e-ISSN: 2458-8091

Journal Homepage: http://dergipark.gov.tr/ejsdr

EJSDR

European Journal of Sustainable Development Research

DergiPark AKADEMİK

CODEN: EJSDR

# Data Mining Techniques in Database Systems

**Ledion Liço[1]***

[1*]*Polytechnic University of Tirana, Faculty of Information Technology, Tirana, Albania,*

**Corresponding Author email: ledionlico@hotmail.com*

## Publication Info

*Paper received:*
01 June 2016

*Revised received:*
19-23 October 2016

*Accepted:*
01 March 2017

## Abstract

At the current stage the technologies for generating and collecting data have been advancing rapidly. The main problem is the extraction of valuable and accurate information from large data sets. One of the main techniques for solving this problem is Data Mining. Data mining (DM) is the process of identification and extraction of useful information in typically large databases. DM aims to automatically discover the knowledge that is not easily perceivable. It uses statistical analysis and artificial intelligence (AI) techniques together to address the issues. There are different types of tasks associated to data mining process. Each task can be thought of as a particular kind of problem to be solved by a data mining algorithm. The main types of tasks performed by DM algorithms are: Classification, Association, Clustering, Regression, Anomaly Detection, Feature Extraction, Time Series Analyses.

In this paper we will perform a survey of the techniques above. A secondary goal of our paper is to give an overview of how DM is integrated in Business Intelligence (BI) systems .BI refers to a set of tools used for multidimensional data analysis, with the main purpose to facilitate decision making. One of the main components of BI systems is OLAP. The main OLAP component is the data cube which is a multidimensional database model that with various techniques has accomplished an incredible speed-up of analyzing and processing large data sets. We will discuss the advantages of integrating DM tools in BI systems.

## Key words

*Data Mining, BI, OLAP, AI, OLAM*

## 1. INTRODUCTION

Data mining is the process used to describe knowledge in databases. Data mining process is very much useful for extracting and identifying useful information and subsequent knowledge from large databases. It uses different techniques such as statistical, mathematical, artificial intelligence and machine learning as the computing techniques. Its predictive power comes from unique design by combining techniques from machine learning, pattern recognition, and statistics to automatically extract concepts, and to determine the targeted interrelations and patterns from large databases. Organizations get help to use their current reporting capabilities to discover and identify the hidden patterns in databases. The extracted patterns from the database are then used to build data mining models, and can be used to predict performance and behavior with high accuracy [2]. Descriptive and Predictive data mining are the most important approaches that are used to discover hidden information[1] Data Mining has become an established discipline within the scope of computer science. The origins of data mining can be traced back to the late 80s when the term began to be used, at least within the research community. In the early days there was little agreement on what the term data mining encompassed, and it can be argued that in some sense this is still the case. Broadly data mining can be determined as a set of mechanisms and techniques, realized in software, to extract hidden information from data. The word hidden in this definition is important; SQL style querying, however sophisticated, is not data mining. By the early 1990s data mining was commonly recognized as a sub-process within a larger process called Knowledge Discovery in Databases or KDD. The most commonly used definition of KDD is that attributed to (Fayyad et al. 1996).. The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad et al. 1996). As such data mining should be viewed as the sub-process, within the overall KDD process, concerned with the discovery of hidden information". Other sub-processes that form part of the KDD

process are data preparation (warehousing, data cleaning, pre-processing, etc) and the analysis/visualization of results. For many practical purposes KDD and data mining are seen as synonymous, but technically one is a sub-process of the other. There are two important models in Data Mining: The Descriptive Model and The Predictive Model.

