

Controlling Health Admissions By Using Data Mining Methods

¹D.KUMAR, ²GOPU NAGARJUNA REDDY

Assistant Professor, M.Tech Student

Dr.Samuel George Institute of Engineering & Technology, Markapur, India,AP

Abstract: Database mining is defined as the process of mining for implement, formerly unidentified, and potentially essential data from usefully huge databases by efficient knowledge diction techniques. Owing to the great advantages various organizations are using data mining technology is a vital part for everyone. We propose new predicting the patient length of stay (LOS) data mining techniques. The used data was collected from the pediatric emergency department (PED) in Lille regional hospital centre. Due to the increased stress in day-to-day life the growth of demand of insurance increased. Data mining helps insurance firms to discovery useful patterns from the customer database. We find on key online privacy and security issues and concerns, the role of self-regulation and the user on privacy and security protections, data protection laws regulatory trends, and the outlook for privacy and security legislation. The purpose of the paper aims to present how data mining is useful in the insurance industry, how its techniques produce good results in insurance sector and how data mining enhance in decision making using insurance data. We will take many data mining techniques is used in security information and event management system to enhance the capabilities of the system.

Index Terms: Data mining, security information event management system, Machine Learning, prediction, length of stay

1. INTRODUCTION

Security data and event management system is specific term in computer security referring to the collection of data typically graphical files from many sources into a central repository for analysis. [1]. The objective is to propose a Decision Support System (DSS) based on a data mining approaches, in order to support operational and tactical decisions [2]. Data mining, a knowledge discovery in databases is a process of analyzing data from different perspectives and summarizing to useful information [3]. The aim to see the data mining techniques and particularly classification methods to develop models for prediction of patient length of stay (LOS) at the emergency department [4]. The most scenario of any insurance firm, effective management of customer data is essential one. With the help of data mining techniques, the customer data handled effectively [5]. Knowledge driven methods operate on knowledge repositories that include scientific literature, published clinical trial results, medical journals, textbooks, as well as clinical practice guidelines [6]. The key objectives of the IRDA include promotion of competition so as to enhance customer satisfaction and ensure the financial security of the insurance market. Various types of insurance such as Life insurance, Property insurance, health insurance, vehicle insurance and other insurance used the data mining models depends upon their requirements [7]. Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful data from databases. Knowledge discovery is needed to make sense and use of data.

Though data mining and knowledge discovery in databases is frequently treated data mining is actually part of the knowledge discovery process [8].

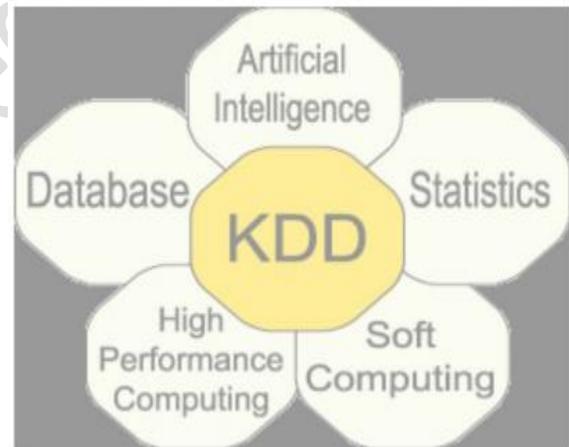


Figure 1: KDD Environment.

2. RELATED WORK

The new useful reaction is possible, understandable-leading to insight many step and process. [9]. In contrast for supervised learning a model is built prior to the analysis. We then apply the algorithm to the data in order to find the parameters of the model. The objective of building models using supervised learning is to predict an outcome or category of interest [10]. Evolution of database technology, data collection, database creation, IMS and Network DBMS, relational data model, Relational DBMS, advance database Models object oriented database, data collection centre, warehousing, multimedia database and recent web database needs to process the approach of data mining. [11]. Used artificial

neural network ensembles to predict the disposition and length of stay in children presenting. The results show that the neural network ensembles correctly predicted disposition in 81% of test cases [12]. This algorithm is compared with various classification algorithms such as Neural Network, Logistic Regression, Random Forest and Decision Tree, SVM, were used to compare the proposed algorithm. Different parameters such as Precision, Accuracy, Recall and Area under the Curve (AUC) were to measure the performance of the proposed algorithm [13]. We need information but what we have is a huge amount of data flooding around. Because of the amount of data is so enormous that human cannot process it fast enough to get the information out of it at the right time, the data mining technology has been established to solve this problem potential [14].

