International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 09 , September 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

A STUDY OF PREDICTIVE ANALYTICS AND ITS MODELING TECHNIQUES

Dr.(Mrs). Ananthi Sheshasaayee[1], K.Bhargavi[2]

[1]Research Supervisor, PG and Research Department of Computer Science, Quaid E Millath Government College for Women, Chennai-600 002, India.

[2]Research Scholar PG and Research Department of Computer Science, Quaid E Millath Government College for Women, Chennai-600 002, India.

ABSTRACT

Predictive analyticsis the enhancement of business intelligence and data discovery which predicts the future using statistical methods on data and it also works bey ond the complexity limits of many OLAP implementations.The Predictive analytics not only answers what is likely to happen next and what to do next, it also predicts how and when to do and when to explain what if scenarios for making better decisions. By implementing the predictive analytics the organizations can gain competitive advantage of predicting the future better in beforehand. Predictive analytics helps to increase the sales by making the organizations to anticipate the customer's needs and purchasing habits which impacts in the reduced inventory cost, increased profit and spotting fraudulent purchases on correct time. By applying combined knowledge of business and statistical methods on the business data, the insights are produced by the predictive analytics which are used by the organizations to understand how people will behave as distributers, buyers, sellers and customers.The insights produced by multiple predictive models are used for making strategic decisions by the senior management. Without the correct tools and techniques, the implementation of predictive analytics gets harder for the organizations. It is important for them to know which technique to use, when to use and on which data. In this paper, the different types of predictive analytics models, different stages in building models, types of algorithms and methodologies used for building models are discussed.

Keywords: Algorithms,Business insights,Models, Predictive Analytics, Statistical Methods.

I.INTRODUCTION

The predictive analytics is the process of handling different types of data and making decisions by applying different statistical methods like modeling, machine learning, data mining and game theory on the data for a corresponding situation. It will first analyze the historical data and then predict the future, the historical data are gathered and converted by different techniques like correlating and filtering. Predictive analytics is used to automatically analyze large amounts of data with different variables; it includes clustering, decision trees, market basket analysis, regression modeling, neural nets, genetic algorithms, text mining, hypothesis testing, decision analytics, and more.

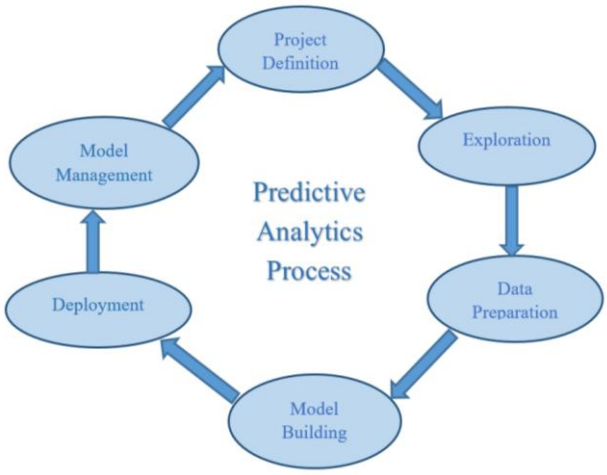
The predictive analytics does the prediction in four different steps like (1).collecting and pre-processing, (2). Transforming the pre-processed data into a format that can be applied on the selected method for predicting the

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

future events. (3). Developing the learning model using the transformed data. (4). Reporting the predictions by using the model that was developed. The predictive analytics avoids the guessing's in the decision making process and applies the scientifically proved methods to take correct decisions in the shortest time.By revealing the hidden patterns and relationships between the data, the predictive analytics delivers the insights that are proactive. The main reason behind the companies that use predictive analytics are to predict the future trends, to understand customer behavior, to develop the business performance and to take strategic decisions.

The insights for strategic decision makings like probing new markets, purchasing, confinement, finding opportunities for up-selling, cross-selling, improving security and fraud detection are developed by multi-related predictive models. The predictive analytics is otherwise called as supervised data mining technique in which data is modelled from historical data to find patterns within the data sets and applied to predict a value using some set of parameters.The data that can be readily used for analysis are structured data like age, gender, marital status, income and sales. The data like textual data in call center notes, social media content and other types of open text that has to be extracted from the text along with sentiment before applying on the model are unstructured data.The type of resultant data concludes whether the regression or classification algorithm used for prediction.The accuracy and the usability of the resulted data greatly depends on the data analysis level and the quality of assumptions.

