iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XI **Month of publication:** November 2020

DOI: https://doi.org/10.22214/ijraset.2020.32250

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

An Overview of Predictive Analysis: Techniques and Applications

Shubham Vartak[1], Atharva Sapre[2]
[1, 2]*Computer Engineering, Atharva College of Engineering, Mumbai, India*

Abstract: ***Predictive analytics may be a category of knowledge analytics aimed toward making predictions about future outcomes supported by historical data and analytics techniques like statistical modeling and machine learning. The science of predictive analytics can generate future insights with a big degree of precision.***
Predictive analytics is more accurate and reliable than previous tools. With the help of sophisticated predictive analytics tools and models, any organization can now use past and current data to reliably forecast trends and behaviors.
The future events and behavior of variables are often predicted using the models of predictive analytics. Organizations today use predictive analytics in a virtually endless number of the ways.
The technology helps users in fields as diverse as finance, healthcare, retailing, hospitality, pharmaceuticals, automotive, aerospace and manufacturing.
We will review the process of predictive analysis, its technique, and its applications in this paper.
Keywords: Predictive Analysis, Statistical modelling, Machine learning, Technology, Finance, Healthcare, Retailing, Hospitality, pharmaceuticals, Automotive.

I. INTRODUCTION

Predictive analytics, a branch in the domain of advanced analytics, is used in predicting the future events. It analyzes the current and historical data in order to make predictions about the future by employing the techniques from statistics, data mining, machine learning, and artificial intelligence [1].
With the ever increasing online content generated at an accelerated rate, petabytes of data is produced each day. The explosion in the user generated data from social media and businesses and organizations, there has always been a search for effective management and storage of the enormous data created [2].
It brings together the information technology, business modeling process, and management to make a prediction about the future. Businesses can appropriately use big data for their profit by successfully applying predictive analytics. It can help organizations in becoming proactive, forward looking and anticipating trends or behavior based on the data. It has grown significantly alongside the growth of big data systems [3].
Predictive analytics has not a limited application in e- retailing. It has a wide range of applications in many domains. Insurance companies collect the data of working professionals from a third party and identify which type of working professional would be interested in which type of insurance plan and they approach them to attract towards its products [4].
Banking companies apply predictive analytics models to identify credit card risks and fraudulent customers and become alert from those types of customers.
Organizations involved in financial investments identify the stocks which may give a good return on their investment and they even predict the future performance of stocks based on the past and current performance.
Many other companies are applying predictive models in predicting the sale of their products if they are making such type of investment in manufacturing.
Pharmaceutical companies may identify the medicines which have a lower sale in a particular area and become alert on expiry of those medicines [5].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

II. PROCESS OF PREDICTIVE ANALYSIS

Based on the ongoing and ancient data, data analysts can predict the future using Predictive analytics that call for various steps. The diagram shown below illustrates the working process of predictive analytics.

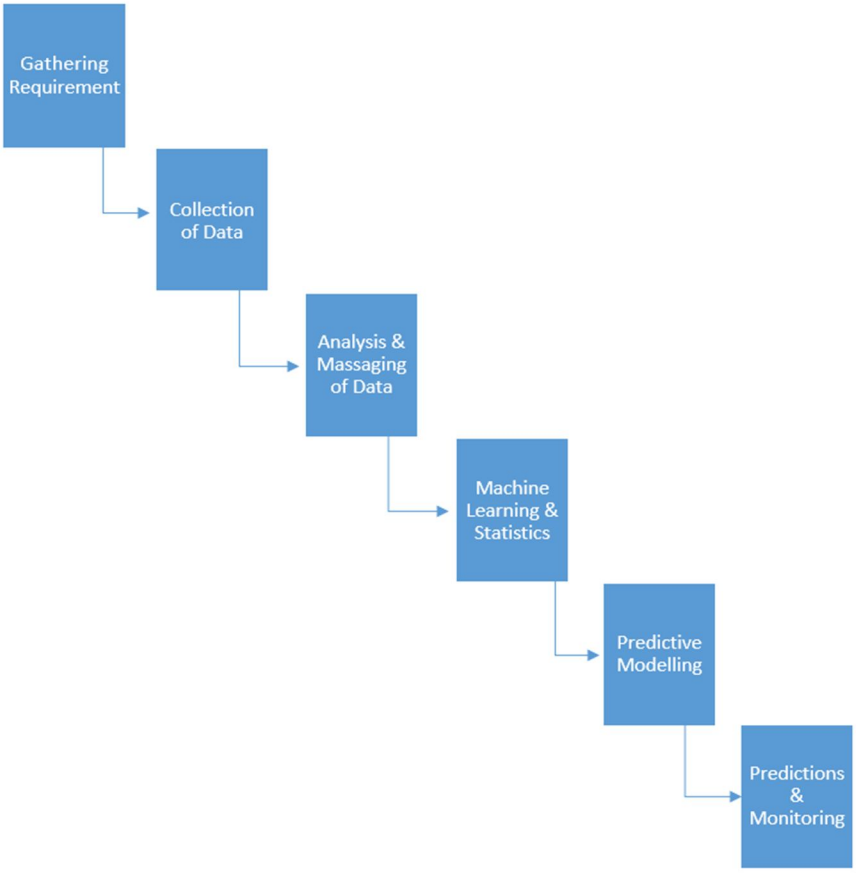


Figure 1: Predictive Analytics Process

A. Requirement Collection

The objective of prediction must be transparent to construct a predictive model. The variety of knowledge that will be obtained must be stated by the use of the prediction. For instance, a food company wants to foresee the sale of a particular product in a given area to circumvent expiry of that product. The data analyst meets the clients and discusses the requirements and the benefits involved in these predictions. What type of data will be required for developing the model is known through the client.

