3rd National Conference on Green Technology and Science for Sustainable Development
© 2020 IJSRST | Volume 5 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
International Journal of Scientific Research in Science and Technology

A Review on Predictive Analytics

Prof. A. A.Dande*, Ashwini R. Gujar, Shubhangi S. Rajput
Department of Computer Science and Engineering, AEC, Chikhali, Maharashtra, India

ABSTRACT

At its core, predictive analytics hold a variety of statistical techniques including machine learning, predictive modelling and data mining and uses statistics both historical and current to estimate, or 'predict', future outcomes. These results might be behaviors a customer is likely to exhibit or possible changes in the market, for example. Predictive analytics help us to understand possible future events by analyzing the past. The most widely used predictive models are Decision trees are a simple, but powerful form of multiple variable analysis. Regression is one of the most popular methods in statistics. Neural networks Patterned after the operation of neurons in the human brain, neural networks also called artificial neural networks are a variety of deep learning technologies. The algorithms are defined as 'classifiers', identifying which set of classification data belongs to. Other classifiers are Time Series Algorithms, Clustering Algorithms, Outlier Detection Algorithms, Ensemble Models, Factor Analysis, Naïve Bayes, Support vector machines. Typically, an organization's data scientists and IT experts ar e tasked with the development of choosing the right predictive models – or building their own to meet the organization's needs. Today, however, predictive analytics and machine learning is no longer just the domain of mathematicians, statisticians and data scientists, but also that of business analysts and consultants. More and more of a business' employees are using it to develop understanding and improve business operations – but problems arise when employees do not know what model to use, how to place it, or need information right away. Applications of predictive analytics are Banking and Financial Services, Security, Retail.

Keywords : Predictive Analytics, Predictive Models, Classifiers

I. INTRODUCTION

A common misapprehension is that predictive analytics and machine learning are the same thing. This is not the case. Where the two do overlap, however, is predictive modelling – but more on that later.

At its core, predictive analytics encompasses a variety of statistical techniques including machine learning, predictive modelling and data mining and uses statistics both historical and current to guess, or 'predict', future outcomes. These results might be behaviours a customer is likely to exhibit or possible changes in the market, for example. Predictive analytics help us to understand possible future experiencs analysring the past.

Machine learning, on the other hand, is a subfield of computer science that, as per Arthur Samuel's definition from 1959, gives 'computers the ability to learn without being explicitly programmed'. Machine learning evolved from the study of pattern recognition and explores the notion that algorithms

can learn from and make predictions on data. Predictive analytics is an area of statistics that deals with withdraw information from data and using it to predict trends and behavior patterns. The improvement of predictive web analytics calculates statistical probabilities of future events online.

Predictive analytics statistical techniques involve data modeling, machine learning, AI, deep learning algorithms and data mining.. For example, identifying suspects after a crime has been committed, or credit card fraud as it occurs. It is important to note, however, that the perfection and usability of results will depend greatly on the level of data analysis and the quality of assumptions.

Fig 1 : Predictive Analytics Working Models

This distinguishes it from forecasting. For example, "Predictive analytics—Technology that learns from experience (data) to predict the future behavior of individuals in order to drive better decisions." In future industrial systems, the value of predictive analytics will be to estimate and stop potential issues to achieve near-zero break-down and further be unsegregated into prescriptive analytics for decision optimization. Furthermore, the converted data can be used for closed-loop product life cycle development which is the vision of the Industrial Internet Consortium.

WORKING

Predictive analytics is operated by predictive modelling. It's more of an perspective than a process. Predictive analytics and machine learning go concurrently, as predictive models typically include a machine learning algorithm. These models can be instruct over time to respond to new data or values, delivering the results the business needs. Predictive modelling largely extend over with the field of machine learning.

There are two types of predictive models. They are categorizing models, that predict class membership, and Regression models that predict a number. The algorithms execute the data mining and statistical analysis, determining trends and patterns in data. Predictive analytics software solutions will have built in algorithms that can be used to make predictive models. The algorithms are called as 'classifiers', identifying which set of categories data belongs to.

Advanced analytics are used to survey the way in which specific business issues relate to data on past, present, and predict future actions. Advanced analytics include statistical, mathematical, and other algorithmic techniques such as those in the diagram below, and are more complex than the basic analytics used to calculate frequencies, cross- tabs, and query and discribes cubes.

From this advanced analysis produce insight that is used to determine which actions will drive the results. Through measuring unreliability surrounding these concern, predictive analytics allow proactive threat management, serving as a conductor for refining key decision making processes across controlled, iterative testing of possible actions and their likely studied—and unintended—consequences. By establish predictive analytics within front- line functional systems, specific consortium goals relating to rewards increases, cost reductions, process improvements, and competitive authority can be accomplished.

II. PREDICTIVE MODELS

Predictive modeling is a activity that uses data mining and chances to forecast results. Once data has been composed for applicable predictors, a statistical model is construct.