Fig 1:Predictive Analytics Process



II.TYPES OF PREDICTIVE ANALYTICS MODELS

Even though there are many models used in predictive analytics, there are three main types of models are often used by the organizations. 1. Predictive models. 2. Descriptive models 3. Decision models. These three models are mostly used together in different methods of predictive analytics to take different customer decisions

1. Predictive model: the predictive models will analyse the previous act of a customer to predict the future behaviour. This models are often called up during live transactions by embedding in operational processes to predict the fraudulent transactions, customer behaviours, and the reason behind the customer behaviours. These models counterbalance the relationship between the hundreds of variables and segregates each customer's possible uncertainty, which helps for predicting the action of a customer.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

2. Descriptive model:descriptive models finds different no of relationships among customersand products and classifies customers and products into groups unlike predictive model which predicts single customer behaviour.
3. Decision models. The decision models predicts the results of complex decisions taken by the organizations same like predictive models which predicts customer behaviour. By finding the relationships among all the components like data, decisions and projecting results of a decisions, the decision models will predict the outcome for the action taken. Unlike predictive model, the decision model will consider about industrial drivers and compulsions. Combining the optimization with decision modelling will help to make decision approaches that decides which action should be taken on each customer or transactions to enhance the outcomesscientifically and to meet the described constraints.

III.BUILDING EFFECTIVE MODELS

The modelling process includes many stages and some of them are,

1. Purpose: The objective of the project is described in this stage.
2. Collection of data: This stage includes the collection of data from various resources respecting the project.Pre-processing the data: The variables, tokens, and other terms which are going to be used in project are explained and it is also explained that where and when these variables can be used. The variables are filtered according to the requirement.
3. Segregating the data: the data is segregated into a training sets and a validation sets in order to develop the model and to check how the model works. This stage is the part supervised learning process in prediction problem and classification problems, which can be used to develop other models and also the value of the resulted variables can be used in various places. The multiple models are tested using different types of settings on each model by the data mining techniques. By testing the models, the performance of the models on particular data is found and thishelps to choose the best performing model.
4. Selecting the technique: Various techniques are selected for creating the models by using the data which is partitioned into training and validation sets.
5. Finding fitted and predicted value: The fitted values are found by applying algorithm on training data sets and the predicted values are found by applying algorithm on validation data sets.
6. Depicting the results: By using prediction algorithms with different settings the best model is chosen based on the lowest errors on validation data.
7. Deployment of the model: the best model chosen is applied on new data to predict the future.

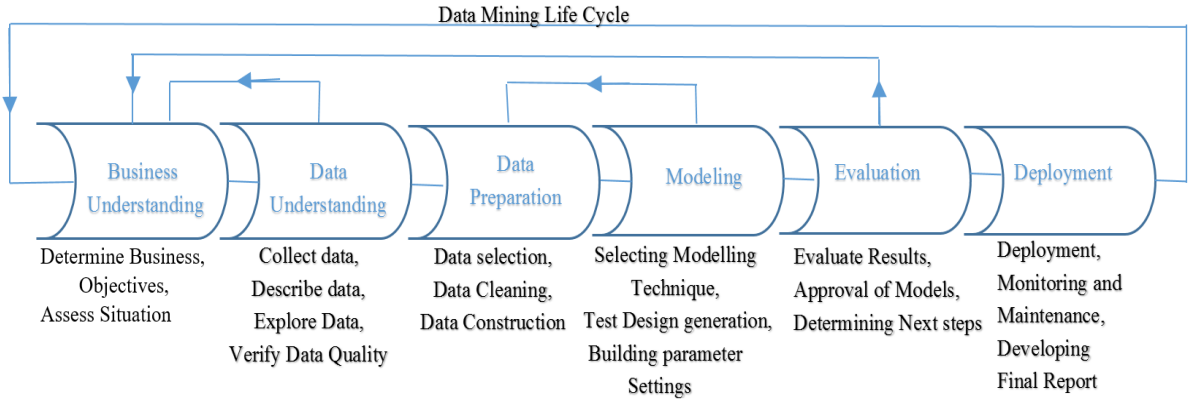
IV.METHODOLOGIES USED FOR BULDING MODELS

Even though the predictive analytics requires great skill sets, experience and creativity to develop a predictive model, a proper methodology is needed to help the process. Many methodologies are proposed to build models based on industrial engineering frameworks or business improvement concepts. Out of these methodologies the CRISP, DMAIC, SEMMA provides the basic outline which must be followed to build the effective and efficient models.

International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 09 , September 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

1. GENERIC CRISP-DM Reference Model: the CRISP methodology contains the phases like business understanding, data understanding, and data preparation, modeling of predictors, evaluation and deployment of models on new data. It provides the complete processing models and data mining methodology along with a complete plan for handling the data mining projects.