B. Data Collection

Once the requirements are known, the analyst will obtain datasets from various sources depending on the need. It may also require all the names of consumers who utilize or go through the company's products. Data can be structured or unstructured. The verification of this data is done by the analyst.

C. Data Analysis and Massaging

The analysis of the gathered data is done and the data is made ready to be used in the model. Structured data is derived from the unstructured data in the given process. When the whole data is accessible in the structured form, its standard is checked. Any inaccurate data or missing parameters of the attributes inside the dataset is given attention. The refinement of the data is an important factor for the productivity of the model. Transforming the unprocessed data into a format that is applicable for analytics is defined as data massaging or munging.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

D. Statistics, Machine Learning

A variety of machine learning and statistical techniques are applied by the predictive analytics process. Most important techniques used in the analytics are Probability theory and Regression analysis. Likewise, many predictive analytics use tools like artificial neural networks, decision trees, and support vector machines. Most of the predictive models are built on statistical or/and machine learning techniques. Therefore, the fundamentals of statistics and machine learning are applied by the analysts to construct predictive models. Machine learning techniques have an edge over the traditional statistical techniques. However, statistics must be included in the development of any predictive model.

E. Predictive Modeling

In this particular process, by the virtue of the statistical and machine learning techniques and the example dataset a model is constructed. Once developed, the model is tested on the test dataset that is a constituent of a main dataset to verify the viability of the model. If the test results are successful, the model is termed as fit. A fit model can give accurate predictions on the desired inputs in the system. Multi-models are also widely used these days in a variety of applications.

F. Prediction and Monitoring

The model is deployed at the client's site after successful tests are carried out. This model is used by the clients for day to day predictions and decision-making processes. The model generates the results and reports and not the managerial process. The model is always supervised to make sure that it is giving correct results and has a good accuracy.

Therefore, we have observed that predictive analytics is not the only step to make predictions about the future. It involves various steps that include multiple processes from requirement gathering to deployment and monitoring. These steps help in effective use of the system in decision-making processes. An easy way to comply with IJRASET paper formatting requirements is to use this document as a template and simply type your text into it.

III.PREDICTIVE ANALYTICS OPPORTUNITIES

Predictive analytics is an add-on of Data Mining technology. Tremendous amounts of mathematical theories are used by both of these concepts. Large chunks of data can be studied using Data Mining technology [6].

Due to developments in the technology, predictive analytics is thriving. With the objective of maximizing their profits, numerous businesses are opting for predictive analytics. The reasons for are same are discussed below:

1) Due to the emergence of various software and ongoing developments in the same has made the software user-friendly.
2) The competitiveness among the companies for growing with profits and financial circumstances force them to use predictive analytics.
3) Processing is made available by user-friendly, economical and swift computers.
4) Finding insights from a huge number of data has become viable due to the uprise in the volume and category of data, this makes companies use predictive analytics.

Predictive analytics is not being limited to the statisticians and mathematicians because of the development of easy to use and interactive software and its availability. Business analysts and managerial decision processes use predictive analytics on a large scale.

Few of the most common opportunities in the field of predictive can be listed as:-

A. Marketing Campaign Optimization

With the help of predictive analytics, the feedback of consumers after purchasing the product can be known. Cross-sale opportunities can be promoted using these models. These types of models assist businesses in alluring and retaining the most beneficial clients [7].

B. Operation Improvement

Predictive analytics can be used for pulling off auguring on directory and managing the assets. To decide the prices of apartments, real estate companies can use predictive models. To aggrandize its occupancy and gain more money, hotels can forecast the number of guests on a particular day by the virtue of predictive models. The efficacy and profits can be affected in a good way using such models [8].

C. Clinical Decision Support System

Development of medicines can be enhanced using predictive analytics. Also, expert systems can be used to diagnose patients. These models are created using predictive analytics [9].

D. Reduction of Risk

Credit score can be obtained by predictive analytics which can predict if the customer will be able to pay for a service or not. This credit score is produced by the model by using numerous data in relation with customer's credibility. Insurance companies, credit card issuers use this kind of models to identify illicit consumers [10].

E. Detecting Fraud

By combining various analysis methods, curtailing & detection of criminal behavior patterns may be made better. Cybersecurity has become a major concern. To regulate the actions on the network in real time, behavioral analytics can be implemented. Shady activities can be caught using such models. Moreover, potential threats can be prevented too [11].