The model may employ a simple linear equation, or it may be a compounded neural network, mapped out by experienced software.

The most widely used predictive models are:

Decision trees:- Decision trees are a easy, but powerful form of multiple changing analysis. They are processed by algorithms that recognize various ways of splitting data into branch-like segments. Decision trees sepration data into subsets based on classification of input variables, helping you to understand someone's path of decisions.

Types of Decision Tree
Types of decision tree is build on the type of target variable we have. It can be of two types:

1. Binary Variable Decision Tree: Decision Tree which has binary target variable then it called as Binary Variable Decision Tree. Example:- In above outline of student problem, where the target variable was "Student will play cricket or not" i.e. YES or NO.
2. Continuous Variable Decision Tree: Decision Tree has continuous target variable then it is called as Continuous Variable Decision Tree. Example:- Let's say we have a problem to guess whether a customer will pay his continuation premium with an insurance company (yes/ no).Here we know that income of customer is a outstanding variable but insurance company does not have income details for all customers. Now, as we know this is an main variable, then we can build a decision tree to predict customer income based on occupation, product and various other variables. In this case, we are predicting values for continuous variable.

Terminology connected to Decision Trees:

Let's look at the basic term used with Decision trees:
Root Node: It represents entire population or sample and this further gets divided into two or more same sets.

Splitting: It is a process of dividing a node into two or more sub- nodes.

Decision Node: When a sub-node splits into further sub- nodes, then it is called decision node.
Leaf/ Terminal Node: Nodes do not split is called Leaf or Terminal node.

Pruning: When we remove sub-nodes of a decision node, this process is called pruning. You can say opposite process of splitting.

Branch / Sub-Tree: A sub section of entire tree is called branch or sub-tree.

Parent and Child Node: A node, which is divided into sub-nodes is called parent node of sub-nodes where as sub- nodes are the child of parent node. These are the terms commonly used for decision trees. As every algorithm has advantages and limitations, some of these discuss below for decision trees.

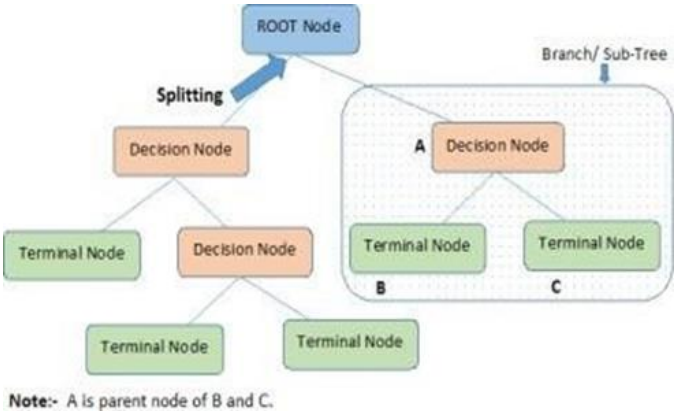

Fig 2 : Decision Tree Mode

ADVANTAGES

1. Easy to Understand: Decision tree results is very easy to understand even for people from nonanalytical background. It does not require any statistical knowledge to read and explain them. Its graphical representation is very intuitive and users can simply relate their hypothesis.

2. Useful in Data examination: Decision tree is one of the fastest way to find most significant variables and relation between two or more variables. With the help of decision trees, we can create new variables / attribute that has better power to predict target variable. For example, we are working on a problem where we have data available in hundreds of variables, there decision tree will help to find most significant variable.

3. Less data cleaning required: It requires less data cleaning compared to some other modeling techniques. It is not influenced by outliers and missing values to a fair degree.

4. Data type is not a obstructiont: It can handle both numerical and unconditional variables.

5. Non Parametric Method: Decision tree is considered to be a non-parametric method. This means that decision trees have no guess about the space issuing and the classifier structure.

DISADVANTAGE

1. Overfit: Over fitting is one of the most actual trouble for decision tree models. This problem gets resolved by use of random forests, which we will discuss some other day.

2. Not fit for continuous variables: While working with continuous numerical variables, decision tree release information when it classify variables in different categories.

REGREESSION

Regression (linear and logistic):- Regression is one of the most desired methods in statistics.

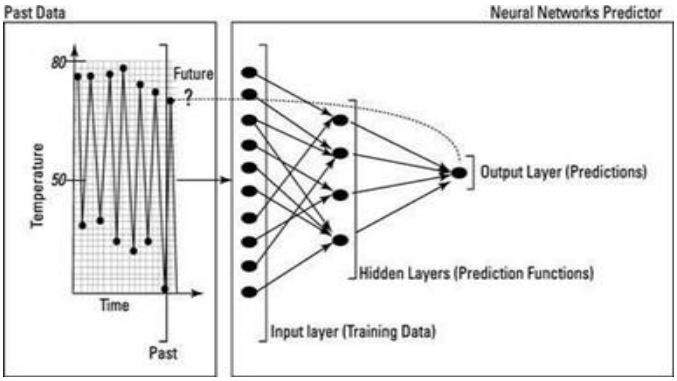

Fig 3 : Neural Network Data Processing

Regression analysis guess relationships between variables, finding key patterns in large and various data sets and how they relate to each other.

Regression analysis is a form of predictive modelling technique which explore the relationship among a dependent and independent variable. This technique is used for predicting, time series modelling and finding the causal outcome relationship between the variables. For example, relationship between rash driving and number of road accidents by a driver is best studied through regression. Regression analysis is main tool for modelling and analyzing data. Here, we fit a curve / line to the data points, in such a way that the differences between the distances of data points from the curve or line is less. Following are some benefits using regression analysis.

1. It specify the remarkable relationships between dependent variable and independent variable.

2. It specify the robustness of clash of multiple independent variables on a dependent variable.

Regression analysis also permit us to compare the effects of variables measured on different scales, such as the effect of price changes and the number of promotional activities.

NEURAL NETWORKS

Neural networks are a diversity of extensive learning technologies. They're typically used to solve compound pattern identification problems and are extremely useful for analyzing large data sets. The neural network is a complex algorithm used for predictive analysis, which is biologically affected by the structure of the human brain. A neural network provides a very simple model in comparison to the human brain. Neural networks can be used to make projection on time series data such as weather data. A neural network can be designed to notice pattern in input data and produce an output free of noise. The neural- network algorithm consist of three layers:

• The input layer feeds past data values into the next (hidden) layer. The black circles represent nodes of the neural network.

• The hidden layer encapsulates several complex functions that create predictors; often those functions are hidden from the user. A set of nodes (black circles) at the hidden layer represents mathematical functions that modify the input data; these functions are called neurons.

• The output layer collects the projections made in the invisable layer and produces the final result: the model's prediction.

III. APPLICATIONS

Predictive analytics is mainly used for security, marketing, operations, risk and fraud detection. Following are some applications that's how predictive analytics and machine learning are utilized in various industries.

1. Banking and Financial Services

In the banking and financial services industry, predictive analytics and machine learning are used in coexistence to detect and decrease fraud, measure market threat, identify opportunities and much, much more.

2. Security

With cybersecurity at the top of every business' agenda in 2017, it should come as no surprise that predictive analytics and machine learning play a key part in security.

3. Retail

Retailers are using predictive analytics and machine learning to better understand consumer behaviour; who buys what and where? These questions can be willingly answered with the right predictive models and data sets, helping retailers to plan onwards and stock items based on seasonality and consumer trends.

IV. CONCLUSION

Predictive analytics is a smart way to add more insight and clarity into your business decisions. Even though it may take a lot of time to collect useful data and devise a plan to sort through it, when you see the results of what it can do, it will be worth it.

Predictive analytics is the practical result of Big Data and Business Intelligence. Big data has revolutionized the way in which companies can now leverage their data and use it to their advantage. It has made it possible for line-managers to use nontransactional data to make strategic decisions. Earlier, only data scientists and statisticians who had the required mathematical skills could understand the predictive analytics technique, but big data has made it easier to capture and store massive volumes of data for faster analysis.

V. REFERENCES

[1]. Predictive analytics and machine learning By Katrina Wakefield,"https://www.sas.com/en_gb/insights /articles/analytics

[2]. /a-guide-to-predictive-analyticsand-machine-learning.html", Last visted [29/09/2019]

[3]. Enabling the Predictive Enterprise with Predictive Analytics By SPSS ,“http://www.spss.com. hk/ predictive_analytics / work.htm”, Lastvisted[23/09/2019]

[4]. Time Series Analysis,https://algorithmia.com/, Last visted[27/09/2019]

[5]. How Predictive Analysis Neural Networks Work by By Anasse Bari, Mohamed Chaouchi, Tommy Jung,https://www.dummies.com/programming/bigdata/data- science/how-predictive-analysis-neural- networks-work/, Last visted[30/09/2019]

[6]. Microsoft Time Series Algorithm By Microsoft, https://docs.microsoft.com/en-us/analysis-services/data- mining/microsoft-time-series-algorithm,Last visted[25/09/2019]

[7]. Getting Started with R By Syncfusion, “https://help.syncfusion.com/predictive-analytics”,Last visted[29/09/2019]

[8]. Predictive Modeling: The Only Guide You'll Need By Resources,“https:// www.microstrategy .com/us/ resources/ introductory-guides/predictive-modeling-theonly-guide-you-need”, Last visted[25/10/2019]

[9]. Decision Tree – Simplified! By Sunil Ray, https://www.analyticsvidhya.com/blog/2015/01/ decisiontree- simplified/, Last visted[28/10/2019]