-ranked algorithm. On other locations, on average, it was second-ranked with not very high difference in values of accuracy.

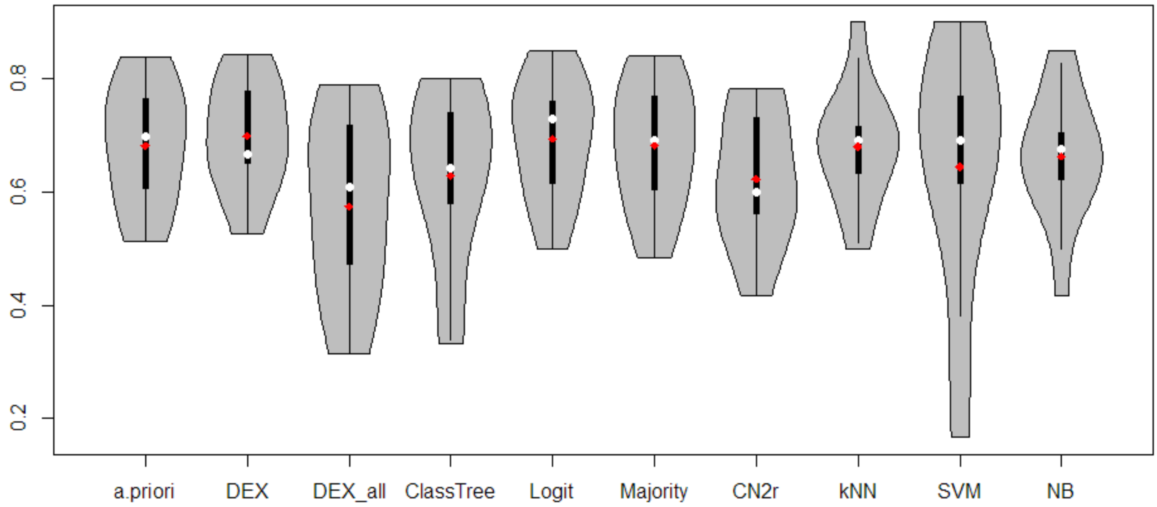

Figure 5: Violin plot of classification accuracy for the morning models

Figure 6 shows distribution of classification accuracy for the afternoon models. Here, the situation is even clearer than in the first part of the skiing day. We see that location-specific DEX model is superior to all others. Especially good results are achieved on locations SUN, MAR, MAK, MAS, PAN where our DEX model strongly outperformed other models.

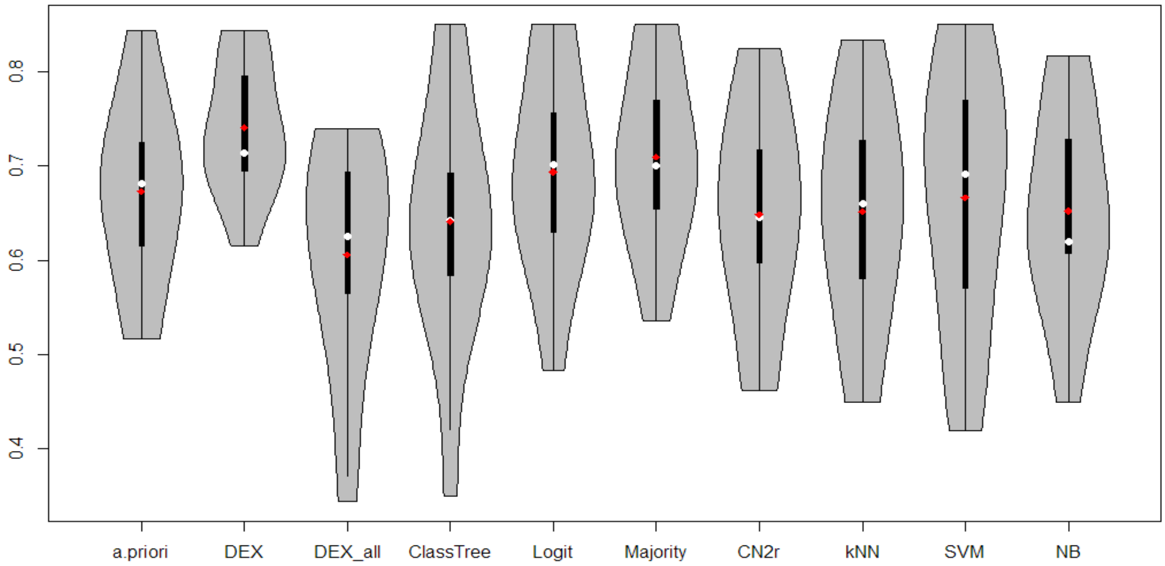

Figure 6: Violin plot of classification accuracy for the afternoon models

We conlude that our approach is major improvement to the previous research of Bohanec and Delibašić (2015) as DEX_all models showed much lower accuracy rates, both compared to location-specific DEX models, and other ML algorithms.

5. CONCLUSION

In this paper we created a near-real-time decision support system for early warning on skiing injury accident occurrence using recently proposed data-enriched DEX-based modeling approach.

The first results show that our approach, on average, reaches or outperforms traditional machine learning algorithms in terms of classification accuracy, whilst models obtained using proposed approach are consistent and easy to understand.

Further research should focus on adapting the methodology to fit the hourly time bins, which should finally lead to a real-time decision support system.

REFERENCES

Bohanec, M., Bratko, V., Rajkovič, I., Zupan, B., and Žnidaršić, M. "DEX methodology: three decades of qualitative multi-attribute modelling." *Informatica* 37 (2013): 49-54.

Bohanec, M. and Delibašić, B. "Data-mining and Expert Models for Predicting Injury Risk in Ski Resorts." ICDSST*-15, International Conference on Decision Support System Technology.* Belgrade, 2015.

Conn, J.M., J.L Annest, and J. Gilchrist. "Sports and recreaion related injury eposides in the US population 1997-99." *Injury Prevention* 9(2) (2003): 117-123.

Dalipi, Fisnik, and Sule Yildirim-Yayilgan. "The Impact of Environmental Factors to Skiing Injuries: Bayesian regularization Neural Network Model for Predicting Skiing Injuries." The 6th IEEE International Conference on Computing Communications and Networking Technologies. Dallas, USA, 2015.

Delibašić, B., and Obradović, Z. "Towards a DGSS Prototype for Early Warning for Ski Injuries." *Data Engineering Workshops (ICDEW), 2012 IEEE 28th international Conference on Data Engeneering Workshops.* Arlington, VA: IEEE, 2012. 68-72.

Demšar, J., and et al. "Orange: Data mining toolbox in Python." *Journal of Machine Learning research* 14 (2013): 2349-2353.

Jakulin, A. *Attribute interactions in Machine learning.* 2005. stat.columbia.edu/~jakulin/Int/ (accessed 9 17, 2015).

Nerin, M.A., P. Maza, P. Bada, and P. Garijo. "Safe Mountains: Ans Initiative of Prevention of Mountain Accidents in Central Pyrenees (Aragon, Spain)." *Wilderness & Environmental Medicine* 18 (2007): 235-250.

Vanat, L. *International Report on Snow & Mountain Tourism: Overview of the key industry figures for ski resorts.* 2014. http://www.isiaski.org/download/20140517_ISIA_Vuokatti_1b_presentation_vanat.pdf (last access date: 3.9.2015).

Warme, W.J., Jr.J.A. Feagin, and P. King. "Ski injury statistics, 1982 to 1993, Jackson Holes Ski Resort." *The American Journal of Sports Medicine* 23 (5) (1995): 597-600

EUROPE 2020 STRATEGY – A MULTIVARIATE APPROACH

Nemanja Milenković*[1], Aleksandar Đoković[1], Aleksandra Vukmirović[2]
[1]Faculty of Ogranizational Sciences, University of Belgrade
[2]Belgrade Business School, Higher education institution for applied studies
*Corresponding author, e-mail: milenkovic@fon.bg.ac.rs

Abstract: *In 2010 the European Union (EU) and Member States started a strategy for sustainable growth for the next ten years – the strategy "Europe 2020". Strategy is focused on three mutually reinforcing priorities: smart growth, sustainable growth and inclusive growth. It establishes five major goals for the EU: employment, research and development, climate change and sustainable energy, education and combating poverty and social exclusion. The European Commission has identified thirteen indicators divided into five groups based on which it will measure the success of the Europe 2020 Strategy. The I-distance method is applied on the above indicators. The aim of this paper has been to present one synthesized indicator that is able to quantitatively demonstrate any country's development. The I-distance method is thoroughly explained and has been applied to 28 EU countries. Crucial indicators for ranking are also elaborated.*

Keywords*: strategy Europe 2020, European Union, I-distance method, indicators, ranking*

1. INTRODUCTION

The global economic crisis has wiped out years of economic and social progress and pointed to the structural weaknesses of European economy. In 2010 the European Union (EU) and Member States started a strategy for sustainable growth for the next ten years – the strategy "Europe 2020". This strategy is a continuation of the Lisbon Strategy announced at the beginning of this century, which was aimed at improving conditions for sustainable economic development (Balcerzak, 2015). In contrast to the Lisbon Strategy, Europe 2020 Strategy does not any more insist on global competitiveness, but on the creation of a "new economy" to solve primarily European economic problems (Hoedl, 2011). It is dedicated to short-term challenges associated with the crisis, but also the need for structural reforms by means of measures to encourage the growth needed to prepare the EU for the future. The implementation of the strategy Europe 2020 takes place in a rather difficult and turbulent period marked by the particularly the negative impact of the global financial crisis, which significantly affected the further development of the project to build economic and monetary union in the European integration space (Terem et al., 2015).

The aim of the Europe 2020 Strategy is not only to overcome the crisis, but also to address the shortcomings of the existing model of growth and to create conditions for smart and sustainable growth and development. Europe 2020 Strategy is focused on three mutually reinforcing priorities (European Commission, 2010): smart growth (developing an economy based on knowledge and innovation), sustainable growth (promoting a more resource efficient, greener and more competitive economy) and inclusive growth (fostering a high-employment economy delivering social and territorial cohesion).

It establishes five major goals for the EU by the end of 2020. They include: employment, research and development, climate change and sustainable energy, education and combating poverty and social exclusion. The objectives of the strategy are supported by seven flagship initiatives (Innovation Union, Youth on the Move, A Digital Agenda for Europe, Resource efficient Europe, Industrial policy for the globalization era, Agenda for new skills and jobs, European platform against poverty). Together they provide a framework to enable the EU and regional centers to jointly strengthen efforts in areas that support the priorities of the Europe 2020 Strategy (Vuković 2011; Jakab and Alderslade, 2015).

The European Commission has identified thirteen indicators divided into the following five groups on the basis of which they will measure the achievements of the five main objectives:

1. Employment: (I) Employment rate (age group 20-64);
2. Research and development: (II) Gross domestic expenditure on R&D (% of GDP);
3. Climate change and energy: (III) Greenhouse gas emissions (in %, base year 1990), (IV) Share of renewable energy in gross final energy consumption (in %), (V) Primary energy consumption (in %, base year 2005), (VI) Final energy consumption (in %, base year 2005), (VII) Greenhouse gas emissions in non-ETS sectors (base year - 100 million tonnes CO_2 equivalent);

4. Education: (VIII) Early leavers from education and training (% of the population aged 18-24), (IX) Tertiary educational attainment (age group 30-34);
5. Poverty and social exclusion: (X) People at risk of poverty or social exclusion (% of total population), (XI) People living in households with very low work intensity (% of total population aged less than 60), (XII) People at risk of poverty after social transfers (% of total population), (XIII) Severely materially deprived people (% of total population).

The advantage of this policy document resides in the fact that it has well-established indicators that enable measuring achievement, and thereby the possible success of the respective European policies (Bere et al., 2015). The aforementioned indicators are capable of expressing the diversities of countries and their performances.

The aim of this paper is to explore the current situation in the EU using these thirteen indicators. To this end we have used multivariate I-distance methodology to reduce the initial set of indicators and to extract and identify the most important indicators that will contribute to the achievement of the priorities and the main objectives of Europe 2020 Strategy.