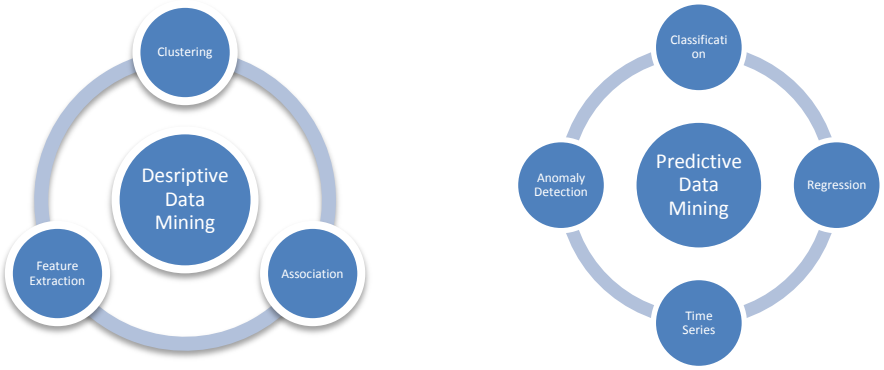


*Figure 1. Descriptive (unsupervised) and Predictive (supervised) Data Mining*

Descriptive model looks at data and analyzes past events for insight as to how to approach the future. This technique is also known as unsupervised learning. Descriptive analytics looks at past performance and understands that performance by mining historical data to look for the reasons behind past success or failure. Almost all management reporting such as sales, marketing, operations, and finance, uses this type of post-mortem analysis. Descriptive models quantify relationships in data in a way that is often used to classify customers or prospects into groups. Unlike predictive models that focus on predicting a single customer behavior, descriptive models identify many different relationships between customers or products. Descriptive models do not rank-order customers by their likelihood of taking a particular action the way predictive models do. The Descriptive model uses techniques as Clustering, Association Rules, Summarizations, and Feature Extraction.

Predictive models turns data into valuable, actionable information. Predictive analytics uses data to determine the probable future outcome of an event or a likelihood of a situation occurring. This technique is also known as supervised learning. Supervised data mining techniques are appropriate when you have a specific target value to predict about your data. The targets can have two or more possible outcomes, or even be a continuous numeric value.In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding decision making for candidate transactions. The Predictive data mining model includes Classification, Anomaly Detection, Regression and Analysis of Time Series.

## 2. THE DESCRIPTIVE MODEL

There are 3 important techniques used in the descriptive model: Clustering, Association, Feature Extraction.

### *2.1. Clustering*

Clustering is an important technique in data mining and it is the process of partitioning data into a set of clusters such that each object in a cluster is similar to another object in the same cluster, and dissimilar to every object not in the same cluster. Dissimilarities and similarities are assessed based on the attribute values describing the objects and often involve distance measures. Clustering analyses the data objects without consulting a known class label. This is because class labels are not known in the first place, and clustering is used to find those labels. Good clustering exhibits high intra-class similarity and low inter-class similarity, that is, the higher the similarity of objects in a given cluster, the better the clustering. The superiority of a clustering algorithm depends equally on the similarity measure used by the method and its implementation. The superiority also depends on the algorithm's ability to find out some or all of the hidden patterns. The different ways in which clustering methods can be compared are partitioning criteria, separation of clusters, similarity measures and clustering space. Clustering algorithms can be categorized into partition-based algorithms, hierarchical-based algorithms, density-based algorithms and grid-based algorithms.

*Table 1.Some of the most used Clustering Algorithms*

| Partition-based algorithms | Hierarchical-based algorithms | Density-based algorithms | Grid-based algorithms |
|---|---|---|---|
| K-Means | Agglomerative(BIRCH,CHAMALEON) | DBSCAN | STING |
| K-Medoids(PAM,CLARA) | Divisive | DENCLUE | CLIQUE |

These methods vary in (i) the procedures used for measuring the similarity (within and between clusters) (ii) the use of thresholds in constructing clusters (iii) the manner of clustering, that is, whether they allow objects to belong to strictly to one cluster or can belong to more clusters in different degrees and the structure of the algorithm[3].

### *2.2. Association*

Another important data mining technique is association rule mining. Association rule technique searches for relationships among variables. For example, a shop might gather data about how the customer is purchasing the various products. With the help of association rule, the shop can identify which products are frequently bought together and this information can be used for marketing purposes. This is sometimes known as market basket analysis. The patterns discovered with this data mining technique can be represented in the form of association rules. Rule support and confidence are two measures of rule interestingness. Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Such thresholds can be set by users or domain experts.