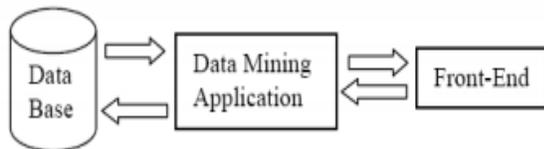


Fig. 2. Data mining architecture

3. SYSTEM ARCHITECTURE

This knowledge contributes a lot of benefits to business strategies, scientific, medical research, governments and individual. The architecture contains new secure communication, database connectivity, organized data management and efficient data analysis for generating global mining model [15]. It uses a combination of mathematical and computational techniques to aid description and classification, and to extract knowledge of data set [16]. It used for identification of similar classes of objects. It's used for grouping based on the customer's behavior. It is applicable for customer segmentation and targeted marketing. The algorithm splits the data at each node based on the variable that separates the data unless an optimal model is not obtained [17].

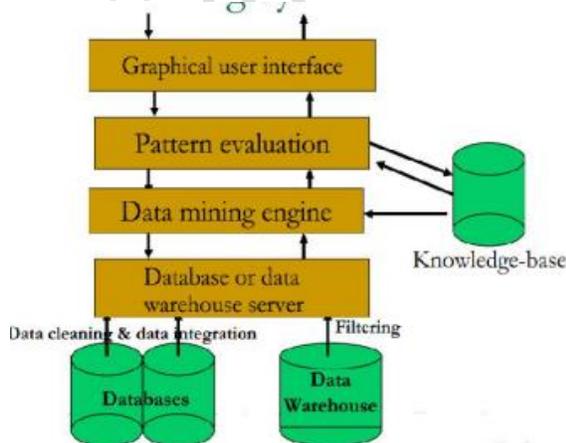


Figure 3: New Architecture of the Proposed System

4. METHODOLOGY

One of the security key models in data mining technology in business or technological one but a social one. It is the issue of individual privacy [18]. There is a wealth of knowledge to be gained from computerized health records. the overwhelming bulk of data stored in these databases makes it extremely difficult for humans to sift through it and discover knowledge [19]. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Mining Methodology and User Interaction

1. Mining different kinds of knowledge in database
2. Interactive mining of knowledge at multiple levels of
3. Abstraction Incorporation of background knowledge
4. Data mining query language and ad-hoc data mining
5. Expression and visualization of data mining results
6. Handling noise and incomplete data
7. Pattern evaluation

A. Logistic Model Tree Algorithm

Although many interesting real world domains demand for regression tools, machine learning researchers have traditionally concentrated their efforts on classification problems. Classification algorithms only deal with nominal variable and cannot handle ones measured on a numeric scale. To use them numeric attributes must first be "discretized" into smaller number of distinct ranges [20]. We decide to take into account the intuitive staff's experience in order to estimate the intervals more realistic than the values obtained by Weak tools for dies cartelization.

1. Logistic Model Tree (LMT)
2. Multi-class alternating decision tree using the Logit Boost strategy (LAD Tree)
3. Decision tree (C4.5 - J48),
4. Decision tree with naive Bays classifiers at the leaves (NB Tree)
5. Random Forest (RF),
6. Decision/regression tree using information gain/variance and prunes it using reduced-error pruning (REP Tree)
7. Multilayer Perception (MP),
8. SVM (SMO).

For these eight methods, the first step corresponds the target variable LOS in order to apply some of these

methods. The model valuation is carried out using 10-fold cross validation on the PED data set [21].

5. DATA MINING TASK

The data mining tasks are different types depending on the use of data mining result the data mining tasks are classified [22].

A. Exploratory Data Analysis In the repositories vast amount of information's are available .This data mining task will serve the two purposes (i)Without the knowledge for what the customer is searching, then (ii) It analyze the data these techniques are interactive and visual to the customer.