2. THE I-DISTANCE METHOD

In order to create a synthesized development indicator, selected indicators are incorporated into the analysis through use of the statistical I-distance method. The analyses carried out using the statistical I–distance method are numerous. I–distance was applied in previous researches in order to evaluate the academic ranking of the world's universities (Jeremic et al., 2011; Jovanovic et al., 2012). It was also used for measuring European countries' health systems (Jeremic et al., 2012), sustainable development (Radojicic et al., 2012), public health (Seke et al., 2013) and ICT Development (Dobrota et al., 2012).

I-distance is a metric distance in an *n*-dimensional space. Ivanovic had originally devised this method to rank countries according to their level of development based on several indicators. Many socio-economic development indicators were considered and the problem was how to use all of them in order to calculate a single synthetic indicator, which would thereafter represent the rank (Ivanovic, 1973).

For a selected set of variables $X^T = (X_1, X_2, ..., X_k)$ chosen to characterize the entities, the I-distance between the two entities $e_r = (x_{1r}, x_{2r}, ..., x_{kr})$ and $e_s = (x_{1s}, x_{2s}, ..., x_{ks})$ is defined as

$$D(r,s) = \sum_{i=1}^{k} \frac{|d_i(r,s)|}{\sigma_i} \prod_{j=1}^{i-1} \left(1 - r_{ji.12...j-1}\right) \tag{1}$$

where $d_i(r,s)$ is the distance between the values of variable X_i for e_r and e_s e.g. the discriminate effect $d_i(r,s) = x_{ir} - x_{is}, i \in \{1,...,k\}, \sigma_i$ the standard deviation of X_i, and $r_{ji.12...j-1}$ is a partial coefficient of the correlation between X_i and X_j, $(j < i)$ (Ivanovic, 1977).

The construction of the I-distance is iterative; it is calculated through the following steps:

- Calculate the value of the discriminate effect of the variable X_1 (the most significant variable, that which provides the largest amount of information on the phenomena that are to be ranked,
- Add the value of the discriminate effect of X_2 which is not covered by X_1,
- Add the value of the discriminate effect of X_3 which is not covered by X_1 and X_2,
- Repeat the procedure for all variables.

Occasionally, it is not possible to achieve the same sign mark for all variables in all sets. As a result, a negative correlation coefficient and a negative coefficient of a partial correlation may occur. This makes the use of the square I-distance even more desirable. The square I-distance is given as

$$D^2(r,s) = \sum_{i=1}^{k} \frac{d_i^2(r,s)}{\sigma_i^2} \prod_{j=1}^{i-1} \left(1 - r_{ji.12...j-1}^2\right) \tag{2}$$

The entity with the minimal value for each indicator or a fictive maximal or average values entity can be set up as the referent entity. The ranking of entities in the set is based on the calculated distance from the referent entity (Ivanovic and Fanchette, 1973).

By using the calculated I^2-distance, the intensity of the observed phenomena and rank entities can be observed. When a correlation coefficient of each indicator with the I^2-distance is calculated with the ranking indicators of those values, the importance of each indicator can also be examined. As the correlation coefficient is stronger, the amount of information that is provided with the observed indicator is also greater, when the $p < 0.05$ indicator is significant. Otherwise, the indicator is not important in measuring the phenomena observed. One of two reasons might explain this: either this indicator is not relevant in measuring the phenomena observed, or its discriminate effect is already contained in previous variables. Whatever the reason, the indicator must be excluded from further analysis, since, in order to select only significant indicators, it is necessary to calculate the I^2-distance and its correlation with the indicators used several times, excluding the one insignificant indicator that has the smallest correlation coefficient (Milenkovic et al., 2014). Through use of the stepwise method, one indicator is eliminated in every calculation until the results show that all the indicators used are significant, whereupon the results are obtained.

3. RESULTS

The database used in this paper includes the values of the thirteen indicators for 28 countries of the EU. We used the data for the year 2014, since it is the last year for which all indicator values for specific countries in the EU exist. To implement I-distance methodology we used the statistical package SPSS version 22. The results achieved by the square I-distance ranking method in the first calculation for evaluating development are presented in Table 1.

Table 1. The results of the I^2-distance method, I^2-distance value, and rank – first calculation

Country	I^2-distance	Rank
Sweden	84.54	1
Estonia	49.42	2
Finland	41.09	3
Luxembourg	39.51	4
Malta	37.35	5
Austria	35.05	6
Czech Republic	32.55	7
Denmark	31.80	8
Lithuania	31.29	9
Netherlands	29.86	10
Germany	27.28	11
Latvia	30.34	12
Poland	28.72	13
Slovenia	25.74	14
France	24.06	15
United Kingdom	21.85	16
Cyprus	23.31	17
Slovakia	22.42	18
Romania	22.05	19
Belgium	18.76	20
Ireland	18.93	21
Hungary	15.97	22
Croatia	17.07	23
Portugal	13.52	24
Bulgaria	12.22	25
Spain	10.82	26
Greece	10.86	27
Italy	10.18	28

This data set was further examined and a correlation coefficient of each indicator with the l^2-distance value was determined. The results are presented in Table 2 (using the Pearson correlation test).

Table 2. The correlation between the l^2-distance and the initial indicators

Indicator	***r***
Severely materially deprived people	0.822**
Employment rate	0.695**
People living in households with very low work intensity	0.635**
People at risk of poverty or social exclusion	0.537**
Share of renewable energy in gross final energy consumption	0.529**
Primary energy consumption	0.519**
Gross domestic expenditure on R&D	0.466*
Final energy consumption	0.390*
Tertiary educational attainment	0.366
People at risk of poverty after social transfers	0.259
Greenhouse gas emissions in non-ETS sectors	0.135
Early leavers from education and training	0.115
Greenhouse gas emissions	0.039

The correlation coefficients between the l^2-distance and initial indicators demonstrate which indicators are important in analyzing a country's development. As has been noted above, the stepwise method excludes one insignificant indicator that possesses the smallest value of the correlation coefficient. Calculating the l^2-distance should be repeated stepwise until the results show that all selected indicators are statistically significant. In this case, the stepwise method eliminated those five indicators that were insignificant in the first calculation and the results achieved in the sixth calculation are those, which are final. The result need not include all indicators that had been significant in first calculation, but may include those indicators that had been insignificant in first calculation. The results are presented in the following tables.

Table 3. The results of the l^2-distance method, l^2-distance value, and rank – last calculation

Country	**l^2-distance**	**Rank**
Sweden	69.54	1
Estonia	47.18	2
Malta	38.04	3
Finland	35.33	4
Luxembourg	30.93	5
Austria	30.89	6
Czech Republic	29.42	7
Germany	29.07	8
Denmark	26.76	9
Netherlands	25.75	10
Poland	25.35	11
Latvia	24.07	12
Slovenia	20.42	13
Lithuania	19.80	14
France	19.29	15
United Kingdom	18.19	16
Slovakia	17.39	17
Romania	17.10	18
Belgium	13.52	19
Cyprus	12.32	20
Portugal	10.86	21
Bulgaria	10.09	22

Ireland	10.08	23
Hungary	7.28	24
Croatia	6.15	25
Italy	4.44	26
Spain	3.98	27
Greece	1.01	28

Once again, a correlation coefficient of each indicator was examined with the I^2-distance, the results of which are presented in Table 4.

Table 4. The correlation between I^2-distance and final indicators

Indicator	***r***
Employment rate	0.754**
Severely materially deprived people	0.737**
People living in households with very low work intensity	0.650**
Primary energy consumption	0.617**
People at risk of poverty or social exclusion	0.576**
Gross domestic expenditure on R&D	0.501**
Final energy consumption	0.488**
Share of renewable energy in gross final energy consumption	0.463*

As it could be seen, all observed indicators from the last calculation are statistically significant, meaning that this is the final calculation of the last two tables presenting the results in examining development of the countries observed. The most important indicator is Employment rate (r=0.754, p<0.01). The employment rate varies very significantly across countries and regions – the percentage of employed people aged 20-64 is reaching 68.95% on average in the EU in 2014. Sweden have the biggest employment rate of 80.00% while on the other side percent of employed people in Greece is 53.30%. The increasing share of young people neither in employment nor in education or training is another major source of concern. Next two indicators are from poverty and social exclusion group of indicators: percent of severely materially deprived people (r=0.737, p<0.01) and percent of people living in households with very low work intensity (r=0.650, p<0.01), which proves the assertion that fight against poverty is fundamental in reaching Europe 2020 Strategy goals (Jessoula, 2015; Copeland and Daly, 2014).

In Table 4 the final results of the I^2-distance Method, I^2-distance Value, and Rank are presented. The highest values of I^2-distance belong to Sweden, which tops the list, confirming previous research (Hudrlikova, 2013). It is interesting that Sweden has already reached RES (renewable energy sources) target. Estonia, which is ranked right below Sweden, has reached its overall objectives for RES-Electricity in 2013 (Cross et al., 2015).

In contrast to Sweden and Estonia, ranking at the bottom of this list are countries with the lowest development level – Greece, Spain and Italy. Previous researchers showed that in these countries the progress has not been observed or has been negligible (Klonowska-Matynia et al., 2015). Bearing in mind the values of selected indicators (especially for Greece), these results are to be expected. Investment in education has to be strengthened in particular within these Mediterranean countries, in terms of both quantity and quality (Gros and Roth, 2012).

4. CONCLUSION

The reasons for having a Europe 2020 Strategy are equally pressing in 2016 as they were in 2010. Emerging from the worst economic and financial crisis in a generation, the EU needs to strengthen its smart, sustainable and inclusive growth strategy so that it can deliver on the expectations of its citizens and maintain its role in the world. At the moment, the European Union is on its way to fulfill the objectives in the fields of education, climate and energy policy, but not in the area of employment, research and development, and poverty reduction. Furthermore, relative poverty rates are difficult to predict when employment increases and the impact is not necessarily a beneficial one (Marx et al., 2012).

In this paper, the I-distance method has been applied in order to measure the development level of EU countries, as based on Europe 2020 Strategy indicators. The research itself started with initial indicators and several calculations led to the final set of indicators. The progress in achieving the objectives can be monitored using the described I-distance methodology, both for the EU as a whole and for each Member State individually. With this approach, not only can countries be ranked, but the differences between them can be better explored as well. In addition, the approach utilized in this paper can identify importance of each Europe 2020 Strategy indicator.

REFERENCES

Balcerzak, A.P. (2015). Europe 2020 Strategy and structural diversity between old and new member states. Application of zero unitarization method for dynamic analysis in the years 2004-2013. *Economics & Sociology,* 8(2), 190-210.

Bere, R.C., Precup, I.B. & Silvestru, C.I. (2015). On Growth Poles from EU Countries in The Framework of Europe 2020. *Procedia Economics and Finance*, 23, 920-925.

Copeland, P. & Daly, M. (2014). Poverty and social policy in Europe 2020: ungovernable and ungoverned. *Policy and Polytics,* 42(3), 351-365.

Cross, S., Hast, A., Kuhi-Thalfeldt, R., Syri, S., Streimkiene, D., & Denina, A. (2015). Progress in renewable electricity in Northern Europe towards EU 2020 targets. *Renewable & Sustainable Energy Reviews,* 52, 1768-1780.

Dobrota, M., Jeremic, V., & Markovic, A. (2012). A new perspective on the ICT Development Index. *Information Development,* 28(4), 271-280.

European Commission. (2010). Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth: Communication from the Commission. Publications Office.

Gros, D., & Roth, F. (2012). The Europe 2020 strategy: Can it maintain the EU's competitiveness in the world? Centre for European Policy Studies (CEPS), Brussels.

Hoedl, E. (2011). Europe 2020 Strategy and European Recovery. *Problemy Ekorozwoju,* 6(2), 11-18.

Hudrlikova, L. (2013). Composite indicators as a useful tool for international comparison: the Europe 2020 example. *Prague Economic Papers*, 22(4), 459-473.

Ivanovic, B., (1973). A method of establishing a list of development indicators. Paris, United Nations Educational, Scientific and Cultural Organization.

Ivanovic, B., & Fanchette, S., (1973). Grouping and ranking of 30 countries of Sub-Saharan Africa, two distancebased methods compared. Paris, United Nations Educational, Scientific and Cultural Organization.

Ivanovic, B., (1977). *Classification Theory*. Belgrade, Institute for Industrial Economic.

Jakab, Z., & Alderslade, R. (2015). Health 2020–Achieving Health and Development in Today's Europe. *Global Policy*, *6*(2), 166-171.