*Definition*. Let $I=\{I_1, I_2, ..., I_m\}$ be a set of items. Let D, the task relevant data, be a set of database transactions where each transaction T is a set of items such that $T \subseteq I$. Each transaction is associated with an identifier, called TID. Let A be a set of items. A transaction T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \rightarrow B$, where $A \subset I$, $B \subset I$ and $A \cap B = \Phi$ . The rule $A \rightarrow B$ holds in the transaction set D with support s, where s is the percentage of transactions in D that contain $A \cup B$. The rule $A \rightarrow B$ has confidence c in the transaction set D if c determines how frequently items in B appear in transactions that contain A .That is, support$(A \rightarrow B)$ = Prob$\{A \cup B\}$ and confidence$(A \rightarrow B)$ = Prob$\{B/A\}$.[4]

*Table 2. An example of market basket transactions*

| TID | Item |
|---|---|
| 1 | {Jeans,T-Shirt,Shoes,Chocolate} |
| 2 | {Jeans,Shoes,Coat,Sunglasses} |
| 3 | {Watch,Bag,,Jeans,T-Shirt} |
| 4 | {Belt,Jeans,Shirt,Shoes} |

Lets assume that that itemset A includes {Jeans,T-Shirt} and itemset B includes {Shoes}.The rule {Jeans,T-Shirt}$\rightarrow${Shoes} has a support value of 2/4 =0.5 and a confidence value of 2/3=0.67. Rules that satisfy both a user-specified minimum support threshold and a minimum user-specified confidence threshold (minconf) are called strong.

Some of the most important association algorithms are : AIS, SETM, Apriori, Aprioritid, Apriorihybrid, FP-Growth. From the recent studies it is observed that, FP-growth performs better ind terms of speed and accuracy than the older AIS, SETM, Apriori, Aprioritid, Apriorihybrid [5]

### *2.3. Feature Extraction*

Feature extraction creates new features based on attributes of your data. These new features describe a combination of significant attribute value patterns in your data. Models built on extracted features may be of higher quality, because the data is described by fewer, more meaningful attributes.. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Representing data points by their features can help compress the data (trading dozens of attributes for one feature), make predictions (data with this feature often has these attributes as well), and recognize patterns. Additionally, features can be used as new attributes, which can improve the efficiency and accuracy of supervised learning techniques (classification, regression, anomaly detection, etc.).

Some of the most commonly used techniques for feature extraction are: Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) .Principal Component Analysis (PCA) is the most popular statistical method. This method extracts a lower dimensional space by analyzing the covariance structure of multivariate statistical observations.Linear Discriminant Analysis (LDA) technique mainly projects the high-dimensional data into lower dimensional space. LDA aims to maximize the between-class distance and minimize the within-class distance in the dimensionality reduced space. [6] Many extensions to of LDA technique have been proposed in the past like NLDA(Null space LDA) and OLDA (Orthogonal LDA).An asymmetric principal component analysis (APCA) was proposed by Jiang et 2009 to remove the unreliable dimensions more effectively than the conventional PCA.

## 3. THE PREDICTIVE MODEL

There are 4 important techniques used in the descriptive model: Classification, Regression, Time Series and Anomaly Detection.

### *3.1. Classification*

Classification techniques in data mining are capable of processing a large amount of data. It can be used to predict categorical class labels and classifies data based on training set and class labels and it can be used for classifying newly available data. The term could cover any context in which some decision or forecast is made on the basis of presently available information. Classification procedure is recognized method for repeatedly making such decisions in new situations. Creation of a classification procedure from a set of data for which the exact classes are known in advance is termed as pattern recognition or supervised learning. Contexts in which a classification task is fundamental include, for example, assigning individuals to credit status on the basis of financial and other personal information, and the initial diagnosis of a patient's disease in order to select immediate treatment while awaiting perfect test results. Some of the most critical problems arising in science, industry and commerce can be called as classification or decision problems.

Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. Next the algorithm is given a data set not seen before, called prediction set, which contains the same set of attributes, except for the prediction attribute – not yet known. The algorithm analyses the input and produces a prediction. [7]For example in a hospital where the target attribute is the illness of the patient, in the hospital database the training set will include the symptoms of the previous recorded patients and the illness as a target. The algorithm than is given a prediction set with the data from the new patient except the illness which is the one attribute needed to predict. The prediction accuracy defines how "good" the algorithm is.How well predictions are done is measured in percentage of predictions hit against the total number of predictions. A decent rule ought to have a hit rate greater than the occurrence of the prediction attribute.

The commonly used methods for data mining classification tasks can be classified into the following groups [8] .

- Decision Trees (DT's)
- Support Vector Machine (SVM)
- Genetic Algorithms (GAs) / Evolutionary Programming (EP)
- Fuzzy Sets
- Neural Networks
- Rough Sets

Multi-Objective Genetic Algorithms (MOGA) have been also used recently to address classification data mining tasks.

### *3.2. Regression*

Regression is a data mining (machine learning) technique used to fit an equation to a dataset. It is more used when the target attribute has a numeric value. It can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. The main types of regression methods are:

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

The simplest form of regression, linear regression, uses the formula of a straight line ($y = mx + b$) and determines the appropriate values for m and b to predict the value of y based upon a given value of the coefficients, *m* and *b* (called *regression coefficients*),specify the slope of the line and the *Y*-intercept, respectively. Multivariate linear regression is an extension of (simple) linear regression, which allows a response variable, *y*, to be modeled as a linear function of two or more predictor variables [11].

Often the relationship between x and y cannot be approximated with a straight line. In this case, a nonlinear regression technique may be used. Alternatively, the data could be preprocessed to make the relationship linear. Nonlinear regression models define y as a function of x using an equation that is more complicated than the linear regression equation. The term multivariate nonlinear regression refers to nonlinear regression with two or more predictors ($x_1$, $x_2$, …,$x_n$). When multiple predictors are used, the nonlinear relationship cannot be visualized in two-dimensional space.

Unfortunately, many real-world problems are not simply prediction. For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural networks) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

### *3.3. Time Series*

A time series is a collection of observations made sequentially through time. At each time point one or more measurements may be monitored corresponding to one or more attributes under consideration. The resulting time series is called univariate or multivariate respectively. In many cases the term sequence is used in order to refer to a time series, although some authors refer to this term only when the corresponding values are non-numerical.

The most common tasks of TSDM methods are: indexing, clustering, classification, novelty detection, motif discovery and rule discovery In most of the cases, forecasting is based on the outcomes of the other tasks. A brief description of each task is given below. [12].
Indexing: Find the most similar time series in a database to a given query time series.
Clustering: Find groups of time series in a database such that, time series of the same group are similar to each other whereas time series from different groups are dissimilar to each other.
Classification: Assign a given time series to a predefined group in a way that is more similar to other time series of the same group than it is to time series from other groups.
Novelty detection: Find all sections of a time series that contain a different behavior than the expected with respect to some base model.
Motif discovery: Detect previously unknown repeated patterns in a time series database.
Rule discovery: Infer rules from one or more time series describing the most possible behavior that they might present at a specific time point (or interval).
This method of DM, unveils numerous facets of complexity. The most prominent problems arise from the high dimensionality of time-series data and the difficulty of defining a form of similarity measure based on human perception.

### *3.4. Anomaly Detection*

Data object is considered to be an outlier if it has significant deviation from the regular pattern of the common data behavior in a specific domain. Generally it means that this data object is "dissimilar" to the other observations in the dataset. It is very important to detect these objects during the data analysis to treat them differently from the other data. Anomaly Detection is the process of finding outlying record from a given data set. This problem has been of increasing importance due to the increase in the size of data and the need to efficiently extract those outlying records which could indicate unauthorized access of the system, credit card theft or the diagnosis of a disease.

Anomalies can be classified into either point anomalies contextual anomalies or collective anomalies. The earlier is when single data records deviate from the remainder of the data sets. This is the simplest kind and the one which is most addressed by the existing algorithms. Contextual anomalies is when the record has behavioral as well as contextual attributes. The same behavioral attributes could be considered normal in a giving context and anomalous in another. Whilst the collective anomalies is when a group of similar data are deviating from the remainder of the data set. This can only occur in data sets where the records are related to each other. Contextual anomalies can be converted into point anomalies by aggregating over the context. The algorithms implemented in the extension all explicitly handle point anomalies [10].

According to the anomaly detection survey [9] the techniques can be grouped into one of the following main categories *classification based*, *nearest-neighbor*based *,clustering based* and *statistical based*. Classification based algorithms are mainly supervised algorithms that assumes that the distinction between anomalous and normal instances can be modeled for a particular feature space. Nearest-neighbor based algorithms assume that anomalies lie in sparse neighborhoods and that they are distant from their nearest neighbors. They are mainly unsupervised algorithms. Clustering based algorithms work by grouping similar objects into clusters and assume that anomalies either do not belong to any cluster, or that they are distant from their cluster centers or that they belong to small and sparse clusters. Statistical approaches label objects as anomalies if they deviate from the assumed stochastic model.Anomaly Detection methods are widely used for fraud or suspicious transtaction detection in finanacial organizations.

## 4. DATA MINING TOOLS IN BI SYSTEMS

BI(Business Intelligence) refers to a set of tools used for multidimensional data analysis, with the main purpose to facilitate decision making. One of the main components of BI systems is OLAP(Online Analytical Processing). The main OLAP component is the data cube which is a multidimensional database model that with various techniques has accomplished an incredible speed-up of analyzing and processing large data sets. In our last paper [13]. we studied OLAP and compared different implementations of it such as ROLAP,MOLAP,HOLAP in terms of performance and data accuracy. Inoursimulationswecompared twotechnologies: ROLAPthatperformsqueryagainstDWandHOLAPwhich usesintelligentcubes . It was highlighted the efficiency of these intelligent cubes that reduced drastically the response time of the system. They also use a very good compression and the space occupied in memory is small. As the cubes are stored on the OLAP server, which means that will take reports even if the server where the database is hosted is down.

Generally the usage of intelligent cubes when databases are large increases the efficiency, the performance and allows to have reports at any time even with the disadvantage of a memory occupied larger.

Another fundamental advantage of OLAP tools is that the user gets a multidimensional information and the reporting is flexible.

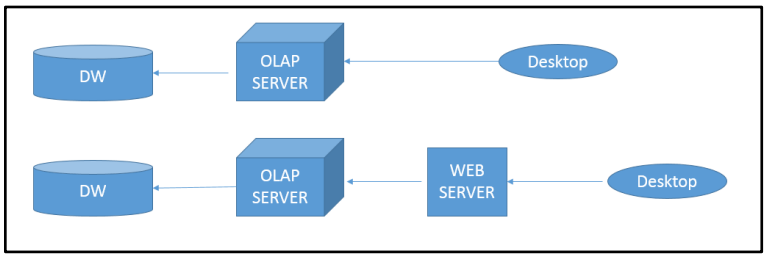


*Figure 2. OLAP architecturewiththreeandfourlevels*

We concluded that OLAP is powerful tool for data extraction in BI systems and has a very good time efficiency in queries against large databases (data warehouses). OLAP is very flexible with columns and rows, and it is possible to report in several dimensions. Although small organizations with a limited database might not need all the capacity of OLAP tools.

OLAP applications are widely used by Data Mining techniques. In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema).The star schemas in data warehouses increases the performance due to the small number of connections that have to do to get a report.Several works in the past proved the likelihood and interest of integrating OLAP with data mining and as a result a new promising direction of Online Analytical Mining (OLAM) has emerged. The term OLAM was frstly introduced by Han in [14].Issues for On-Line Analytical Mining of Data Warehouses were analyzed by HAN,Chee and Chinag in[15].The purpose of integrating OLAP with data mining is because of the high quality of data in data warehouses,available information processing infrastructure surrounding data warehouses,OLAP-based exploratory data analysis and on-line selection of data mining functions..An architecture that integrated OLAP and OLAM was proposed in this paper.