B. Descriptive Modeling It describe all the data, it includes models for overall probability distribution of the data, partitioning of the dimensional space into groups and models describing the relationships between the variables.

C. Predictive Modeling This model permits the value of one variable to be predicted from the known values of other variables.

D. Discovering Patterns and Rules. This task is primarily used to find the hidden pattern as well as to discover the pattern in the cluster. In a cluster a number of patterns of different size and clusters are available ..

E. Retrieval by Content The primary objective of this task is to find the data sets of frequently used in the for audio/video as well as images It is finding pattern similar to the pattern of interest in the data set.

5.1 Techniques for Continue Target Value

Only regression allows us to use continue variable of the target variable. Nine methods available are used [23].:

1. **LR:** Linear regression
2. **DS (Decision Stump):** Class for building and using a decision stump. Usually used in conjunction with a boosting algorithm regression classification
3. **M5P:** Induction of model trees for predicting continuous classes
4. **REP Tree:** Builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with backfitting).
5. **SM Oreg:** SVM Regression
6. **IRM:** Isotonic regression model
7. **MLP:** MultiLayer perceptron
8. **PRLM:** Pace regression linear models
9. **K Star:** K-nearest neighbor's classifier (lazy)

Classification models are not very useful when the target values are continuous. We also observe that independently of the used algorithms, we obtain quasi the same performances.

6. RESULTS AND DISCUSSION

For the evaluation of the methods accuracy, kappa, sensitivity and specificity this performance metrics are used. As shown in table, the Random forest

performs best across all performance measures. A small difference is observed the remaining two methods decision tree and gradient boosted machine. For admission of any patient through the emergency department the triage process plays an important role. It is based on the partition of the original sample into ten subsamples, using nine as training data and one for testing data. As shown by results, this study can be benefit to PED manager to predict LOS and detect the beginning of strain situation. This information can be used to make the decision more proactive.

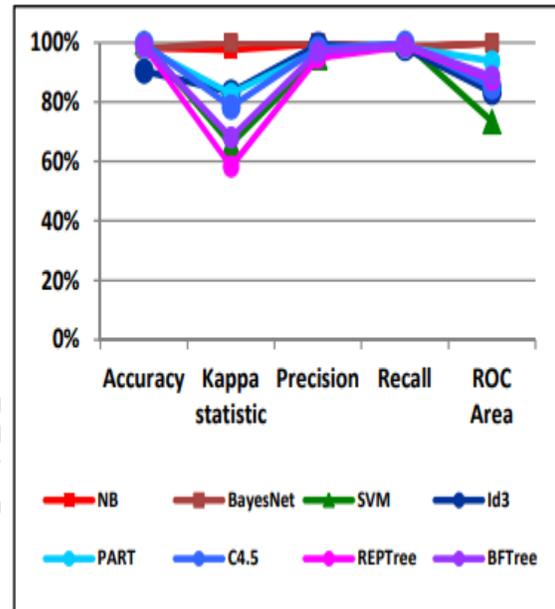


Figure 4: Rate of maximum

7. CONCLUSION AND FUTURE WORK

Data mining is key features of many homeland security initiatives is used as a means for detecting fraud, assessing risk, and product retailing, data mining involves the use of data analysis tools to find previously unknown, valid patterns and relationships in large data sets. The different algorithm of data mining used to extract the patterns and thus the knowledge from this variety databases. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain. .It highlighted the importance and role of data mining techniques are useful in insurance sector for managing the customer data and gain business advantage. The benefit not only includes prediction of medical condition using the previous history of a patient from the database but also hospital management systems such as emergency division. The future of the insurance lies in the increasing the product plans and improving security levels using advanced data mining algorithms.