Jeremic, V., Bulajic, M., & Martic, M., (2011). A fresh approach to evaluating the academic ranking of world universities. *Scientometrics* 87(3), 587-596.

Jeremic, V., Bulajic, M., Martic, M., Markovic, A., Savic, G., Jeremic, D., & Radojicic, Z. (2012). An Evaluation of European Countries Health Systems through Distance Based Analysis. *Hippokratia,* 16(2), 170-174.

Jesoulla, M. (2015). Europe 2020 and the Fight against Poverty - Beyond Competence Clash, Towards "Hybrid' Governance Solutions? *Social Policy & Administration,* 49(4), 490-511.

Jovanovic, M., Jeremic, V., Savic, G., Bulajic, M., & Martic, M. (2012). How does the normalization of data affects the ARWU ranking? *Scientometrics,* 93(2), 319-327.

Klonowska-Matynia, M. & Sasin, M. (2015). EU Countries Socio-Economic Development in the Context of Europe 2020 Strategy. *Rocznik Ochrona Srodowiska,* 17(1), 771-791.

Marx, I., Vandenbroucke, P. & Verbist, G. (2012). Can higher employment levels bring down relative income poverty in the EU? Regression-based simulations of the Europe 2020 target. *Journal of European Social Policy,* 22(5), 472-486.

Milenkovic, N., Vukmirovic, J., Bulajic, M. & Radojicic, Z. (2014). A multivariate approach in measuring socio-economic development of MENA countries. *Economic Modelling,* 38, 604-608.

Radojicic, Z., Isljamovic, S., Petrovic, N., & Jeremic, V. (2012). A novel approach to evaluating sustainable development. *Problemy Ekorozwoju,* 7(1), 81-85.

Seke, K., Petrovic, N., Jeremic, V., Vukmirovic, J., Kilibarda, B., & Martic, M. (2013). Sustainable development and public health: rating European countries, BMC Public Health, 13, 1-7.

Terem, P., Cajka, P., & Rysova, L. (2015). Europe 2020 Strategy: Evaluation, implementation, and prognoses for the Slovak Republic. *Economics & Sociology,* 8(2), 154-171.

Vuković, M. (2011). Europe 2020 Strategy – priorities and goals. Faculty of Political Sciences, University of Belgrade,.Yearbook 5, 493-513.

A FUZZY LOGIC-BASED SYSTEM FOR ENHANCING SCRUM METHODOLOGY

Mihailo Stupar*[1], Pavle Milošević[1], Bratislav Petrović[1]
[1]University of Belgrade, Faculty of Organizational Sciences
*Corresponding author, e-mail: stupar.mih@gmail.com

Abstract: *In this paper, we propose a decision support system for enhancing scrum methodology based on fuzzy logic. The proposed system consists of three main components: a fuzzy inference system, an aggregation operator and a feedback function. In the basic scrum, requirements which describe a certain task do not have a clear interpretation. Also, the basic model does not take into account the experience of the developers as well as the logical dependencies of input variables. Fuzzy inference systems are particularly useful for this purpose, because it incorporates logic in inference process and inputs are presented using linguistic quantifiers. Aggregation function is used to aggregate predictions in a single value that uniquely represent a specific task while a feedback is employed to adjust an input variable to improve system performance. Further, the proposed system is simulated with randomly generated inputs in order to analyze it behavior. The predictions of the system are more accurate and with less deviation in the final iterations.*

Keywords*: scrum methodology, decision support system, fuzzy logic, fuzzy inference system, aggregation operator, feedback function*

1. INTRODUCTION

Scrum as an agile methodology is very popular within the software and product developers. Scrum is ideal for project with aggressive deadlines, complex requirements, and significant degree of uniqueness (Almseidin et al. 2015). Today, development teams are often distributed all over the world, but people still use this methodology for leading the project (Sutherland et al. 2007). In the scrum, projects move forward through the series of iterations called sprints, and each sprint is typically two to four weeks long.

At the start of each sprint, a whole team of developers has a meeting (sprint planning meeting) where they decide which tasks should be included in the following sprint. They all have a list of features/tasks (made by experts) that should be implemented by the end of the project. These features/requirements are collected in the Product Backlog (Duechting et al. 2007). Each task is weighted with some number of points. In the scrum, these points are called story points. The number of story points that come into the sprint is well known, but the number of tasks entered in the sprint depends on the developers' estimation how difficult each task is. In the sprint planning meeting, they argue about these costs of the tasks, and how many tasks from the Product Backlog will be included. They all give a value (weight) to each task independently. Later, their manager (scrum master) looking at the values for the task decides (with agreement with the team) how many story points this task actually costs.

Although the scrum methodology is successfully applied in various fields, there are certain issues that should be addressed to story point estimation. Developers often think that number of story points for each task means the number of hours/days to complete this task. The first assumption with the idea bout story points is that they are just points, relative values instead of absolute values. When someone looks into the values, he could be able to compare tasks based on story points and he should not be able to say how many hours/days are needed to complete them. Another issue is related to the experience of developers. Some developers are experienced and tasks are often too easy for them. On the other hand, some of them are bad in estimation. For example, they often give smaller value that it actually is.

Scrum master needs to think about all these parameters when decides what is the final value of story points for each task. Because fuzzy logic provides a useful tool to deal with problems in which the phenomena are imprecise and vague (Lin et al. 2006), it is ideal for our purpose. Our proposed system based on fuzzy logic could enhance efficiency in scrum planning phase. The system accepts developer's task estimation and scrum master's knowledge as inputs, and the output is weight (story point value) of the observed task. Also, the system should become more and more stable over time, and get more precise prediction.

This paper is structured as follows. In Section 2 we describe fuzzy logic and FIS in general, and how is it used previously in scrum methodology. In Section 3 we introduce our fuzzy logic-based system, an aggregation operator and a feedback function that should improve accuracy of the system. Further, we create a simulation that shows how our system works, and this is explained in Section 4. Finally, Section 5 contains conclusion and ideas for further work.

2. FUZZY LOGIC

Fuzzy logic is a generalization of classical logic in a sense that it can process all values from the interval [0,1] (Zadeh, 1965, Zadeh, 2008). It may be seen as an attempt to formalize human capability to reason and make rational decisions in an environment of imprecision and uncertainty. In fuzzy logic, operators of intersection and union are realized using different functions that are referred to as t-norms and t-conorms (or s-norms), respectively. Min operator is the standard choice for fuzzy intersection, while algebraic product and Lukasiewicznorm are also frequently used (Ross, 2010). Max operator, probabilistic sum and Lukasiewicz t-conorm are it's corresponding as t-conorms. The most common negation operator in fuzzy logic is a standard negation $\bar{x} = 1 - x$.

2.1. Fuzzy inference system

A fuzzy inference system (FIS) is a system based on fuzzy logic that utilizes a set of rules to map inputs to outputs. It is the most commonly seen fuzzy methodology. Two most important types of the FIS are Mamdani (Mamdani, 1977) and Takagi–Sugeno (Takagi and Sugeno, 1985). In the Mamdani system, both input and output are fuzzy sets and it is easier to interpret. In the Takagi–Sugeno system inputs are fuzzy sets, while the output is a linear combination of its inputs. This type of FIS is more accurately one, but more computationally expensive. The fuzzy inference systems are particularly useful in problems where inputs are expressed as linguistic expressions. The FIS is based on IF-THEN rules, fuzzy conditional statements that incorporate logic. They are a collection of linguistic statements that describe how the FIS should make a decision. Just like an algebraic variable takes numbers as values, a linguistic variable takes words or sentences as values. Their IF part is a logical condition that should be fulfilled in order to THEN part be realized. IF-THEN rules are commonly specified by a field expert.

The fuzzy inference process consists of three main steps: fuzzification, rule evaluation and defuzzification. The first step in fuzzy inference is to convert linguistic expressions to values on the unit interval [0,1] using a set of input membership functions. Further, fuzzy rules are evaluated using appropriate operators for t-norm, t-conorm and negation. Results of all IF-THAN rules are aggregated into a single fuzzy set. Finally, defuzzification is applied. It is a process that converts a fuzzy set or a fuzzy number into a crisp value, representing the final output.

2.2. Fuzzy logic in the scrum methodology

Fuzzy logic proved to be particularly useful for building expert system based on logical dependent variables. It is especially suitable when inputs are expressed as linguistic statements. Due its characteristics fuzzy logic is widely used as a tool for enhancing scrum methodology.

In (Sedehi and Martano, 2012) they introduced a new model based on fuzzy logic that is used to evaluate and monitor scrum projects. Their model has linguistic variables as inputs and output is the level of success of the (part of) scrum project. In (Kurian, 2011) is created Sugeno based fuzzy model that should determine and react to changes in an agile process, such as product/software development process. Similar approach stands in (Lin et al. 2006), where they used Mamdani based model to register changes in the enterprise world.

3. DESIGN OF FUZZY EXPERT SYSTEM

The goal of this paper is to build fuzzy expert system that can be a valuable support or even replace an expert (scrum master) during the each sprint planning phase. Rules that scrum master follow in the decision making process can be easily expressed linguistically, so the fuzzy logic system is suitable for this kind of problem (Lin et al. 2006). Proposed system consists of three components: fuzzy inference system (FIS), aggregation function and feedback function. Experts' knowledge should be used only in the first iteration in order to set up initial parameters.

3.1. Fuzzy inference system

Our fuzzy logic system has three input variables that describe the experience of developers, their estimation skills and their rating for observed task. The output is a value (story point value) of the observed task.

In case of the experience (EXP) input variable, every developer has a specific status in the company, which is based on years of experience. Instead of using four groups (junior, intermediate, senior, and expert) which is common in literature (Orlowsky et al. 2006), we decide to exclude expert group from our system. Experts are more often leaders and they define tasks that should be done instead of task implementation. So, our EXP variable consists of three membership functions, each one is for one level of experience. Every membership function in this paper is a PI - shaped function defined by four parameters. These parameters are specified by an expert using the fuzzy visualization tool in MATLAB (Sivanandam et al. 2007). In the Table 1 parameters for every membership function for EXP variable are shown. Values represent the years of experience and they are comma separated.

Table 1: Membership function parameters for EXP variable

Variable	Values of parameters
Junior	0, 0, 1.183, 2.64
Intermediate	1.52, 2.69, 3.706, 5.04
Senior	3.579, 5.18, 6.24, 8.16

Second input variable describes an accuracy/quality of developer's estimation how much the specific task is complex (EST). This variable may be of great importance for the inference process since some developers constantly miscalculate the task complexity. It is linguistic in nature and it is represented with three membership functions: Overestimated, Well Estimated, and Underestimated, while the assessments are on the interval $[1,7]$. Membership functions' parameters are shown in Table 2.

Table 2: Membership function parameters for EST variable

Variable	Values of parameters
Overestimated	-1.62, 0.779, 1.85, 3.452
Well Estimated	2.071, 3.26, 4.61, 5.99
Underestimated	4.516, 6.29, 7.3, 9.7

Bearing in mind that the accuracy of each developer's estimations may vary over time, EST values will be adjusted after each iteration. The proposed system has a sort of feedback at the end of the each iteration that should improve the validity of EST variable. Detailed explanation of feedback will be given later in the text.

Finally, the third input variable is a weight/rating that the developer gives to each task (WEI). Although human expression can be interpreted with 9 linguistic terms (Lin et al. 2006), in agreement with scrum master we decide to represent this variable with three values (easy, medium, heavy), while the expert's assessments are on the interval $[1,7]$. This task weight representation is intuitive, close to human perception and based on natural language. All these variables are PI-shaped as well, and their parameters are represented in the Table 3.

Table 3: Membership function parameters for WEI variable

Variable	Values of parameters
Easy	-1.16, 0.76, 1.75, 3.53
Medium	1.94, 3.34, 4.5, 5.93
Heavy	4.67, 6.12, 7.24, 9.16

The output of the fuzzy system is a numeric value that represents the weight of the task in story point context, and all values are split into four groups: Easy, Medium, Complex, and Very Complex. Although the valid story point values in basic scrum methodology belong to modified Fibonacci array (1/2, 1, 2, 3, 5, 8…) (Downey and Sutherland, 2013), we decided to represent output using a uniform distribution. If we use Fibonacci array, we would have a very small difference between easy and medium group and a very big gap between complex and very complex groups (see Table 4). As a result, the tasks would be weighted as a Complex or Very Complex with high probability.