IV. CATEGORIES OF PREDICTIVE ANALYTICS MODELS

The term predictive analytics means Predictive Modeling, which is defined as the scoring of data using predictive models and then forecasting them as output. But in general, it is used as a term to refer to the disciplines related to analytics. These disciplines include the process of data analysis and used in business decision making. These disciplines can be categorized as the following:-

A. Predictive Models

Predictive modelling uses statistics to predict outcomes [12]. Most often the event one wants to predict is in the future, but predictive modelling can be applied to any type of unknown event, regardless of when it occurred. For example, predictive models are often used to detect crimes and identify suspects, after the crime has taken place [13]. In many cases the model is chosen on the basis of detection theory to try to guess the probability of an outcome given a set amount of input data, for example given an email determining how likely that it is spam [14].

B. Descriptive Models

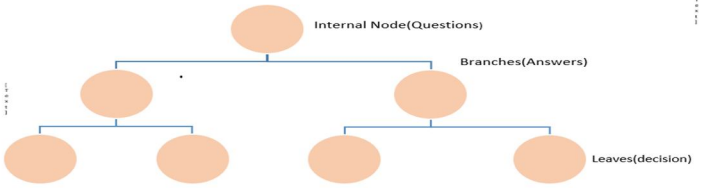
Descriptive models quantify relationships in data in a way that is often used to classify customers or prospects into groups. Unlike predictive models that focus on predicting a single customer behavior (such as credit risk), descriptive models identify many different relationships between customers or products. Descriptive models do not rank-order customers by their likelihood of taking a particular action the way predictive models do. Instead, descriptive models can be used, for example, to categorize customers by their product preferences and life stage. Descriptive modeling tools can be utilized to develop further models that can simulate a large number of individualized agents and make predictions [14].

C. Decision Models

The relationship between the data, the decision, and the result of the forecast of a decision are described by the decision models. In order to make a prediction on the result of a decision which involves many variables, this relationship is described as the decision model. To maximize certain outcomes, minimize some other outcomes, and optimize, these models are used. It is used in developing business rules to produce the desired action for every customer or in any circumstance [15].

The predictive analytic model is defined precisely as a model which predicts at a detailed level of granularity. It generates a predictive score for each individual. It is more like a technology which learns from experience in order to make predictions about the future behavior of an individual. This helps in making better decisions. The accuracy of results by the model depends on the level of data analysis [15]

V. PREDICTIVE ANALYTICS TECHNIQUES



A. Decision Tree

Decision tree techniques which are based on ML, use classification algorithms from data mining to identify the possible threats and rewards of pursuing several different courses of action. Potential outcomes are then presented as a flowchart which helps humans to visualize the data through a tree-like structure.

1) A decision tree has three major parts: a root node, which is the starting point, along with leaf nodes and branches. The root and leaf nodes ask questions.

2) The branches link the root & leaf nodes, depicting the flow from questions to answers. Generally, each node has numerous additional nodes extending from it, representing potential answers. The answers may not be complex, rather they can be simple as "yes" and "no."

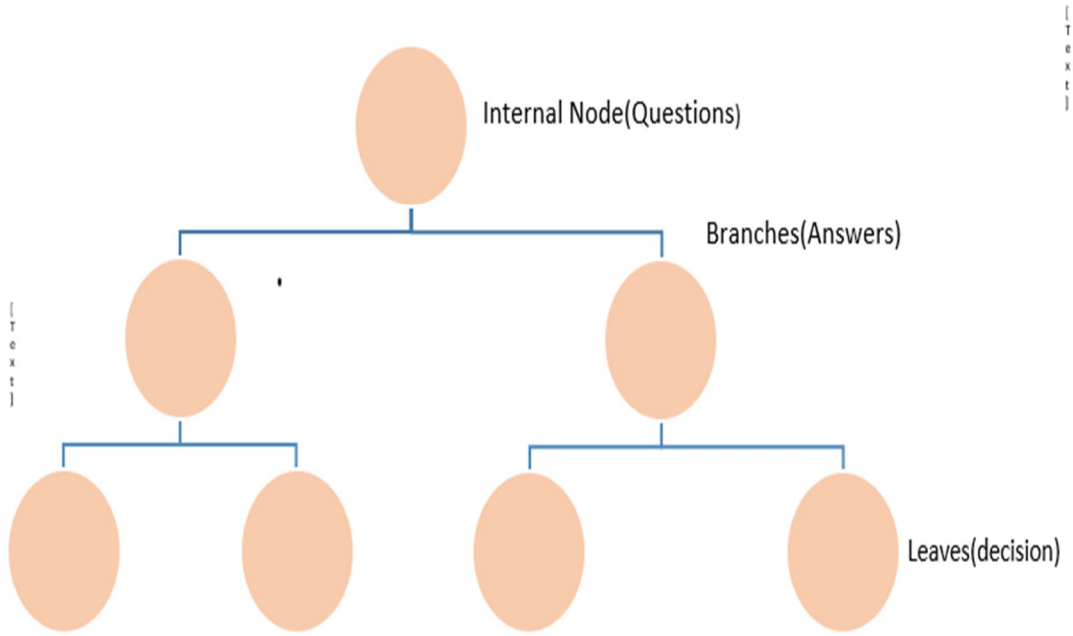


Figure 2: Decision Tree

B. Neural Network:

As traditional ML-based predictive analytics techniques like multiple linear regression aren't always good at handling big data. For example, big data analysis frequently requires an understanding of the sequence or timing of events. These techniques are much more adept at managing sequence and internal time orderings. Neural networks can make better predictions on time series information; for example, weather data. However, neural networking stands out at some types of statistical analysis, its applications range much further than that.