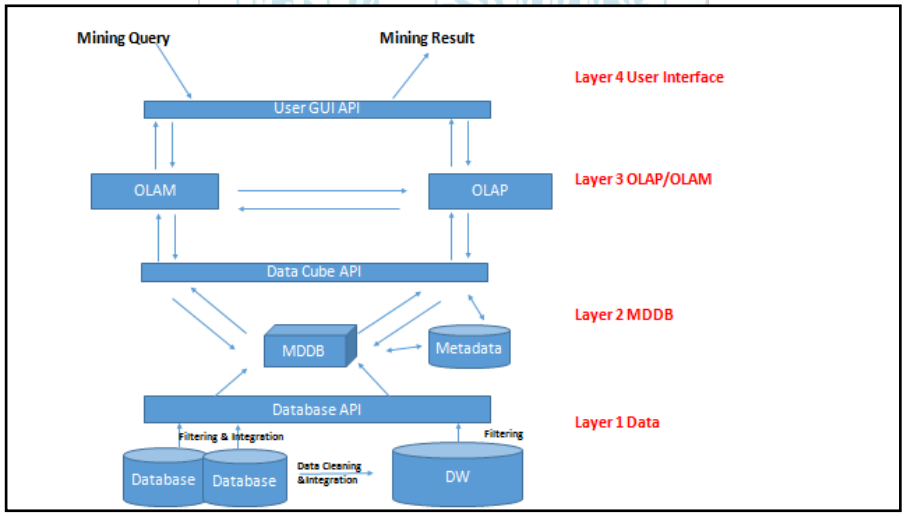


*Figure 3. OLAM Architecture*

Hua [16] proposed and developed an interesting association rule mining approach called Online Analytical Mining of association rules. It integrated OLAP technology with association rule mining methods and leads to flexible multidimensional and multi-level association rule mining. Dzeroski et al. [17] combined OLAP and Data Mining in a different way to discover patterns in a database of patients. Two data mining techniques, clustering and decision tree induction were used. Clustering was used to group patients according to the overall presence/absence of deletions at the tested markers. Decision trees and OLAP were used to inspect the resulting clustering and to look for correlations between deletion patterns, populations and the clinical picture of infertility. Dehne et al. [18] studied the applicability of coarse grained parallel computing model (CGM) to OLAP for data mining. Authors presented a general framework for the CGM which allows for the efficient parallelization of the existing data cube construction algorithm for OLAP. Experimental data showed that this approach yield optimal speed up even when run on a simple processor cluster via a standard switch. The study shows that OLAP and data mining, if combined together, can produce greater benefits in a number of diverse

research areas. Usman and Asghar in [19] combined enhanced OLAP and data mining techniques but their main focus was on a particular data mining technique known as Hierarchical Clustering. Furthermore, they used data mining as a pre-processing step to get better understanding of data before passing it to the automatic schema builder and which then generates schema for OLAP engine. They proposed an integrated OLAM architecture which integrates enhanced OLAP with data mining and to provide automatic schema generation from the mined data. This proposed architecture improved the performance of OLAP and added extra intelligence to the OLAP system.Experimental results proved that the proposed architecture improved the cube construction time, empowered interactive data visualization, automated schema generation, and enabled targeted and focused analysis at the front-end.

Although there is little research in this area recently and integration of other different techniques of data-mining with OLAP needs to be done. One of the largest computer technologies companies Oracle has integrated OLAP and data mining capabilities directly into the database server. Oracle OLAP and Oracle Data Mining (ODM) are options to the Oracle Database.

## 5. CONCLUSION

Data mining offers numerous ways to uncover hidden patterns within large amounts of data. These hidden patterns can potentially be used to predict future behavior. Good data is the first requirement for good data exploration. There are various techniques and algorithms that can be used to perform data-mining but their use depends on the application. Predictive data mining techniques are appropriate when you have a specific target value you'd like to predict about your data. Predictive analytics can be used for forecasting customer behavior and purchasing patterns to identifying trends in sales activities. On the other hand descriptive data mining does not focus on predetermined attributes, nor does it predict a target value. Rather, descriptive data mining finds hidden structure and relation among data. Descriptive analytics are useful because they allow us to learn from past behaviors, and understand how they might influence future outcomes. In this a paper a survey of the most important data mining techniques and algorithms for both models was made.

Also a review of the existing work in combining different techniques of Data Mining with OLAP was made and the advantages were mentioned.Combining OLAP and data mining techniques can provide a very effective way for extracting hidden or useful information in large datasets .This combination gives us an intelligent system improved in performance with data mining capabilities. We conclude that there is little research in this area recently and integration of other different techniques of data-mining with OLAP needs to be done.Our future work consists in combining and testing various data mining techniques in combination with OLAP in databases and comparing their efficiency in terms of data retrieval speed and data quality.

## REFERENCES

[1] FransCoenen ,Data Mining: Past, Present and Future,*The Knowledge Engineering Review, 2004, Cambridge University Press*

[2]Pradnya P. Sondwale, Overview of Predictive and Descriptive Data Mining Techniques,International Journal of Advanced Research in Computer Science and Software Engineering,April 2015

[3] Mihika Shah,Sindhu Nair ,A Survey of Data Mining Clustering Algorithms,International Journal of Computer Applications (0975 – 8887) Volume 128 – No.1, October 2015

[4] Irina Tudor, Association Rule Mining as a Data Mining Technique, Petroleum-Gas University of Ploieşti,Buletin Vol. LX No. 1/2008

[5] Trupti A. Kumbhare et al An Overview of Association Rule Mining Algorithms, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 927-930

[6] N. Elavarasan, Dr. K.Mani,A Survey on Feature Extraction Techniques, International Journal of Innovative Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2015

[7]FabricioVoznika Leonardo Viana"DATA MINING CLASSIFICATION" Springer, 2001

[8] A.Shameem Fathima,D.Manimegalai,Nisar Hundewale ,A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, November 2011

[9]Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly Detection: A Survey. Technical report, University of Minnesota, 2007.

[10]Victoria Hodge and Jim Austin. A survey of outlier detection methodologies. Artificial Intelligence Review, 22:85126, 2004

[11] Han - Data Mining Concepts and Techniques 3rd Edition - 2012.pdf

[12]Shanta Rangaswamy,Time Series Data Mining Tool,International Journal of Research in Computer and Communication Technology, Vol 2, Issue 10, October- 2013

[13] Zanaj,Lico A multidimensional analyses in Business Intelligence systemsIJCSIS May 2012

[14]J. Han, "Towards online analytical mining in large databases," ACM SIGMOD Record, vol. 27, no. 1, pp.97-107, March 1998.

[15] J. Han, S. H. S. Chee and J. Y. Chiang, "Issues for online analytical mining of data warehouses," in Proc. Of the SIGMOND Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD), Seattle, 1998, pp. 2:1-2:5.

[16]H. Zhu, "Online analytical mining of association rules," Master Thesis, Simon Fraser University, 1998, pp. 1-117.

[17].S. Dzeroski, D. Hristovski and B. Peterlin, "Using data mining and OLAP to discover patterns in a database of patients with Y chromosome deletions,"inProc. AMIA Symp., 2000, pp. 215–219.

[18]F. Dehne, T. Eavis and A. Rau-Chaplin, "Coarse grained parallel on-line analytical processing (OLAP) for data mining, in Proc. of the Int'l Conf. on Computational Science (ICCS), 2001, 589-598.
[19]Usman ,Asghar,An Architecture for Integrated Online Analytical Mining,JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 3, NO. 2, MAY 2011
[20] Han - Data Mining Concepts and Techniques 2rd Edition - 2006.pdf
[21]NGUYEN Feature Extraction for Outlier Detection in High-Dimensional Spaces, JMLR: Workshop and Conference Proceedings 10: 66-75 The Fourth Workshop on Feature Selection in Data Mining

## BIOGRAPHY

LedionLiço has a Scientific Master Degree in Electronic Engineering from the Polytechnic University of Tirana(UPT) , Faculty of Information Technology (2011) .He is currently following the PHD program near UPT. He is working as Head of IT Division near Gener2 company and he is also involved as a part-time lecturer in the UPT University. Tirana, Albania.