Table 4: Story point distribution using Fibonacci array

Group name	Values
Easy	1/2, 1, 2
Medium	3, 5, 8
Complex	13, 21, 34
Very Complex	55, 89

It's obvious that this distribution cannot give good results because fuzzy system threats all membership functions equally and Very Complex membership function takes almost half of the full output interval. Instead of using these values, the parameters of output membership functions are rescaled (Table 5).

Table 5: Output membership functions' parameters

Group name	Values of parameters
Easy	0, 0, 2.69, 15.56
Medium	3.53, 10.92, 16.07, 25.34
Complex	16.8, 23.96, 30.02, 38.78
Very Complex	27.95, 39.98, 50, 55

Using these inputs and output we created set of rules for our fuzzy system. We follow scrum master thoughts and recommendations during the scrum planning phase. This phase is interactive and he/she leads it with his/her suggestions. These advices are transformed into the following set of rules:

1. IF (EST = WELL_EST) and (EXP is not JUNIOR) and (WEI is EASY) THEN (OUT is EASY)
2. IF (EST = OVER_EST) and (WEI is not EASY) THEN (OUT is EASY)
3. IF (EST = UNDER_EST) and (EXP is not SENIOR) and (WEI is EASY) THEN (OUT is MEDIUM)
4. IF (EST = OVER_EST) and (WEI is not EASY) THEN (OUT is MEDIUM)
5. IF (EST = OVER_EST) and (EXP is not SENIOR) and (WEI is HEAVY) THEN (OUT is COMPLEX)
6. IF (EST = UNDER_EST) and (EXP is not JUNIOR) and (WEI is HEAVY) THEN (OUT is COMPLEX)
7. IF (EST = WELL_EST) and (EXP is not JUNIOR) and (WEI is HEAVY) THEN (OUT is V_COMPLEX)
8. IF (EST = UNDER_EST) and (WEI is not EASY) THEN (OUT is V_COMPLEX)

The proposed FIS is Mamdani-type. This type of FIS ensures the necessary transparency of the decision making process. AND operator is evaluated using *min* function, OR operator with *max* function and in the defuzzification process we use centroid method.

In the Figure 1 is shown how OUT values change for different EXP and EST values and constant WEI value. If a developer is a good estimator (EST value is close to 4), the OUT value is high (since WEI is high) except in the case of inexperienced developers (EXP < 2). Since inexperienced developers do not have enough knowledge yet, they may perceive medium tasks as difficult once. On the other hand, developers who usually overestimate their tasks have OUT values smaller than predicted (about 2).

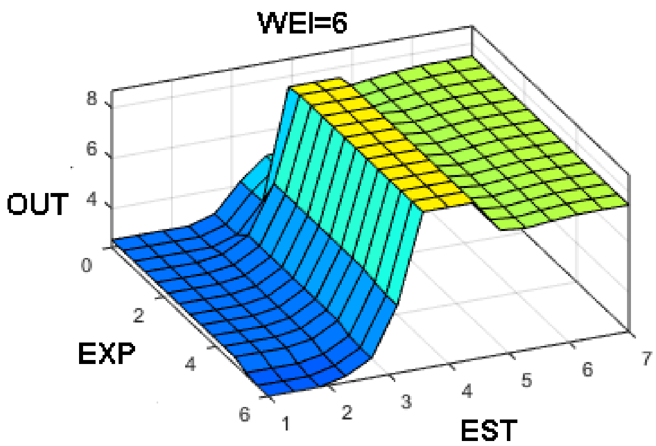

Figure 1: Output values for WEI=6

3.2. Aggregation function

Each developer has its own input vector and the fuzzy system gives a separate output for each task. In other words, for each task we get as many outputs as there are developers in our system. In second layer we aim to aggregate these predictions in a single value that uniquely represent specific task. Since characteristics of each developer are taken into account in FIS, all FIS output values should be treated equally. Therefore, we propose a simple average as the aggregation function.

Scrum master insights regarding certain team and/or project may be expressed by using various aggregation functions in this layer. For example, if the optimistic estimation is needed, the proper aggregation function is *max* function. For calculating pessimistic estimation *min* function should be used.

3.3. Feedback

The third and very important layer is feedback function which is represented in the Figure 2. At the end of each iteration/sprint we can see how difficult each task was, and this variable is represented with REAL_OUT value. Using this REAL_OUT and ESTIMATED_OUT values for the i^{th} spring we can update EST value for each developer in the $(i+1)^{th}$ sprint. For example, let assume that developer D is a good estimator and that he states that some task T is a heavy one (WEI). After the sprint if it is turned out that the task T is easy in general (REAL_OUT), developer D should not be concerned as a good estimator in the next iteration (EST).

The feedback is realized as the quadratic function of the difference between real and estimated difficulty. The function is weighted in order to get a weaker slope of the function. Using this function, we will award or penalize developer EST value if he has had a god or bad estimation.

Using this set of rules and feedback function, the system should improve itself over time and become stable after few sprints. Furthermore, we can assess developer's estimations in order to analyze their tendency to overestimate or underestimate the task over a period of time. At the beginning, we can assume that all developers are good estimators, and by the end of the project we will see how good they actually are. These (trained) values should be starting points in the next project so we could expect better results from the start.

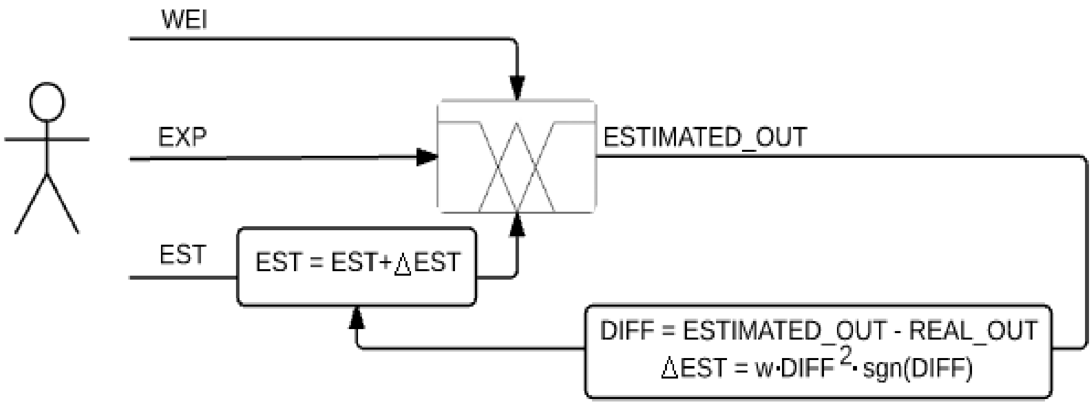

Figure 2: Design of the fuzzy expert system

4. EXPERIMENT

In this section we present the simulation in order to illustrate how the proposed system works. We will simulate the work on the project that consists of two sprints with six tasks. Five experts with various levels of experience participate in the project.

First, the initial values of the estimation accuracy EST and experience EXP variables for each developer are assigned. Further, we assign output values REAL_OUT for each task in one sprint. REAL_OUT represents the actual complexity of the task, that includes both objective circumstance and human factor. On the otherhand, it will be also used as a base for WEI value calculation in this simulation.

We are going to use these output values in order to find out the mean squared error of each task in the estimation process. EST, EXP and REAL_OUT values are taken from a uniform distribution. Finally, the weight/rating that the developer gives to each task WEI is calculated based on EST and OUT values.

WEI value calculation. Values of variable WEI cannot be randomly generated because it depends on characteristic of the tasks (summarized in REAL_OUT) and correctness of developer's estimations (explained with EST). Though, we calculate WEI in such a way that resembles a developer's evaluation.

$$OFFSET = EST - CENTER\_EST \tag{1}$$

$$WEI = OFFSET \cdot c - REAL\_OUT, \quad c = 0.35 \tag{2}$$

First, in (1) we calculate offset or distance from the center of all EST values. EST values are uniformly disturbed on the interval $[1,7]$, so in our case CENTER_EST is 4. Further, WEI value is calculated using the formula (2). The value of coefficient *c* is obtained through testing and set to 0.35.

FIS evaluation. Next, we use EST, EXP and WEI as inputs for our fuzzy inference system in order to get ESTIMATED_OUT as output. We are going to get as many ESTIMATED_OUT values for a single task as we have developers on the project. Note that in the first iteration we use INIT_EST value for EST variable. In further iterations, EST values are going to be calculated with feedback function.

EST value improvement using feedback. After each iteration EST value is updated in accordance with prediction accuracy. Thus, the adaptability of the system and its robustness to changes are assured.

$$\Delta OUT = ESTIMATED\_OUT - REAL\_OUT \tag{3}$$

$$\Delta EST = \Delta OUT^2 \cdot w \cdot \operatorname{sgn}(\Delta OUT), \quad w = 0.05 \tag{4}$$

$$EST_i = EST_{i-1} + \Delta EST \tag{5}$$

In the equation (3), we calculate the difference between predicted and real output weights. Later, we use weighted quadratic function (4) to calculate the value which will be added to the EST value from the previous iteration (5). The value of weighting coefficient *w* is set to 0.05.

The evaluation of the system. Instead of evaluating our system as a set of iterations (Sedehi and Martano, 2012), we will utilize the difference between the aggregated estimated and actual output (ΔFIN_OUT) and mean squared error (MSE) for each task as performance measures. We use ESTIMATED_OUT values from each developer and aggregate them using the average function into ESTIMATED_OUT_AVG. This value will be compared to the REAL_OUT value to obtain the difference between the estimated and actual output. MSE is used as a deviation indicator of individual prediction ESTIMATED_OUT. Firstly, we calculate mean squared error for each developer separately and then apply average function on them to get a single MSE. This value could show how stable our system is.

Table 6: MSE for Sprint 1

TASK	ESTIMATED_OUT_AVG	REAL_OUT	ΔFIN_OUT	MSE
1	3.8620	4	0.138	3.8765
2	5.1015	5	0.1015	8.8885
3	0.8593	2	1.1407	1.3015
4	1.1220	3	1.878	3.5324
5	0.8551	2	1.1449	1.3111
6	1.1198	3	1.8802	3.5687
AVERAGE			**1.0472**	**3.7465**

Table 7: MSE for Sprint 2

TASK	ESTIMATED_OUT_AVG	REAL_OUT	ΔFIN_OUT	MSE
1	3.9489	4	0.0511	1.6600
2	1.2121	3	1.7879	3.2835
3	0.8594	2	1.1406	1.3013
4	1.8630	2	0.137	0.2163
5	4.3738	4	0.3738	3.2725
6	1.1049	1	0.1049	0.1093
AVERAGE			**0.5992**	**1.6405**

In Tables 6 and 7 results for two sprints are presented together with ΔFIN_OUT and MSE. In general, the average accuracy of prediction in Sprint 2 is increased compared to Sprint 1 while the MSE value is decreased. Based on MSE values we can see that our system becomes more and more stable over time. At the beginning, even in cases when it produces a good prediction of the task complexity, i.e. for task 2 in Sprint 1, it has high deviation of individual assessments. For Sprint 2, MSE values are significantly lower. Therefore, we can state that besides greater estimation accuracy, the proposed fuzzy logic-base system improves the stability of individual assessments.

5. CONCLUSION AND FUTURE WORK

Fuzzy logic has been widely used to assist decision makers in a number of different domains. In this paper, it is utilized as a basis for building a decision support system to determine the weight of the tasks in agile methodology such as scrum. The system consists of three modules: a fuzzy inference system, an aggregation operator and a feedback function.

We consider that there is no need to use unique predefined values for estimation in the scrum (as Fibonacci series), but we can efficiently use linguistic variables to get the same goal. Using the proposed fuzzy inference systems we enhance each developer’s story point estimation in accordance with their experience and previous prediction accuracy. The knowledge of scrum master is transformed is fuzzy rules that are easy to interpret and fully resemble human reasoning. The output variables are further aggregated in a final estimation using a simple average. Feedback is applied to update the variable that represents each developer’s quality estimation in order to increase adaptability to changes and poor assessments. We have simulated the proposed system and shown that it becomes more stable over time and gives more accurate predictions of the tasks.