1) For its part, non-traditional data extends way beyond text data such as social media tweets and emails. Deep learning techniques are also required for data input such as maps, audio, video, and medical images. These techniques create layer upon layer of neural networks to analyse complex data shapes and patterns, improving their accuracy rates by being trained on representative data sets.

2) Deep learning techniques are majorly applied in image classification applications like voice and facial recognition and in predictive analytics techniques based on those methods. For example, to keep an eye on viewers' reactions to TV show trailers and decide which TV programs to run in various world markets, BBC Worldwide has developed an emotion detection application. The application leverages an offshoot of facial recognition called face tracking, which analyses facial movements. The objective is to predict the emotions that viewers would encounter when watching the actual TV shows

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

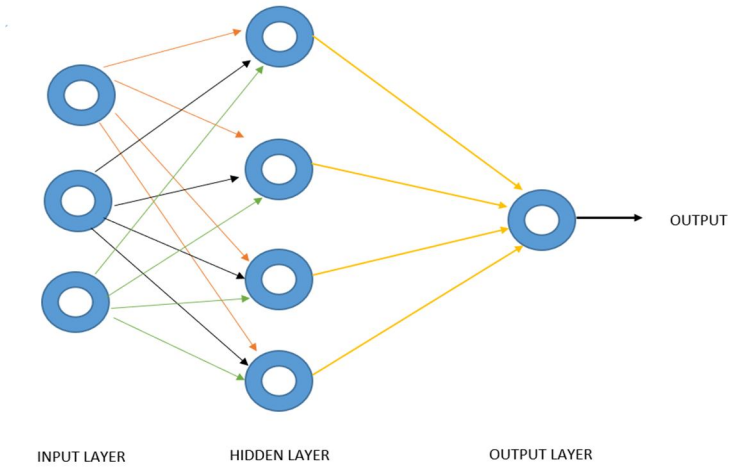


Figure 3: Artificial Neural Network

C. Regression Model

As predictive analytics is a tool for machine learning and big data, regression modelling is a tool for predictive analytics—one of the primary tools in fact. Regression analysis necessitates looking at dependent variables (outcomes) and an independent variable (the action) while also evaluating the power in the association among them. In other words, it looks to understand if there is a relationship between variables and how strong that relationship is.

An instance of a regression model, as it correlates with the market research and the predictive analytics, may incorporate understanding of how the likelihood to purchase is affected online by undemanding product search and the delivery cost. The regression output may show that the ease of product search has a powerful association with a possibility to purchase and as a result, more focus should be placed on making better that variable over delivery cost.

There are numerous regression techniques to be used depending on the kind of classification of predictive analytics and the class of variables involved. Regression models are used in predictive analytics for foreseeing, the linear regression model, and the logistic regression model. The linear regression model is beneficial to model the linear correlation between dependent and independent variables. A linear function is used as a regression function in this model. Whereas, the logistics regression when there are genres of dependent variables. Through this model, unknown values of discrete variables are predicted on the basis of known values of independent variables. It can presume a finite number of values in prediction.

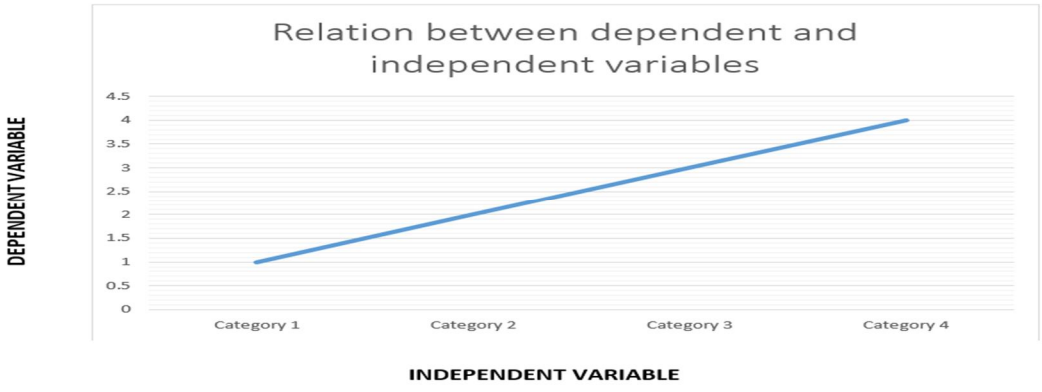


Figure 4: Regression Model

D. Ensemble Learning

Within machine learning & statistics, multiple learning algorithms are utilized by ensemble methods to acquire better predictive performance than could be obtained from any of the constituent learning algorithms alone [16]. Unlike a statistical ensemble in statistical mechanics, which is habitually infinite, a machine learning ensemble comprises only a concrete finite set of alternative models, but usually allows for much more flexible structure to exist among those choices. . The instance of classification using ensemble learning is represented in figure.