Fig 2: CRISP-DM



i. Business understanding: in this phase the objectives and the requirements from the business perspective is focused and understood.

ii. Data understanding: in this phase, the collection and pre-processing of the data is done in order to assimilate and analyze the data.

iii. Data Preparation: the final data set which has to be applied on the models are built from the raw data using iterations which includes the description, selection, cleaning , construction, integration and transformation of data.

iv. Selection of modelingtechniques: The selection of modeling techniques, test design generation, building and validation of models are done in this phase.

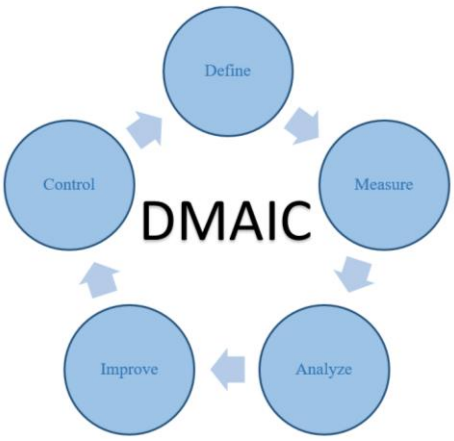
v. Evaluation: In this Phase, the developed models are evaluated thoroughly before deploying. The evaluation process is done keenly and minutely, so that the model will achieve the objective of the business requirement. The main goal of this phase is to confirm that all the business issues are considered properly.

vi. Interpretation of models: The last phase of CRISP_DM is the deployment of the models developed. The information that has to be used are presented in a way so that the client can understand and use it in his business to gain the benefits. The Report generation and documentation are also done in the deployment phase.

2. DMAIC (Define-Measure-Analyze-Improve-Control): The DMAIC methodology is a six sigma methodology to get rid of defects, waste, quality control problems in all business activities. The DMAIC methodology mostly uses step-wise sequential approach, even though some are iterative approach. The DMAIC methodology contains five steps that are.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Fig 3: DMAIC



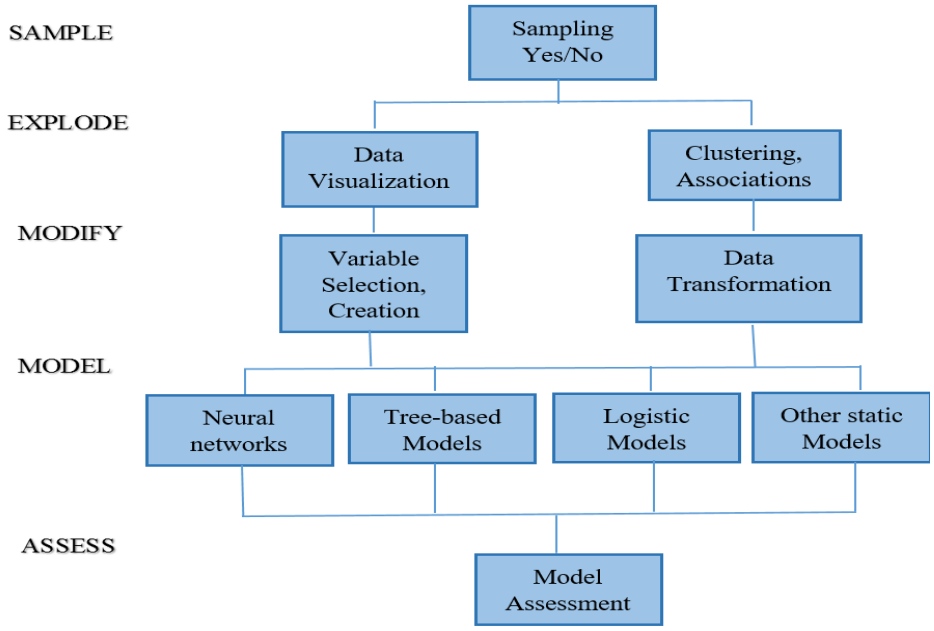
i. Define: In this first step the problem, scope, the requirements of the customers, goals and the timeframe to complete the process are understood and defined.
ii. Measure: In this step, the performance measurement of the existing process is done.
iii. Analysis: The analysis for the requirement of the action to be taken is done in this step by using both statistical and qualitative approaches.
iv. Improve: After the analysis phase, the solution is found and tested for the issues and prioritized according to the requirements of the customers.
v. Control: The last stage of the DMAIC is controlling the product performance.