The proposed system could be more accurate if we take into account additional input variable in the FIS that represents how often requirement are going to change in the sprint. Some tasks/requirements are not fixed during the whole sprint and this may lead to breaking a deadline. The idea for future work is to incorporate that variable in proposed system and to add more rules that will treat these tasks differently from stable ones. Also, we aim to use various aggregation operators is ordered to model different problem situations.

REFERENCES

Almseidin, M., Alrfou, K., Alnidami N., & Tarawneh A. (2015). A Comparative Study of Agile Methods: XP versus SCRUM. *International Journal of Computer Science and Software Engineering,* 4(5), 126-129.

Downey, S., & Sutherland, J. (2013). Scrum metrics for hyperproductive teams: how they fly like fighter aircraft. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, 4870-4878. IEEE.

Düchting, M., Zimmermann, D., & Nebe, K. (2007). Incorporating user centered requirement engineering into agile software development. In *Human-computer interaction. Interaction design and usability*, 58-67. Berlin: Springer.

Kurian, T. (2011). A Fuzzy Based Approach for Estimating Agility of an Embedded Software Process. Retrieved from http://www.siliconindia.com/events/siliconindia_events/Global_Embedded_conf/Globa_Embedded_Conf_PPT_final_tisni.pdf

Lin, C. T., Chiu, H., & Tseng, Y. H. (2006). Agility evaluation using fuzzy logic. *International Journal of Production Economics*, 101(2), 353-368.

Mamdani, E. H. (1977). Application of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Transactions on Computers*, 100(12), 1182-1191.

Orłowski, C., Bach-Dąbrowska, I., Kapłański, P., & Wysocki, W. (2014). Hybrid Fuzzy-ontological Project Framework of a Team Work Simulation System. *Procedia Computer Science*, 35, 1175-1184.

Ross, T. J. (2010). *Fuzzy logic with engineering applications (*3rd ed.). Chichester: John Wiley & Sons.

Sedehi, H., & Martano, G. (2012). Metrics to Evaluate & Monitor Agile Based Software Development Projects-A Fuzzy Logic Approach. In *Software Measurement and the 2012 Seventh International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2012 Joint Conference of the 22nd International Workshop on,* 99-105. IEEE.

Sivanandam, S. N., Sumathi, S., & Deepa, S. N. (2007). *Introduction to fuzzy logic using MATLAB* (Vol. 1). Berlin: Springer.

Sutherland, J., Viktorov, A., Blount, J., & Puntikov, N. (2007). Distributed scrum: Agile project management with outsourced development teams. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*, 274a-274a. IEEE.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338-353.

Zadeh, L. A. (2008). Is there a need for fuzzy logic?. *Information sciences*, 178(13), 2751-2779.

APPLYING THE ANALYTIC HIERARCHY PROCESS TO RANK CITY-BRANCHES

Violeta Cvetkoska*[1], Nina Begicevic Redzep[2]
[1]Ss. Cyril and Methodius University in Skopje, Faculty of Economics – Skopje, Skopje, Republic of Macedonia
[2]University of Zagreb, Faculty of Organization and Informatics, Varazdin, Croatia
*Corresponding author, e-mail: vcvetkoska@eccf.ukim.edu.mk

Abstract: *The aim of the paper is to present developed AHP model for ranking city-branches of a bank. According to the number of active counters the city-branches of the bank are classified into the following three groups: group A with 1-3 active counters, group B with 2-5 active counters and group C with 4-9 active counters. The developed AHP model consists of a goal, criteria and alternatives. The goal is to rank the city-branches in the groups they belong to. The 8 criteria that are defined in the model are: denar savings, foreign currency savings, transaction accounts, exchange operations, public services, master cards, fees for payment operations (commission), and domestic payment operations. The alternatives are the city-branches that belong to each group. Developed AHP model is validated for ranking the city-branches of Commercial bank AD Skopje which are located in Skopje. The programming tool Super Decisions is used to implement the AHP model and the programming tool Expert Choice is used to perform sensitivity analysis. The obtained results are presented and analyzed in the paper.*

Key words: *multi-criteria decision-making, Analytic Hierarchy Process, ranking, sensitivity analysis, bank, city-branches*

1. INTRODUCTION

One of the key sectors that enable the economy to function is the banking sector. Banks play a vital role in the world's economy. They are authorized financial institutions that collect deposits and provide loans to individuals and legal entities, and they are largely responsible for the payment system. At the community level they do their activities through branches and city-branches.

This paper presents the applicability of the multi-criteria decision making method – the Analytic Hierarchy Process (AHP) to rank city-branches of a bank. The Analytic Hierarchy Process was developed by Thomas L. Saaty in the late seventies of the previous century (Saaty, 1977, 1980). It is designed to solve multi-criteria decision problems which can be decomposed into the following elements: goal, criteria, sub-criteria and alternatives. These elements are then structured in a hierarchy. The decision-maker makes a pair comparisons of the elements of each level of the hierarchy and provides judgments about the relative importance of each criterion regarding the goal, afterwards specifying a preference for each alternative regarding each of the criteria. The outputs of AHP are the weights of the criteria and the priorities of the alternatives. This method enables the quantitative and qualitative factors to be considered and it supports individual and group decision-making.

The AHP model for ranking the city-branches has been developed and validated on the case of one of the largest and most renowned banks in the Republic of Macedonia – Komercijalna Banka AD Skopje. The sample consists of the city-branches that are located in Skopje, while the analysis was made for 2011. In order to implement the AHP model and examine whether the ranking of the city-branches is stable, programming tools Expert Choice and Super Decisions were used.

Aside from the introduction, state of the art is given in Section 2. The objectives of the research and the research methodology are stated in Section 3. The Analytic Hierarchy Process is described in Section 4. The developed AHP model is explained in Section 5, while the validation of the model is given in Section 6. The sensitivity analysis is performed in Section 7, and the conclusion is given in Section 8.

2. STATE OF THE ART

Multi-criteria decision-making (MCDM) is a sub-field of Operational Research/Management Science, which refers to making decisions in the presence of a number of criteria that in most cases are conflicting. From the 1960-ies onwards it is considered to be an active research area and it has produced a high number of articles and books (Roy, 2005).

Velasquez & Hester (2013) give a literature review of common MCDM methods, pointing out the advantages and disadvantages of each method and their areas of application. They considered the following methods of MCDM: multi-attribute utility theory (MAUT) (see Fishburn, 1967; Keeney, 1974, 1977), Analytic Hierarchy Process (see Saaty, 1977, 1980), fuzzy set theory (see Zadeh, 1965), case-based reasoning (CBR) (see Aamodt & Plaza, 1994), data envelopment analysis (DEA) (see Charnes et al.,1978; Cooper et al., 2007; Thanasssoulis 2001), simple multi-attribute rating technique (SMART) (see Edwards, 1971, 1977), goal programming (GP) (see Charnes et al., 1955), ELimination and Choice Translating Reality (ELECTRE) (for ELECTRE I see Roy, 1968, for ELECTRE IS see Roy & Skalka (1984), for ELECTRE II see Roy & Bertier (1973) for ELECTRE III see Roy (1978), for ELECTRE IV see Roy & Hugonnard (1982) and for ELECTRE TRI see Roy & Bouyssou (1993)), Preference Ranking Organization METHod for Enrichment of Evaluations (PROMETHEE) (for PROMETHEE I and II see Brans et al. (1984)), simple additive weighting (SAW) (for SAW see Hwang & Yoon, 1981) and technique for order of preference by similarity to ideal solution (TOPSIS) (for TOPSIS see Hwang and Yoon, 1981, Yoon, 1987, Hwang et al., 1993).

Mardani et al. (2015) made a literature review for MCDM techniques and their application. They considered 393 articles, published in more than 120 international peer-reviewed journals from the Web of Science database in the period 2000-2014. The articles are grouped in the following 15 fields: energy, environment and sustainability, supply chain management, material, quality management, GIS, construction and project management, safety and risk management, manufacturing systems, technology management, operation research and soft computing, strategic management, knowledge management, production management, tourism management and other fields. The highest number of articles (109) is found in the application field of operation research and soft computing, followed by the field of energy, environment and sustainability with 53 articles, and only 5 articles are applied in the field of knowledge management. The European Journal of Operational Research has published the highest number of articles (70), followed by the Journal of Expert Systems with Applications, having published 20 articles. According to the frequency of application of decision-making techniques (AHP, ELECTRE, DEMATEL, PROMETHEE, TOPSIS, ANP, aggregation DM methods, hybrid MCDM and VIKOR), the most used one is the AHP (128 articles), followed by: the hybrid MCDM (64 articles), aggregation DM methods (46 articles), TOPSIS (45 articles), ELECTRE (34 articles), ANP (29 articles), PROMETHEE (26 articles), VIKOR (14 articles), and DEMATEL (7 articles).

When a choice of the best alternative from several alternatives has to be made, or alternatives should be ranked so that multiple criteria are taken into consideration on the basis of which alternatives are evaluated, the Analytic Hierarchy Process is one of the most commonly used MCDM methods. It can be used to solve complex problems in education, healthcare, banking, manufacturing, government, sport, etc. In the focus of this paper is the application of the AHP in banking, more specifically to rank city-branches of banks, so below we refer to the references in this field.

Javalgi et al. (1989) apply the AHP for bank management and their empirical analysis was conducted in a major metropolitan area. Arbel & Orgler (1990) apply the Analytic Hierarchy Process to bank strategic planning, i.e. bank mergers and acquisitions strategy. They developed a model that was tested in a bank holding company. Xie & Gong (2008) use fuzzy AHP and Balanced Scorecard to evaluate the performance of commercial banks. Haghighi et al. (2010) use fuzzy AHP in order to examine the impact of 3D-readiness on the development of e-banking in Iran. They have interviewed thirty bank managers and experts in Iran, and have concluded that the most important attribute of the development of e-banking is “industry e-readiness”. Onder et al. (2013) evaluate the financial performance of Turkish banks by using the methods: AHP and TOPSIS. The observed period in the study is 2002-2011. Rezaei et al. (2013) use fuzzy AHP to determine effective factors weight on optimizing the balance sheet of banks. The study is applied in the Refah bank. Nasrabadi et al. (2014) rank five branches of the Sina bank from the perspective of electronic banking by using the AHP. The following were included as criteria: efficiency and system responsiveness, quality and safety of provided data and services, customer and customer-orientation, designing and implementing e-services, and web 2.0 tools recruitment. The relative weight of the criterion efficiency and system response was the highest (0.345), followed by: quality and security of information and services (0.276), customer and customer-orientation (0.169), designing and implementing e-services (0.121), and using web 2.0 tools (0.890).

In the existing literature there was not found an article with an application of the AHP like this presented in our paper thus leading to the conclusion that this is an original application of the AHP.

3. OBJECTIVES AND RESEARCH METHODOLOGY

The objectives of the empirical research are:

- to develop AHP model for ranking the city-branches of banks;
- to validate developed AHP model by performing a ranking of the city-branches in the groups they belong to – case study of Commercial bank AD Skopje
- to perform sensitivity analysis in order to examine the sensitivity, i.e., the stability of the obtained results.

The main goal of this paper was to present developed AHP model for prioritization of city-branches of bank that is a model that can be used for comparisons of city-branches in general; to present results of the AHP model validation in comparisons of city-branches of Commercial bank AD Skopje.

In order to realize the objectives of the research, the follolwing steps were done:

- to conduct an interviews with the Manager of the Independent Domestic Payment Operations Department and with employees in this Department to define criteria for ranking the city-branches and to develop the AHP model;
- to collect judgements of respondents to assess the importance of the criteria in terms of the goal, and priorities of alternatives in terms of each criterion;
- to perform sensitivity analysis by using the programming tool Expert Choice.

In order to identify the criteria, the method of interview was used and for assessment of the importance of criteria and priorities of alternatives the results of the survey were used (it was designed a questionnaire which was distributed to respondents by e-mail).