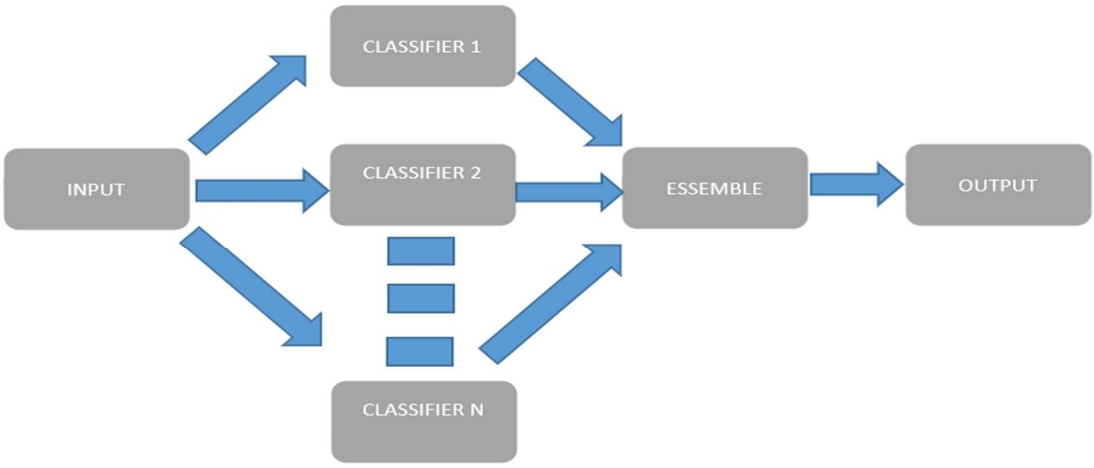


Figure 5: Ensemble Classifier

E. Bayesian Statistics

This technique belongs to the statistics which takes parameters as random variables and uses the term “degree of belief” to define the probability of occurrence of an event. The Bayesian statistics is based on Bayes’ theorem which terms the events priori and posteriori. In conditional probability, the approach is to find out the probability of a posteriori event given that priori has occurred [15].

For instance, an individual has a prior belief of a candidate's chances of winning an election and their confidence can be quantified as a probability. However, another person also may have a separate differing preceding opinion about the same candidate's odds. As new data arrives, both beliefs are (rationally) updated by the Bayesian procedure. Thus, in the Bayesian interpretation a probability is a synopsis of an individual's belief. A key point is that different (intelligent) individuals can have different opinions (and thus different prior beliefs), since they have differing access to data and ways of interpreting it. However, as both of these individuals come across new data that they both have access to, their (potentially differing) prior beliefs will lead to posterior beliefs that will begin converging towards each other, under the rational updating procedure of Bayesian inference.

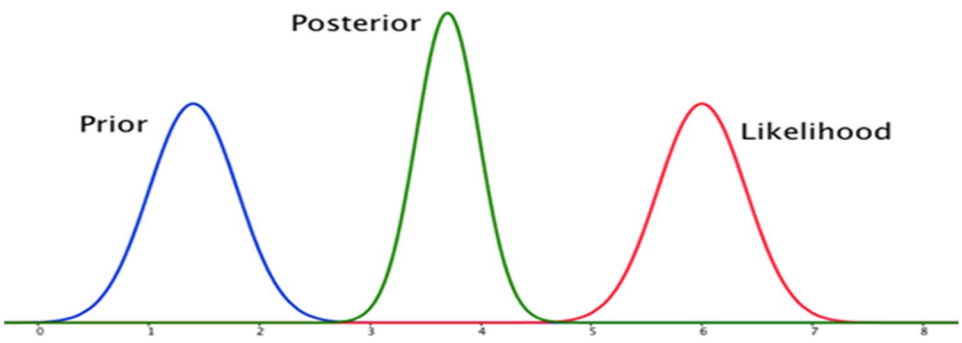


Figure 6: Bayesian Statistics

F. Gradient Boost Model

This technique is used in predictive analytics as a machine learning technique. It is mainly used in classification and regression-based applications. It is like an ensemble model which ensembles the predictions of weak predictive models that are decision trees [17]. Gradient boosting is a common ensemble method, of which the idea is to build the weak learner in the direction of the gradient to get the best results in the least amount of time. This method solves the optimization problem in function space imitating the gradient descent method in numerical space [18].The model is represented in the following figure.

Figure 7: Gradient Boosting

G. Support Vector Machine

It is a supervised kind of machine learning technique popularly used in predictive analytics. With associative learning algorithms, it analyzes the data for classification and regression [19] [20]. However, it is mostly used in classification applications. It is a discriminative classifier which is defined by a hyperplane to classify examples into categories. It is the representation of examples in a plane such that the examples are separated into categories with a clear gap. The new examples are then predicted to belong to a class as which side of the gap they fall [15]. Following is an example of a support vector.