3. SEMMA (Sample-Explore-Modify-Model-Assess): The SEMMA methodology is also similar to six sigma methodology and proposed by SAS institute. SEMMA consists of five phases.

i. Sample: the sampling of the data set is mainly concerned in this phase. The small amount of data is extracted from the large data pool to extract important information.
ii. Explore: in this phase the analysis of data is mainly focused and finding the unexpected trends and deviations are done in order to understand the ideas.
iii. Modify:In this phase the modification of data is done by creation, selection and transformation, reduction of variables and the outliers are checked.
iv. Modeling of data: In this phase many modeling techniques are used which has its own functionality depending upon the conditions and the problems. The software used for automatic searching finds the combination of data which can predict the outcome exactly.
v. Assess: The final phase of SEMMA includes the calculation of accuracy, efficiency and performance and the reliability of the models.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Fig 4: SEMMA



V. MODELING ALGORITHMS

In the modeling of the predictors, the algorithms are used to find which model can be used for particular data to get the appropriate outcome. There are four main types of modelling algorithms that are used.

1). Classification: the classification is a data mining algorithm which is used to predict the value of a target or class variable by developing a model in which numerical variables can be used.

2). Regression: the regression algorithm is used to predict the value of the numerical variable by building a models with one or more predictors.

3). Clustering: the clustering algorithm is used to cluster the similar set of data into groups. The data in each group will be nearly similar.

4). Association rules: the association rule is used to find all the item sets that has greater support and uses the large item sets to develop the desired rules that has greater confidence.

VI.CONCLUSION

Predictive analytics is the process of handling different types of data by developing predictive models using different methodologies and algorithms to predict the future actions for the corresponding situation. In this paper, the different types of predictive analytics models, steps involved in developing models and the main methodologies and algorithms used for developing predictive models are discussed.The main three methodologies for developing predictive models are CRISP-DM, DMAIC and SEMMA and all these three methodologies are iterative. Since the SAS SEMMA was developed with a specific data mining tool like SAS Enterprise miner, the initial planning phase is given less attention and the Deployment phase is entirely ignored. But the CRISP-DM methodology pays more attention on initial phase and has deployment phase.

The DMAIC methodology same like SEMMA can be applied only on a specific tool, it is a six sigma process which delivers good results among the other process improvement techniques in six sigma. The DMAIC

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

methodology helps the business handle problems from the starting to the end while producing good results and it follows logical approach. The DMAIC provides the organizations with accurate baselines and helps the organizations to improve in different areas and to find solutions to the complex problems. Even though there are some dissimilarities between all these three methodologies, all the three process are cyclical. From the above points it can be concluded that the models that are iterative are the best predictive models that can be applied on predictive analytic process.

REFERENCE:

1. Morelli, Theresa, Colin Shearer, and Axel Buecker. "IBM SPSS predictive analytics: Optimizing decisions at the point of impact." *IBM redpaper* 4710 (2010).
2. Analytics, Predictive. "Bringing The Tools To The Data." *An Oracle white paper* (2010).
3. Mishra, Nishchol, and Dr Sanjay Silakari. "Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges." *International Journal of Computer Science and Information Technologies* 3.3 (2012): 4434-4438.
4. Kobielus, James. "The Forrester Wave™: Predictive Analytics And Data Mining Solutions, Q1 2010." *Forrester Research Inc. report* 4 (2010).
5. Siegel, Eric. "Predictive analytics." *Hoboken: Wiley* (2013).
6. Najdenov, Bojan, and Fadi Makhoul. "Predictive Analytics–Examining the Effects on Decision Making in Organizations." (2015).
7. Halper, Fern. "Predictive Analytics for Business Advantage." *TDWI Research* (2014).
8. Gulati, Hina. "Predictive analytics using data mining technique." *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on*. IEEE, 2015.
9. Delen, Dursun, Asil Oztekin, and Haluk Demirkan. "Introduction to Predictive Analytics and Big Data Minitrack." *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. IEEE, 2013.
10. Mishra, Debahuti, et al. "Predictive data mining: Promising future and applications." *Int. J. of Computer and Communication Technology* 2.1 (2010): 20-28.
11. Gualtieri, Mike, et al. "The Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 2013." *Forrester research* (2013).
12. Stewart, Andrew Reed. *Analysis and Prediction of Decision Making with Social Feedback*. Diss. PhD thesis, Princeton University, 2012.
13. Rod Koch CMA, P. M. P. "From business intelligence to predictive analytics." *Strategic Finance* 96.7 (2015): 56.
14. https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=YTW03411USEN
15. https://www01.ibm.com/events/wwe/grp/grp004.nsf/vLookupPDFs/4.%20%20Bill%20Haffey%20Predictive%20Analytics%209.16/$file/4.%20%20Bill%20Haffey%20Predictive%20Analytics%209.16.pdf
16. https://www.cgi.com/sites/default/files/white-papers/Predictive-analytics-white-paper.pdf
17. https://www.ijarcsse.com/docs/papers/Volume_7/6_June2017/V7I6-0305.pdf