4. THE ANALYTIC HIERARHY PROCESS (AHP)

The AHP method is one of the most widely exploited MCDM decision-making methods in cases when the decision, it means the selection of given alternatives and their prioritizing, is based on several tangible and intangible criteria (sub-criteria). The process of complex decision problem solving is based on the problem decomposition into a hierarchy structure which consists of the goal, the criteria, sub-criteria and the alternatives. Hence the AHP is a general theory of relative measurement. It is used to derive relative priorities on absolute scales from both discrete and continuous paired comparisons in multilevel hierarchic structures based on the judgment of knowledgeable and expert people (Saaty, 2001). On the basis of the pair-wise comparisons, relative significance (weights) of elements of the hierarchy structure is calculated. The AHP can combine these judgments into a single representative judgment for the group and also including the importance of the individuals themselves.

The AHP method application can be explained in four steps (Saaty & Begicevic, 2010, Begicevic et al., 2011):

1. The AHP enables decision makers to structure decisions hierarchically. The hierarchy model of the decision problem is developed in such a way that the overall goal of the decision is at the top of the model, strategic objectives in the higher levels, evaluation criteria in the middle levels, and alternative choices at the bottom.
2. After the hierarchy has been determined, the decision makers begin the procedure of prioritizing in order to determine the relative importance of elements on each level. The AHP provides a structured framework for setting priorities on each level of the hierarchy using pair-wise comparisons, a process of evaluating each pair of decision factors at a given level on the model for their relative importance with respect to their parent. On each hierarchy structure level, the pair-wise comparisons should be done by all possible pairs of the elements of this level, starting with the top of the hierarchy and working its way to the lowest level. The decision maker's preferences are expressed by numeric values on 1-3-5-7-9 scale - Intensity of Importance Scale (Table 1).
3. On the basis of the pair-wise comparisons, relative significance (weights) of elements of the hierarchy structure are calculated, which are eventually synthesized into an overall priority list of alternatives. Decision maker is allowed to change preferences and to test the results if the inconsistency level is very high. The consistency of the judgments is tracked using the rigorous math

analytics behind the AHP to validate the decision process. In cases where inconsistency is above 10% it is recommended that the criteria and judgments be revisited (Saaty, 1980).

4. Results are priorities of the alternatives and hierarchy tree with objective's relative significance. The sensitivity analysis is also carried out. Sensitivity analysis is used to determine the sensitivity of the alternatives to changes in the objectives' priorities.

Table 1: Intensity of Importance Scale (Saaty, 2012, p. 6)

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
2	Weak	
3	Moderate importance	Experience and judgment slightly favor one activity over another
4	Moderate plus	
5	Strong importance	Experience and judgment strongly favor one activity over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
8	Very, very strong	
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation
Reciprocals of above	If activity *i* has one of the above nonzero numbers assign to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	A reasonable assumption
Rationals	Ratios arising from the scale	If consistency were to be forced by obtaining *n* numerical values to span the matrix

5. DEVELOPING THE AHP MODEL FOR RANKING THE CITY-BRANCHES

The constructed AHP model consists of a goal, criteria, and alternatives. The goal is to rank the city-branches of the bank in the groups to which they belong, the alternatives are the city-branches that belong to each group, and in order to determine the criteria, the Manager of the Independent Domestic Payment Operations Department and the employees in this Department were chosen as respondents. The determination of the criteria was based on the results of interviews, questionnaires and qualitative analysis of relevant documents.

The activities which take place in the groups of city-branches are: denar savings, bank accounts, foreign currency savings, loans, transaction accounts, foreign exchange operations, other services, public services, master cards, business trip, commission, domestic payment operations, statements of transaction accounts, standing orders for utilities, issuing lists of codes, documents for bank cards, contracts for e-banking and input pensions. According to the value of each criterion for which data from internal reports of the bank departments are used, 8 activities are selected, which have the highest values in the observed period and the same serve as criteria: denar savings, foreign currency savings, transaction accounts, exchange operations, public services, master cards, fees for payment operations (commission), and domestic payment operations. The criteria are described in Table 2.

The AHP model for ranking the city-branches that belong to group A is presented in Figure 1, and in an analogous manner the AHP models for ranking the city-branches of groups B and C can be represented.

Table 2: Description of criteria

Criteria	Description
Denar savings	Deposit and money withdrawal from savings bank book, interest, authorization, representation, opening and replacement of savings book, etc.
Foreign currency savings	Deposit and money withdrawal from savings bank book, pledge of foreign currency deposit, conversion, interest, authorization, representation, opening and replacement of savings book, etc.
Transaction accounts	Cash and non-cash transfers, interests, statement of account, opening transaction account, authorization, representation, etc.
Exchange operations	Purchase/sale of foreign currency, difference in exchange rates-exchange operations, check received on encashment
Public services	Payments for public services, embassies, Western Union
Master Cards	Pay-in and pay-out (master card transfers)
Fees for payment operations (commission)	Fees (commission) calculated and charged while transaction being processed by bank clerk
Domestic Payment Operations	Cash and non-cash payments (payments-in and withdrawals) processed through payment instruments in the domestic payment operations

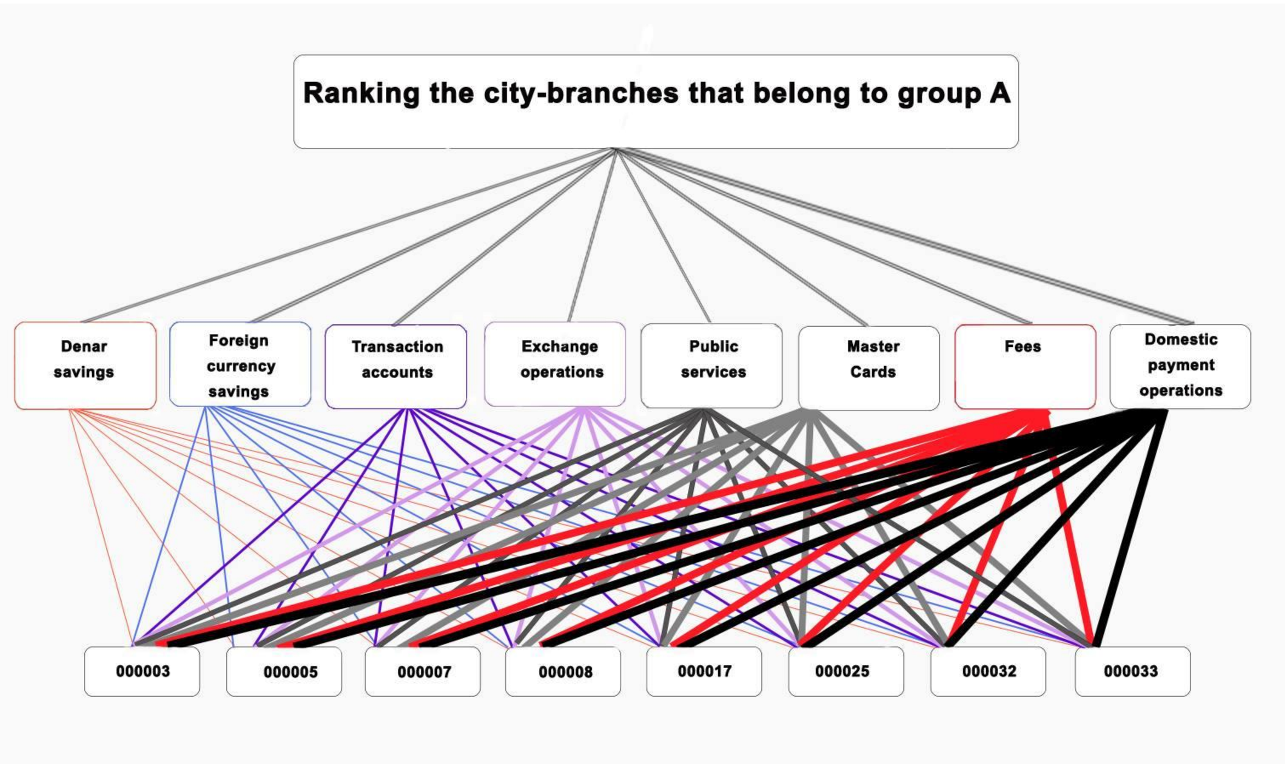

Figure 1: The AHP model for ranking the city-branches that belong to group A

6. VALIDATION OF THE AHP MODEL

The AHP is applied for ranking the city-branches of Commercial bank AD Skopje that are located in Skopje (11 Oktomvri, Avtokomanda, Biser, Buket, Bunjakovec, Butel, Vlae, Vodnjanska, GTC 1, Drachevo, Gjorce Petrov, Ekonomski fakultet, Jane Sandanski, JAT, Kapishtec, Kjubi, Leptokarija, Lisiche, Makpetrol, Madzari, MVR, Nova Makedonija, Novo Lisice, Partizanska, Rasadnik, Ruzveltova, Skopjanka, Stopanska komora na Makedonija, Sudska palata, Topansko pole, Cvetan Dimov, Centrala, Cair, Cento, Cesma, and Shuto Orizari).

Information about the profile of the bank's city-branches are obtained through the conducted interviews with the Chief Operative Officer of the bank and the Manager of the Independent Domestic Payment Operations Department.

According to the number of active counters, the city-branches are classified in the following three groups: group A with 1-3 active counters, group B with 2-5 active counters, and group C with 4-9 active counters. Groups A and B differ only in the number of employees, and group C, despite the number of employees, differs from groups A and B in that it has counters for transaction accounts and the following services can be performed: standing orders for utilities, issuing lists of codes, documents for bank cards, contracts for e-banking and input pensions.

In order to rank the city-branches of Commercial bank AD Skopje in the groups they belong to, the period from 2008 to 2011 was covered, and for an adequate comparative analysis, only the city-branches that belong to each of the groups in 2008 are taken into consideration for the entire observed period. Since the transactions which are done in the Head Office are significantly extensive, more complex and may not be comparable and placed in correlations made between the three groups of branches, they are excluded from the analysis. The research in the bank was conducted in 2011 and the data was collected in the period between 2008-2011 in which was obtained permission to use the data. The data used in the AHP model is not made public and is considered confidential, therefore it is not given in the paper, and the real name of the city-branches have been replaced with numbers. The following eight city-branches belong to group A: 000003, 000005, 000007, 000008, 000017, 000025, 000032 and 000033. Eleven city-branches belong to group B, and they are: 000006, 000009, 000012, 000016, 000020, 000027, 000028, 000030, 000035, 000036 and 000037, while the following ten city-branches belong to group C: 000004, 000010, 000013, 000014, 000018, 000019, 000021, 000023, 000031 and 000034. The city-branch 000010 has no counter for transaction accounts, and the city-branches 000013 and 000014, despite the services on the counters for transaction accounts, conduct international payment for legal entities. In this paper the focus is on the last year of the observed period, i.e. 2011.

After developing the AHP model, a questionnaire was composed in which the respondents (Officers of the Analysis, Information and Support Office) were asked to do pair-wise comparison in each level of the hierarchy and to express their preferences using the intensity of importance scale. The respondents first had to compare the criteria that were given in pairs by using the option of importance (i.e. which one of the two criteria that are compared in a pair is more important for the goal – ranking the city-branches in the groups they belong to in the observed time frame) and afterwards express their preferences with the help of the Saaty's scale. Next, in the same questionnaire, the respondents had to compare the alternatives, i.e. the city-branches that were given in pairs in regards to each criterion, and on the basis of the data for the observed period, to use the option of priority (which of the two city-branches that are compared in a pair is given priority in regards to the criterion) and to express their preferences with the help of the same scale. The prepared questionnaire was quite huge (a total of 320 pages), so it represented a fairly complex task for the respondents. The mentioned questionnaire is not added as an appendix due to its size.

The questionnaire was sent to the respondents by e-mail, and after they filled it, they submitted the quastinnaire by e-mail, determining that the questionnaires were fully completed. The values that were given in the questionnaires by the respondents were entered in the programming tool Super Decisions (Super Decisions Software, 2006) for each group of city-branches separately. There are four ways of assessing the comparisons in pairs in this software tool: graphically, verbally, through a matrix, and through a questionnaire. The data from the questionnaires that were filled by the respondents were entered in the tool Super Decisions through the choice of questionnaire. The obtained results i.e. the weights of the criteria and priorities of alternatives, are presented and analyzed below.

The weights of the criteria are shown in Figure 2. It can be seen from this Figure that only one criterion, i.e. the criterion of public services has a weight 0.045455, while the remaining seven criteria have same weight (0.136364).