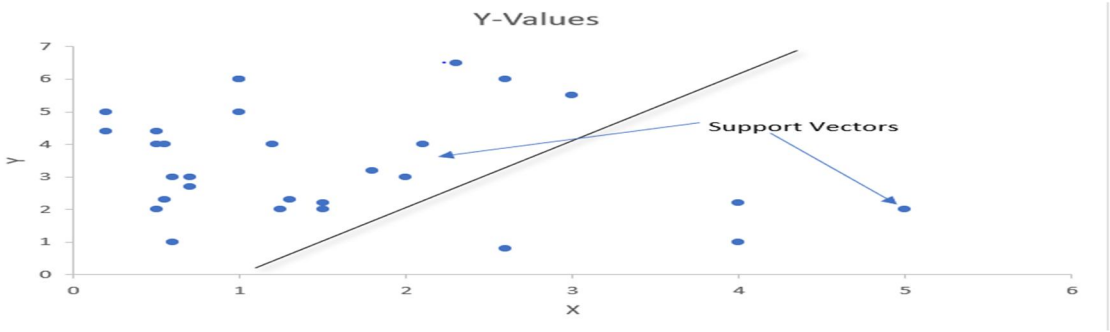


Figure 8: Support Vector Machine

H. Time Series Analysis

One method which deals with time based data is Time Series Modeling. Time series Modelling uses methods of prediction & forecasting As the name suggests, it involves working on time (years, days, hours, minutes) based data, to derive hidden insights out of large data sets. Time series models are very useful models when you have serially correlated data. Most companies and enterprises work on time series data to analyze sales numbers for the next year, website traffic, competition position and much more. However, it is also one of the areas, which many analysts do not understand. Below is an example of a time series graph of sales of a certain company.

Figure 9: Time Series Analysis

I. KNN

K-Nearest Neighbour is one among the only Machine Learning algorithms supported Supervised Learning technique.K-NN algorithm assumes the similarity between the new data and available data and puts the new data into the category that's most almost like the available categories. It stores all the available data and classifies a new data point based on the similarity. This means when new data appears then they are often easily classified into a well suite category by using K- NN algorithm.K-NN algorithm are often used for Regression and also for Classification but mostly it's used for the Classification problems. K-NN may be a non-parametric algorithm, which suggests it doesn't make any assumption on underlying data. Below is an example of regression using the KNN algorithm.

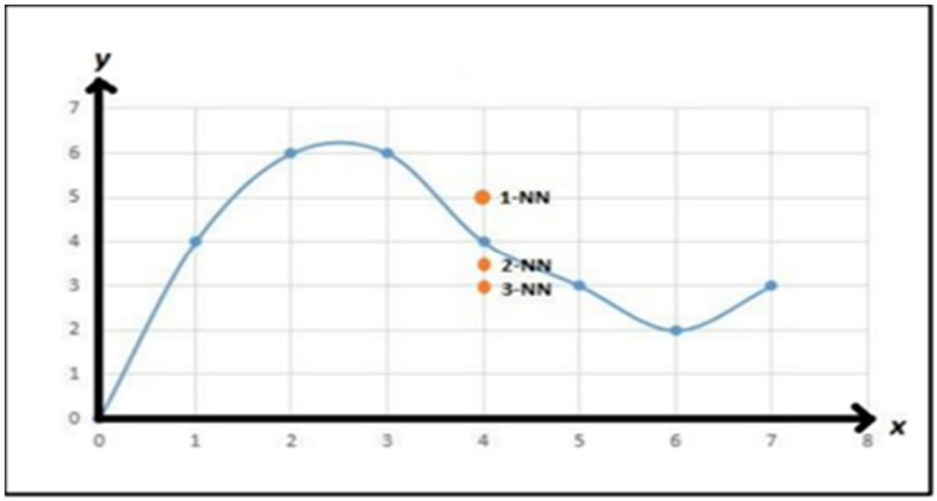


Figure 10: Regression using k-NN

J. Principal Component Analysis

Principal Component Analysis, or PCA, is a spatial property-reduction technique that's typically accustomed to cut back the dimensionality of enormous data sets, by reworking a massive set of variables into a smaller one that also contains most of the data within the large set. Reducing the quantity of variables of a knowledge set naturally comes at the expense of accuracy, however the trick in spatial property reduction is to trade a bit accuracy for simplicity. As a result, smaller information sets are easier to explore and visualize and build information so that it becomes a lot easier and quicker for machine learning algorithms to process extraneous variables. Following is an example of principal component plotting.

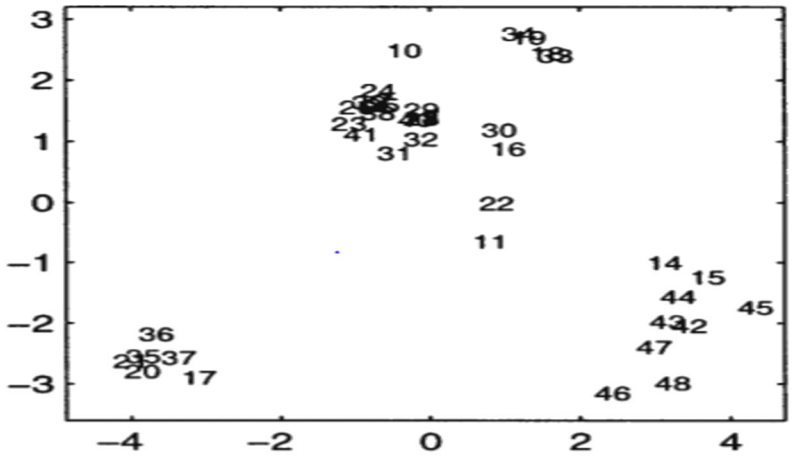


Figure 11: Principal Component Analysis

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

VI. DISTINCTIONS AND FEATURES OF TRADITIONAL AND BIG DATA APPROACHES IN BUSINESS

TABLE I

Approaches	Traditional approach	Big data approach
Business user	Calculate what question to ask	Inspects what questions can be asked
information technology	Structures the data to answer that question	Delivers a platform to enable creative discovery
Main applications	Monthly sales reports, profitability analysis, client surveys	Maximum asset utilization, Brand sentiment, Product strategy