8. REFERENCES

- [1] Han and Kamber, "Data Mining: Concepts and Techniques", Second Morgan Kaufman Publisher, 2006,
- [2] Micheal J. A .Berry and Gordon S.linoff. EBook "Data Mining Technique for marketing Sales and CRM", Wiley Publishing, Inc., Indianapolis. Indiana, 2004.
- [3] Lijia Guo. "Applying Data mining Techniques in Property/Casualty Insurance", Casualty Actuarial Society Forum Casualty Actuarial Society - Arlington, Virginia W inter 2003.
- [4] Katharina Morik and Hanna Kopcke. "Analyzing Customer Churn in Insurance Data". Lecture Notes in Computer Science, vol. 3202, 2004, pp. 325-336 .
- [5] Keyvan Vahidy Rodpysh,, Amir Aghai and Meysam Majdi , " Applying data mining techniques for Customer Relationship management", International Journal of Information Technology, Control and Automation Vol.2, No.3, July 2012
- [6] A. B. Devale and Dr. R. V. Kulkarni, "Application of data mining techniques in life insurance", International journal of Data mining and Knowledge management Process", Vol.2, No.12.,pp. 31-40. July 2012
- [7] S. Balaji and Dr. K. Srinivasta, "Naïve Bayes Classification Approach for Mining Life Insurance Databases for Effective Prediction of Customer Preferences over Life Insurance Products" , International journal of Computer Applications, vol.51, No.3, pp.22-26. August 2012
- [8] Kanwal garg, Dharminder kumar and m.c.garg, "Data mining techniques for identifying the customer behavior of investment in life insurance sector in india", International journal of information technology and knowledge management, Vol.1, pp.51-56.
- [9] Cheng C.H. and Chen Y.S. (2009), "Classifying the segmentation of customer value via RFM model and RS theory", Expert Systems with Applications, Vol.36, pp. 4176-4184.
- [10] Gulshan Vohra and Dr. Bhart Bhushan " Data mining techniques for analyzing the investment behavior of customers in life insurance sector", International Journal of Research in Engineering & Applied Sciences, Vo 2, Issue 1 January 2012
- [11] D. Tomar and S. Agarwal, "A Survey on Data Mining Approaches for Healthcare," International Journal of Bio-Science and BioTechnology, vol. 5, no. 5, pp. 241- 266, 2013.
- [12] N. P. Waghulde, N. P. Patil, "Genetic Neural Approach for Heart Disease Prediction," International Journal of Advanced Computer Research, vol. 4, no. 3, pp. 778-784, 2014
- [13] Zhang, Xingyu and Kim, Joyce and Patzer, Rachel E and Pitts, Stephen R and Patzer, Aaron and

Schrager, Justin D, "Prediction of emergency department hospital admission based on natural language processing and neural networks," Methods of information in medicine., vol.56, pp. 377– 389, 2017.

[14] Boukenze, Basma and Mousannif, Hajar and Haqiq, Abdelkrim,"Predictive analytics in healthcare system using data mining techniques," Computer Sci Inf Technol., vol. 1, pp. 1–9, 2016.

[15] Dinh et al., "The Sydney Triage to Admission Risk Tool (START) to predict Emergency Department Disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia," BMC emergency medicine., vol.16, pp. 46, 2016.

[16] Lin, T. Y., "Anomaly Detection -- A Soft Computing Approach", Proceedings in the ACM SIGSAC New Security Paradigm Workshop, Aug 3-5, 1994,44-53.,1994

[17] Scott W. Ambler, "Challenges with legacy data: Knowing your data enemy is the first step in overcoming it", Practice Leader, Agile Development, Rational Methods Group, IBM, 01 Jul 2001.

[18] Agrawal, R, and R. Srikant, "Privacy-preserving Data Mining," Proceedings of the ACM SIGMOD Conference, Dallas, TX, May 2000.

Guide details:

D.Kumar received B.Tech (CSIT) Degree from JNT University in 2006 and M.Tech (CSE) Degree from JNTUK Kakinada in 2011. He has 11 years of teaching experience. He joined as Assistant Professor in Dr.Samuel George Institute of Engineering & Technology, Markapur, India in 2006. Presently he is working as Associate Professor in CSE Dept. His Interested research areas are Image Processing and Computer Networks. He attended Various National and International Workshops and Conferences.



Student details:

Gopu Nagarjuna Reddy

Email:nagarjunareddy197@gmail.com

Dr.Samuel George Institute of Engineering & Technology, Markapur, India,AP