The priorities of the alternatives that belong to groups A, B and C are shown in Figures 3, 4 and 5 respectively. The Consistency Ratio (CR) in all of the analyzed cases is 0.01 and it follows that the results are consistent.

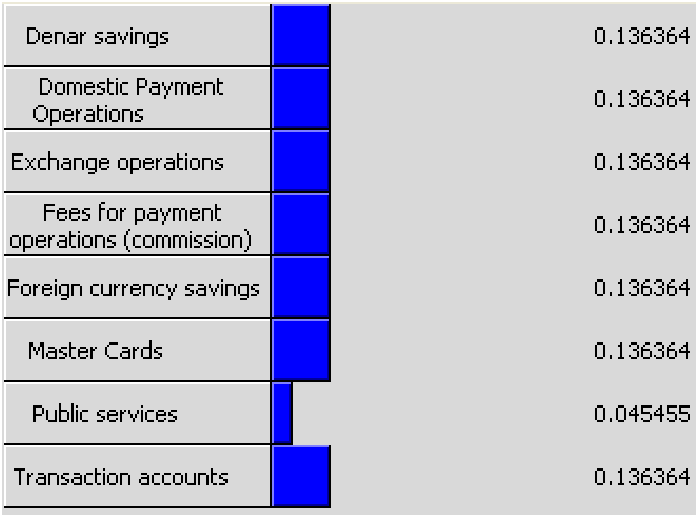

Figure 2: Weights of criteria

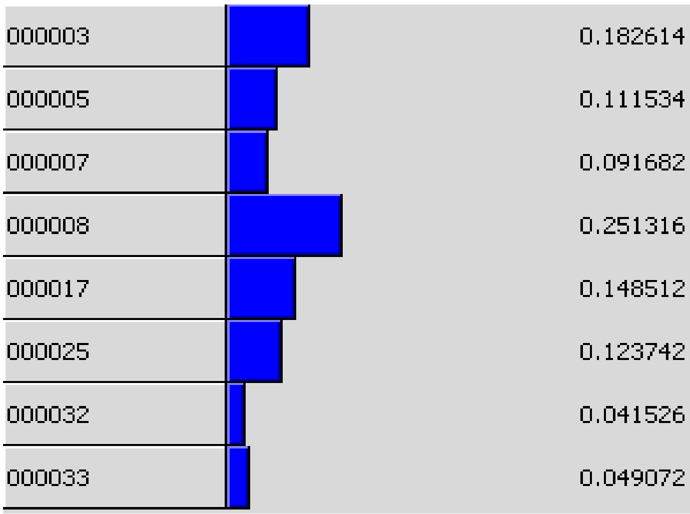

Figure 3: Priorities of city-branches that belong to group A

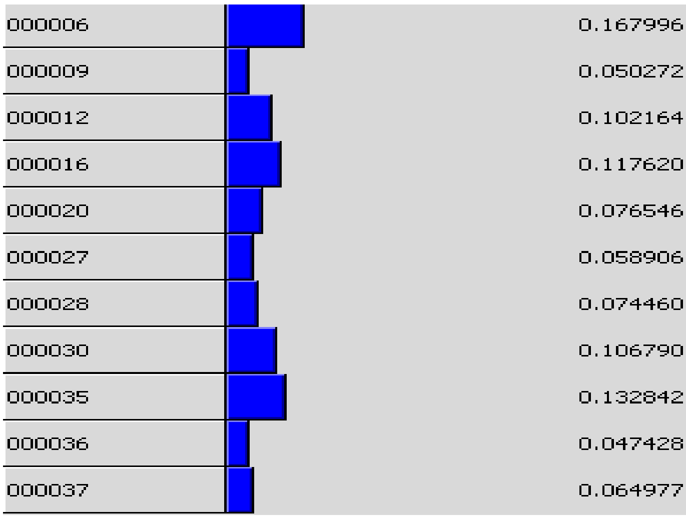

Figure 4: Priorities of city-branches that belong to group B

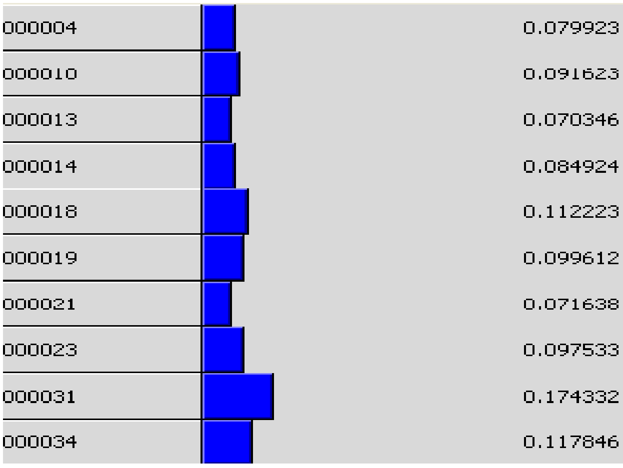

Figure 5: Priorities of city-branches that belong to group C

Based on the priorities of the city-branches their ranking is performed in the groups in which they belong (Table 3). It can be seen from this Table that in group A the highest ranked city-branch is 000008, followed by the city-branch 000003, while the lowest ranked branch is 000032. In group B, the highest ranked city-branch is 000006, followed by the city-branch 000035, while the lowest ranked city-branch is 000036. In group C, the highest ranked city-branch is 000031, followed by the city-branch 000034, and the lowest ranked city-branch is 000013.

Table 3: Rank of the city-branches of Commercial bank AD Skopje

City-branches that belong to group A	Rank	City-branches that belong to group B	Rank	City-branches that belong to group C	Rank
000003	2	000006	1	000004	8
000005	5	000009	10	000010	6
000007	6	000012	5	000013	10
000008	1	000016	3	000014	7
000017	3	000020	6	000018	3
000025	4	000027	9	000019	4
000032	8	000028	7	000021	9
000033	7	000030	4	000023	5
		000035	2	000031	1
		000036	11	000034	2
		000037	8		

7. SENSITIVITY ANALYSIS

Sensitivity analysis offers a stable solution and it enables change of inputs in order to observe the consequences on outputs, i.e. the priorities of the alternatives (Begicevic, Divjak & Hunjak, 2007). In order for a sensitivity analysis to be performed, the programming tool Expert Choice (Expert Choice, 2005) has been used. This tool contains the following five options for a sensitivity analysis: 1. Performance; 2. Dynamic; 3. Gradient; 4. Head to Head and 5. 2 D.

Most interesting for analysis are the three top ranked city-branches from each group for 2011, and also there have been chosen three criteria. Out of the eight criteria, seven of them are equally important for each model, hence resulting in a combination of three criteria (if the criteria had a different importance, then it would have been suitable to choose the three criteria that have the highest weights). Such an approach, with three criteria and three alternatives, could have been realized in the trial version of the Expert Choice software, for which access was given, while in the models that were created in Expert Choice, data from the filled questionnaires by the respondents were used.

For the first AHP model, in which the city-branches that belong to group A are ranked, the top ranked have been chosen to be the city-branches: 000008, 000003 and 000017; the criteria being: foreign currency savings, exchange operations, and domestic payment operations. For the purpose of comparing the elements of the hierarchy in pairs, the data from the filled questionnaires by the respondents were used (explained in Section 6). The weight for each of the criteria is 0.333, while for the alternatives 000008, 000003 and 000017, the overall priorities are 0.483, 0.297 and 0.220, respectively.

Figure 6 is created through choosing the option of Performance from the menu Sensitivity-Graphs of the Expert Choice software. The weights of each criterion and the ranking of the city-branches can be seen in it. Additionally, the order of the three city-branches for each criterion can be seen through this option. In regards to the criterion on foreign currency savings, the city-branches 000008 and 000003 are equally preferred, while regarding the other two criteria, the most preferred is the city-branch 000008. In order to see how the change of the input data (criteria) will reflect on the final results, there has been made a change of the importance of the criterion of foreign currency savings (because of the above-mentioned statement for both the alternatives that are highly ranked), so its importance has increased by 20%, which has seen a slight decrease in the overall priority of the city-branch 000008, a slight increase in the overall priority of the city-branch 000003, the city-branch 000017 having no changes, and the ranking of the city-branches stays the same. In addition, the importance of the criterion of foreign currency savings has changed (from 33.3% increasing to 60%), but there have been no changes in the ranking. Figure 6 also displays the alternatives 000003 and 000017 as being equally preferred in regards to the criterion of exchange operations, and its importance has changed as well to: 40%, 50%, and 60%, but once again, the ranking of the city-branches does not change.

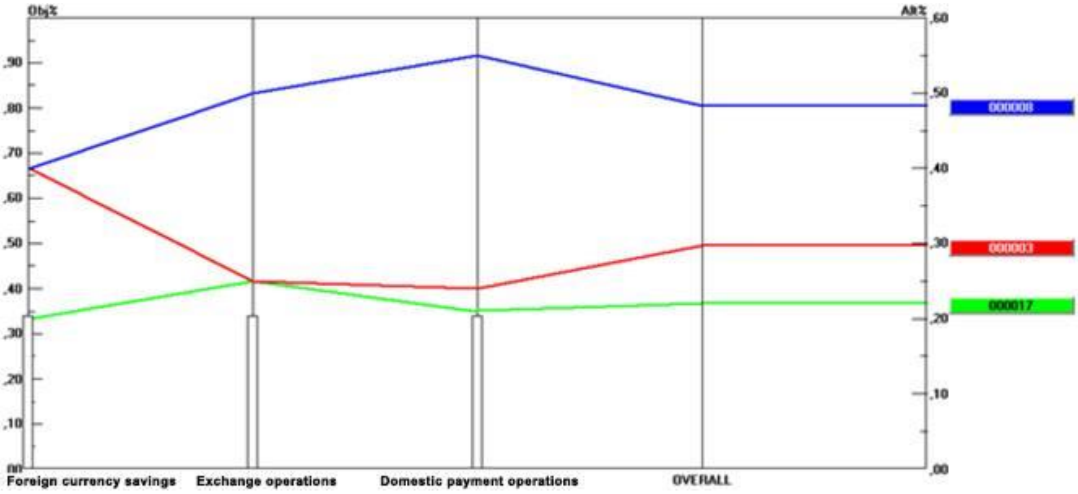

Figure 6: Sensitivity analysis - option Performance for the first AHP model

Through the option of Gradient, one can notice the sensitivity of the overall priorities of the city-branches 000008, 000003 and 000017 on the change of the importance of each criterion separately. The red vertical line shows how large the weight is for the chosen criterion. Figure 7 displays the conclusion that if the weight of the criterion of foreign currency savings increases, there is a decrease in the priority of city-branch 000008, an increase in the priority of city-branch 000003, while the priority of city-branch 000017 does not undergo significant changes.

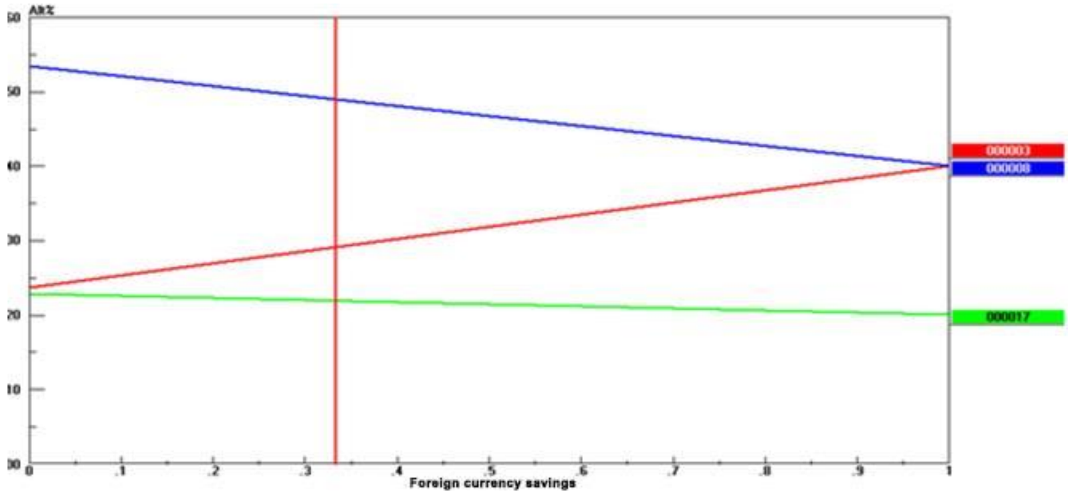

Figure 7: Sensitivity analysis - option Gradient for the criterion of foreign currency savings for the first AHP model

The option of Head-to-head is followed for one pair of city-branches (000017 and 000003). Figure 8 shows that the city-branch 000003 has an advantage over city-branch 000017, regarding the criteria: foreign currency savings and payment operations. What can be noticed in addition is that the weighed advantage of the city-branch 000003 (the gray triangle oriented towards the right) over city-branch 000017 is not considerably high.