VII. APPLICATION OF PREDICTIVE ANALYTICS

There are numerous applications of predictive analytics in many domains. From loan prediction to stock market prediction where a person's capacity to repay the loan can be predicted based on his background and return on a stock, investment can be estimated respectively. Few of the most common applications are listed below:

A. Oil Gas and Utilities

The oil and gas industries are using the predictive analytics techniques in foretelling the failure of equipment in order to reduce the risk. They foresee the need for resources in future using these models. The need for maintenance can be predicted by energy-based companies to avoid any fatal accident in future [21].

B. Retail

The retail industry is aided in identifying the clients and comprehending their needs by predictive analytics. The behavior of a consumer towards a particular product is made known by using these techniques. Once the spending behavior of a client is known, companies can fix prices and give special offers on products. These models also help in understanding if a specific product will do well in a given season. Accordingly, campaigns can be carried out by these companies. Supply chains can be refined as well using predictive analytics. They identify and predict the demand for a product in the specific area may improve their supply of products [22].

C. Banking and Financial Services

In banking and financial industries, there are tremendous uses of predictive analytics. In both the industries data and money is a crucial part and finding insights from those data and the movement of money is a must. The predictive analytics helps in detecting the fraudulent customers and suspicious transactions. It reduces the credit uncertainty on which these industries lend money to its customers. It helps in cross-sell and up-sell opportunities and in retaining and attracting the valuable customers. For the financial industries where money is invested in stocks or other assets, predictive analytics foretells the return on investments and helps in the investment decision making process.

D. Health and Insurance

The pharmaceutical sector uses predictive analytics in drug designing and improving their supply chain of drugs. By using this technique, these companies may predict the expiry of drugs in a specific area due to lack of sale. The insurance sector uses predictive analytics models in identifying and predicting the fraud claims filed by the customers. The health insurance sector uses this technique to find out the customers who are most at risk of a serious disease and approach them in selling their insurance plans which are best for their investment [23].

E. Government and Public Sector

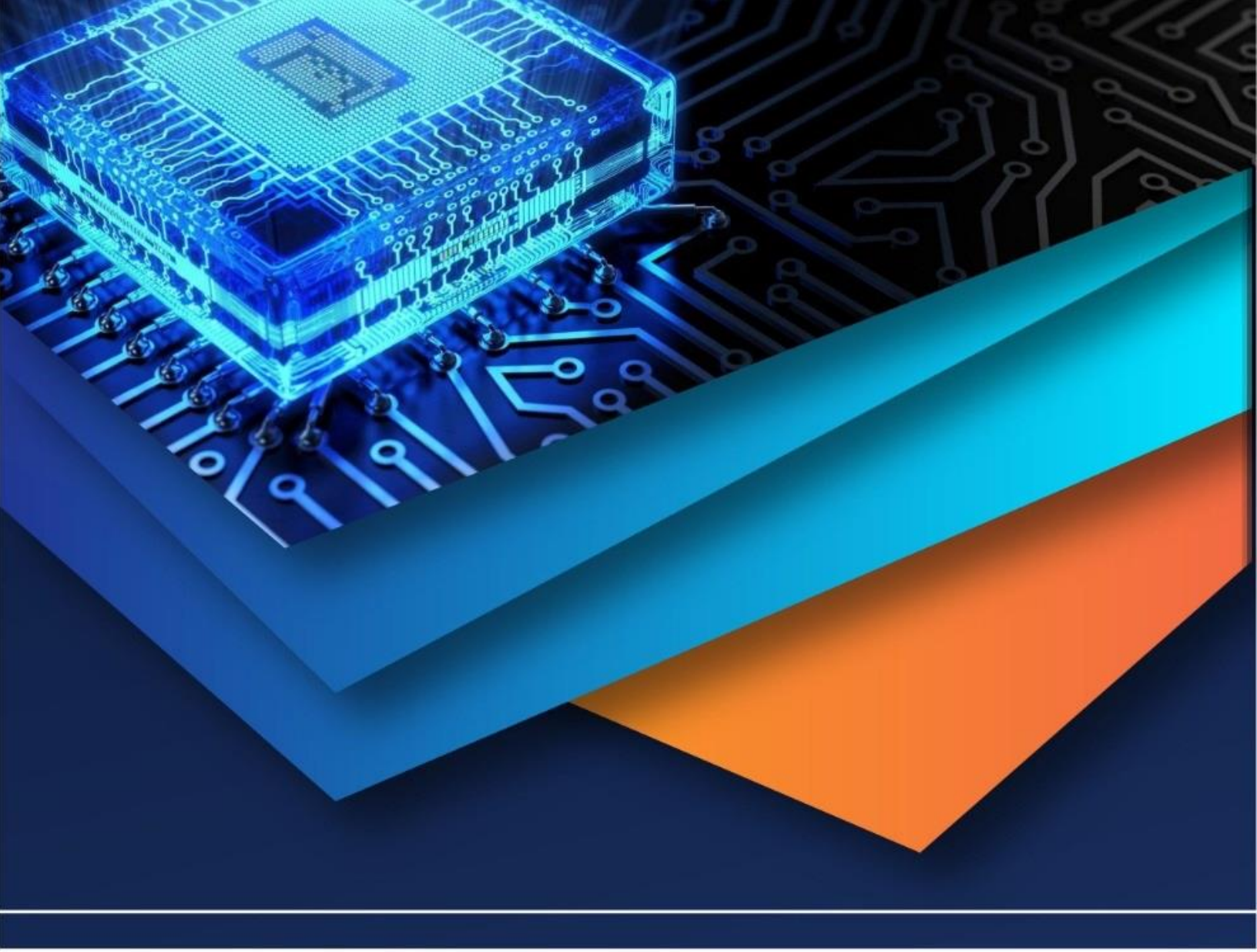
The government agencies are using big data-based predictive analytics techniques to identify the possible criminal activities in a particular area. They analyze the social media data to identify the background of suspicious persons and forecast their future behavior. The governments are using predictive analytics to forecast the future trend of the population at country level and state level. In enhancing cyber security, predictive analytics techniques are being used in full swing [24].