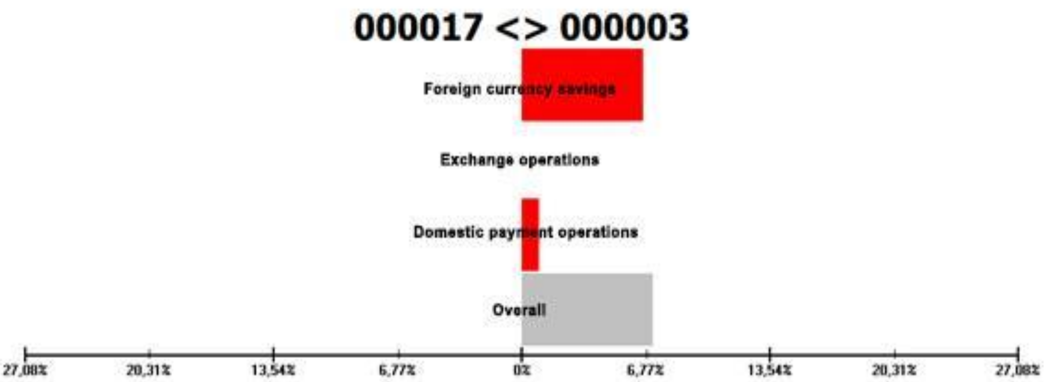

Figure 8: Sensitivity analysis - option Head-to-head for the city-branches: 000017 and 000003

In addition, the sensitivity analysis through the option of 2D is explained. In regards to the chosen criteria: foreign currency savings and exchange operation, the top city-branch is 000008 (Figure 9), and for its advantage over city-branch 00003 the contriubting criterion is exchange operations. City-branches 000003 and 000017 have the same priority in regards to the criterion of exchange operations, but when we take into consideration the criterion of foreign currency savings, then city-branch 000003 is in advantag

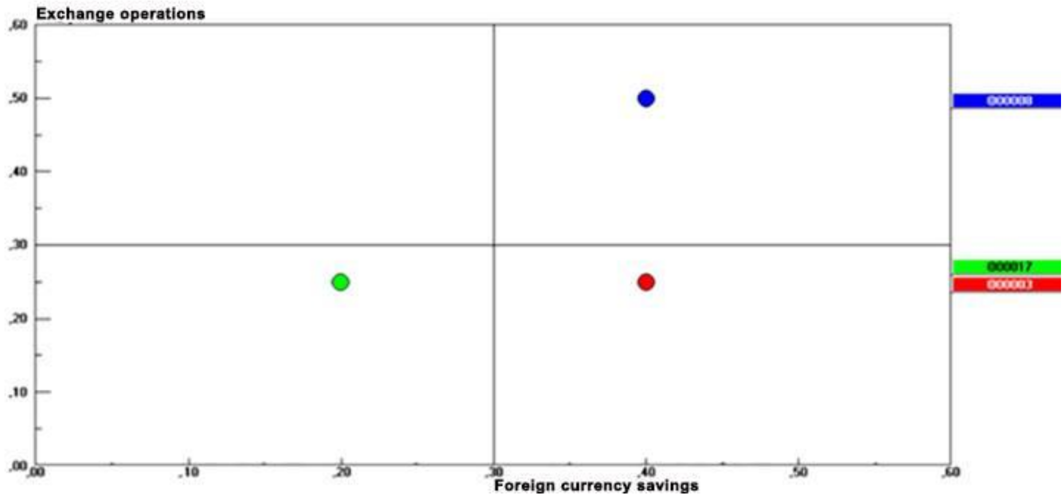

Figure 9: Sensitivity analysis - option 2D for the criteria: foreign currency savings and exchange operations for the first AHP model

For the purpose of analyzing whether the ranking of the three city-branches from Group A is stable, a sensitivity analysis is conducted, through the option of Dynamic, with the significance of each criteria separately increased by 5%. On the basis of the analyzed results it has been concluded that with the increase of the significance of the criteria by 5%, the ranking of the three city-branches from Group A is stable.

For the second AHP model, which ranks the city-branches that are part of Group B, the three top ranking city-branches for 2011 have been distinguished: 000006, 000035 and 000016. At the same time, the following three criteria have been chosen: denar savings, master cards and commission. For the third AHP model, which ranks the city-branches that belong to Group B, the three top ranked city-branches for 2011 have been chosen (000031, 000034 and 000018), while the criteria chosen are: denar savings, foreign currency savings and transaction accounts. In order to analyze whether the ranking of the three city-branches of Groups B and C is stable, a sensitivity analysis has been conducted, through the option of Dynamic, thus concluding that the ranking is stable.

8. CONCLUSION

The AHP can be applied for solving numerous MCDM problems (planning, optimizing, measuring performances, resource allocation, conflict solving, etc.). Through the conducted research presented in this paper, the following scientific contributions have been achieved: the AHP model for ranking the city-branches of a bank was developed and it was validated on the case of Komercijalna Banka AD Skopje. In our experience, the developed AHP model has strongly motivated all of the respondents (experts in banking) because their knowledge and preferences were incorporated in it. The developed AHP model with obtained weights of criteria can be used for prioritization of the city-branches of banks in general. The benefits of using this model are: the calculation of weights of criteria for ranking the city-branches can help bank managers to be more objective in the process of ranking; the ranking procedure is more transparent and simpler; better quality in decision-making at city-branches; improving the performance of city-branches that has a positive influence on the successful operating of the bank in which they belong to. In our further research we plan to develop generic AHP model for comparison of city-branches, as well as to use statistical methods for defining criteria. We will also try to develop the ANP (Analytic Network Process) model for ranking city-branches of the bank.

REFERENCES

Aamodt, A. & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1), 39-59.

Arbel, A. & Orgler, Y.E. (1990). An application of the AHP to bank strategic planning: The mergers and acquisitions process. *European Journal of Operational Research*, 48(1), 27-37.

Begicevic, N, Divjak, B. & Hunjak, T. (2011). AHP-based group decision making using keypads. *International Journal of Economics and Business Research*, 3(4), 443-458.

Begicevic, N., Divjak, B. & Hunjak, T. (2007). Prioritization of e-learning forms: a multicriteria methodology. *Central European Journal of Operations Research*, 15(4), 405-419.

Brans, J. P., Mareschal, B. & Vincke Ph. (1984). PROMETHEE: A new family of outranking methods in multicriteria analysis. In: J. P. Brans (Ed.), Operational Research, IFORS 84 (291-302). Amsterdam: North Holland.

Charnes, A, Cooper, W. W. & Ferguson, R. O. (1955). Optimal estimation of executive compensation by linear programming. *Management Science*, 1(2), 138-151.

Charnes, A, Cooper, W. W. & Rhodes, E. L. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444

Cooper, W. W, Seiford, L. M. & Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software*. (2nd ed.). New York: Springer Science + Business Media, LLC.

Edwards, W. (1971). Social utilities. *Engineering Economist (Summer Symposium Series)*, 6, 119-129.

Edwards, W. (1977). How to use multiattribute utility measurement for social decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 7, 326-340.

Expert Choice, Inc. (2005). http://www.expertchoice.com/ [Accessed in March and May, 2012].

Fishburn, P. (1967). Conjoint measurement in utility theory with incomplete product sets. *Journal of Mathematical Psychology*, 4(1), 104-119.

Haghighi, M, Divandari, A. & Keimas, M. (2010). The impact of 3D e-readiness on e-banking development in Iran: A fuzzy AHP analysis. *Expert Systems with Applications*, 37(6), 4084-4093.

Hwang, C. L. & Yoon, K. (1981). *Multiple attribute decision making methods and applications: A state of the art survey*. New York: Springer-Verlag.

Hwang, C.L., Lai, Y.J. & Liu, T.Y. (1993). A new approach for multiple objective decision making. *Computers and Operational Research*, 20(8), 889-899.

Javalgi, R.G, Armacost, R.L. & Hosseini, J.C. (1989). Using the analytic hierarchy process for bank management: Analysis of consumer bank selection decisions. *Journal of Business Research*, 19(1), 33-49.

Keeney, R. & Fishburn, P. (1974). Seven independence concepts and continuous multiattribute utility functions. *Journal of Mathematical Psychology*, 11(3), 294-327.

Keeney, R. (1977). The art of assessing multiattribute utility functions. *Organizational Behavior and Human Performance*, 19(2), 267-310.

Mardani et al. (2015). Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istrazivanja*, 28(1), 516-571.

Nasrabadi, S.N, Fallah, E. & Hasanzadeh, A. (2014). Application of analytical hierarchy process (AHP) to rank selected branches of Sina bank from the perspective of e-banking. *Interdisciplinary Journal of Contemporary Research in Business*, 6(1), 222-232.

Onder, E, Tas, N. & Hepsen, A. (2013). Performance evaluation of Turkish banks using analytical hierarchy process and TOPSIS method. *Journal of International Scientific Publication: Economy & Business*, 7(1), 470-503.

Rezaei et al. (2013). Effective factors on optimizing banks' balance sheet using fuzzy analytical hierarchy process. *Management Science Letters*, 3(11), 2781-2786.

Roy, B. & Bertier, P. (1973). La methode Electre II: Une application au media planning. In M. Ross (Ed.), *Operational Research 1972* (pp. 291-302). Amsterdam: North Holland Publishing Company.

Roy, B. & Bouyssou D. (1993). Aide multicritere a la decision: Methodes et cas. economica.

Roy, B. & Huggonard, J.-.C.H.R. (1982). Ranking of suburban line extension projects of the Paris metro system by a multicriteria method. *Transportation Research*, 16A(4), 301-312.

Roy, B. (1968). Classement et choix en presence de points de vue multiples (la methode Electre). *Revue Francaise d'Informatique et de Recherche Operationnelle*, 8(1), 516-571.

Roy, B. (1978). Electre III: un algorithme de classement fonde sur une representation floue des preferences en presence de criteres multiples. *Cahires du CERO* , 20(1), 3-24.

Roy, B. (2005). Paradigms and challenges. In Figueira, J, Greco, S. & Ehrgott, M. (Eds), *Multiplecritera decision analysis: State of the arts surveys* (pp. 3-24). New York: Springer.

Roy, B. & Skalka, M. (1984, February). Electre IS: Aspects methodologiques et guide d'utilisation. *Document Du LAMSADE*, 125.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3), 234-281.

Saaty, T. L. (1980). *Multicriteria decision making: The analytic hierarchy process*. Pittsburgh: RWS Publications.

Saaty, T.L. & Begicevic, N. (2010). The scope of human values and human activities in decision making. *Applied Soft Computing*, 10(4), 963-974.

Saaty, T.L. & Vargas, L. (2012). *Models, methods, concepts and applications of the analytic hierarchy process*. (2nd ed.). New York: Springer Science+Business Media.

Saaty, T.L. (2001). *Decision making in complex environments. The analytic network process for decision making with dependence and feedback*. Pittsburgh: RWS Publications.

Super Decisions Software (2006). http://www.superdecisions.com/ [Accessed in March, April and May, 2012].

Thanassoulis, E. (2001). *Introduction to the theory and application of data envelopment analysis: A foundation text with integrated software*. Boston: Kluwer Academic Publishers.

Velasquez, M. & Hester, P. T. (2013). An analysis of multi-criteria decision making methods. *International Journal of Operations Research*, 10(2), 56-66.

Xie, W. & Gong, H. (2008). Using fuzzy analytic hierarchy process and balanced scorecard for commercial bank performance assessment. In ISBIM'08 Proceedings of the 2008 International Seminar on Business and Information Management (pp. 432-435). Washington: IEEE.

Yoon, K. (1987). A reconciliation among discrete compromise situations. *Journal of Operational Research Society*, 38(3), 272-286.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.