VIII. CONCLUSIONS

There has been an extended history of using predictive models within the tasks of predictions. Earlier, the statistical models were used which supported the sample data of a large-sized data set. With the improvements within the field of computing and the advancement of computer techniques, newer techniques have been developed and better and better algorithms have been introduced over the amount of time . The developments within the field of AI and machine learning have changed the world of computation where intelligent computation techniques and algorithms have been introduced. Machine learning models have a really good diary of getting used as predictive models. Artificial neural networks have brought a revolution within the field of predictive analytics. The output of any value can be predicted based on input parameters. Now, with the advancements within the field of machine learning and the development of deep learning techniques, there's a trend nowadays of using deep learning models in predictive analytics and that they are being applied in full swing in this task. This paper opens a scope of development of latest models for the task of predictive analytics. There's also a chance to add additional features to prevailing models to enhance their performance.

REFERENCES

[1] Elkan, C. (2013). Predictive analytics and data mining.
[2] Swani, L., & Tyagi, P. (2017). Predictive Modelling Analytics through Data Mining. International Research Journal of Engineering and Technology (IRJET).
[3] Siegel, E. (2016). Predictive Analytics.
[4] Nyce, C., & American Institute of CPCU/IIA. (2007). Predictive Analytics White Paper.
[5] Eckerson, W. (2007). Extending the Value of Your Data Warehousing Investment (The Data Warehouse Institute., Ed.).
[6] Mohsen Attaran , Sharmin Attaran 2018, 'Opportunities and Challenges of Implementing Predictive Analytics for Competitive Advantage' International Journal of Business Intelligence Research Vol 9, Issue 2
[7] F Reichheld, P Schefter, Retrieved 2018, "The Economics of E-Loyalty", Harvard Business School Working Knowledge.
[8] V Dhar, 2001, "Predictions in Financial Markets: The Case of Small Disjuncts", ACM Transaction on Intelligent Systems and Technology, Vol-2, Issue-3.
[9] J Osheroff, J Teich, B Middleton, E Steen, A Wright, D Detmer, 2007, "A Roadmap for National Action on Clinical Decision Support", JAMIA: A Scholarly Journal of Informatics in Health and Biomedicine, Vol-14, Issue2, Pages-141-145.
[10] M Schiff, 2012, "BI Experts: Why Predictive Analytics Will Continue to Grow", The Data Warehouse Institute
[11] M Nigrini, 2011, "Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations", John Willey and Sons Ltd
[12] Geisser, S. (. (1993). Predictive Inference: An Introduction. Chapman & Hall.
[13] Finlay, S. (2014). Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods (1st ed.).
[14] Predictive modelling. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Predictive_modelling
[15] Kumar, V., & Garg, M. (2018). Predictive Analytics: A Review of Trends and Techniques. International Journal of Computer Applications.
[16] David Opitz, Richard Maclin 1999, "Popular ensemble methods: An empirical study". Journal of Artificial Intelligence Research 11, 169-198
[17] Friedman, J. H. (1999). Greedy Function Approximation: A gradient boosting machine.
[18] Wang, J., Li, P., Ran, R., Zhou, Y., & Che, Y. (2018). A Short-Term Photovoltaic Power Prediction Model. MDPI.
[19] Cortes, C. (1995). support-vector networks. Machine Learning, 125-127.
[20] al, B. H. (2001). Support Vector Clustering. Journal of Machine Learning Research, 125-137.
[21] J Feblowitz, 2013, "Analytics in Oil and Gas: The Big Deal About Big Data", Proceeding of SPE Digital Energy Conference, Texas, USA
[22] K Das, GS Vidyashankar, 2006, "Competitive Advantage in Retail Through Analytics: Developing Insights, Creating Values", Information Management
[23] N Conz, 2008, "Insurers Shift to Customer-Focused Predictive Analytics Technologies", Insurance & Technology
[24] G H Kim, S Trimi, J-H Chung, 2014, "Big-data applications in the government sector", Communications of the ACM, Vol-57, Issue-3, Pages-78-85.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY