iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.48071

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: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

An Insight into Predictive Analytics Techniques

Lakshmi Kumari[1], Dr. Ritu Aggrawal[2]
Department of IT, Management Education and Research Institute

Abstract: ***Predictive Analytics is the branch of advanced analytics which is used to make predictions about unknown future events. It deploys many techniques like Data mining, Artificial Intelligence and Machine Learning to analyze current data to make future predictions. In this paper we will study about different predictive models such as Decision Tree, Regression Analysis and Neural Network. These methods use known results to develop a model so as to predict values for different or new Data. Any industry can use predictive analytics techniques to reduce risk, optimize operations and increase revenue. It can be used in Banking and Financial industry, Governments and the public sector, Retail industry, Health care industry, Manufacturing, Oil and Gas Industry. This paper gives a good insight to whoever wants to know and use the predictive analysis for his academic or business applications.***

Keywords: Machine Learning, Regression analysis, Neural Network, Decision Tree, Big Data

I. INTRODUCTION

Predictive analytics is a form of technology that makes predictions about certain unknowns in the future. It makes use of a number of techniques to make these determinations including artificial intelligence, data mining, machine learning and statistics. Predictive analysis uses machine learning to follow predictive modelling. The model will be trained on a dataset to respond to a new data set or values. These models determine relations, patterns and structures in data that can be used to draw conclusions about how changes in the underlying processes that generate the data will change the results. Predictive models build on these descriptive models and look at past data to determine the likelihood of certain future outcomes, given current conditions or a set of expected future condition.

This paper is organized as follows: Section II Describes various predictive analysis process. Section III describes various predictive analysis methods and Finally, conclusion is presented in section IV.

II. PREDICTIVE ANALYTICS PROCESS

Predictive analytics is the main branch in the field of data analytics, which is uses quantitative methods and expert knowledge to derive meaning information from data and answer fundamental questions about any scientific research, the weather, healthcare, business and other areas of predictions.

There are three major types of predictive analytics. The first type is known as descriptive analytics, which gives an insight of what has happened in a area. The second is Predictive analytics, which helps to predict out about what will likely happen in future. It looks for patterns in data and forwards them to help businesses calculate risks and capitalize on opportunities. The third category, prescriptive analytics automatically takes a next best course of action based on intelligence generated by the other two kinds of analytics. Two additional modes of analytics sometimes figure into the business analytics continuum: diagnostic analytics, which explores why something happened, and real time analytics, which analyzes data as it's generated, collected or updated. [1]

Different variables are combined into a predictive model which is capable of predicting future probabilities with an acceptable level of reliability. The software relies heavily on advanced algorithms and methodologies, such as regression models, time series analysis and decision tree.

The predictive analytics is divided into the following phases:

1) *Define Problem:* This step defines the problem statement and also to identify the data sets that will be used.
2) *Data Collection:* Data collection will collect all important data involved in past or historical data from different authenticated sources .
3) *Data Cleaning:* Data cleaning process will remove unnecessary details from the data and will provide filter data without any noise. So the data will be available without any error or any duplicate values.
4) *Data Analysis:* This step involves the analysis of data in order to identify some similarity or patterns based on past data .Useful trends and patterns will be find to analyze the data.

©IJRASET: All Rights are Reserved

5) *Building Predictive Model:* This step involves using different algorithms to develop models depending on different patterns observed.
6) *Model Validation:* It is a very crucial step in predictive analysis. In this, we check the efficiency of our predictive model by doing different tests. We take a input data set to check the validity of the model used. Accuracy is also determined at this step.
7) *Deployment of Model:* It involves real time testing and implementation of the model. To make it available for predictive analysis .
8) *Monitoring:* We need to monitor the working of our model on regular basis so as to have efficient functioning of the model and to have proper results

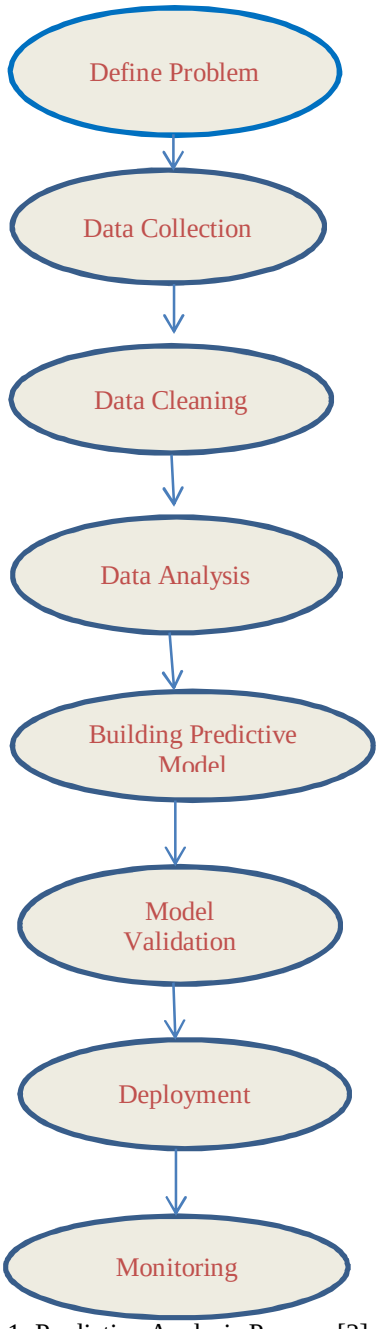


Fig 1: Predictive Analysis Process [2]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

III. PREDICTIVE ANALYTICS TECHNIQUES

These different techniques are used to indentify “unknowns” that previously was not known , leading to an overall need for analysts that can succinctly identify which model best aligns with the type of unknown in each scenario.
The widely used predictive analytics techniques are

A. Linear Regression

Linear regression is a supervised algorithm of machine learning. This technique predicts out the relationships among the variables and also used for forecasting. The variables used in this technique are assumed to be continuous,real or numeric .This technique is used to find out the value of dependent variable based on a independent variable.It finds a linear relationship between dependent and independent variable. Therefore known as Linear Regression.
Two types of Linear Regression are there:

1) *Simple Linear Regression:* In this one single independent variable is used to predict the value of dependent variable then it is know as Simple Linear Regression.
2) *Multiple Linear Regression:* In this dependent variable value is predicted from a multiple number of independent variable.

The line which shows the relationship between independent and dependent variable is known as Regression Line. The line can be positive as well as negative. The regression line is positive when both dependent and independent variable increases on Y-axis and X axis respectively. The line is negative when dependent variable decreases and independent variable increases on Y-axis and X-axis respectively.

B. Decision Tree

Decision Tree is a supervised algorithm.This algorithm uses tree representation to solve problems. In this technique a training model is made to predict the value or class of the target variable by using decision rules derived from training data. We scan from root node of the tree for finding the class label for a record and go on comparing to nodes in that branch and jump to the next node.

There are two types of Decision Tree

1) *Categorical Variable Decision Tree:* If the decision tree has a categorical target vaiable then the decision tree is known as categorical variable decision tree.
2) *Continuous Variable Descision Tree:* If the decision tree has a continuos target variable then I is known as continuous variable decision tree.

Decision trees solve the problem by sorting it down the tree from root node to leaf node. Each node in tree acts as test case for some attributes and each edge following the node is the answer to the test case. Therefore the process is recursive in nature .Decision tree follows sum of product representation. Attribute selection is done so as to define which attribute is to considered as the root node at each level. The two popular techniques for attribute selection are Information Gain and Ginni Index.A decision tree simply asking question and replying with yes/no ,the tree is split into subtree. Therefore a decision tree can have both categorical and numerical value.Decision trees usually copy humans in their thinking ability to make a decision . Decision trees are classification models that partition data into subsets based on categories of input variables. This helps you understand someone's path of decisions. A decision tree looks like a tree with each branch representing a choice between a number of alternatives, and each leaf representing a classification or decision. This model looks at the data and tries to find the one variable that splits the data into logical groups that are the most different. Decision trees are popular because they are easy to understand and interpret. They also handle missing values well and are useful for preliminary variable selection. So, if you have a lot of missing values or want a quick and easily interpretable answer, you can start with a tree.[3]

C. Neural Network

Neural Networks are the biological neural networks that do the processing after being exposed to a number of datasets and examples without any specific rules.ANNs have three node layers where first one is Input layer,second one is hidden layers which can in multiple numbers and the last one is output layer. The nodes connected to one another will have certain weight associated with them and also a minimum threshold value. If the output of any indivisual node is more than threshold value than that node is activated and sends the data to the next layer of the network.The basic unit of neural network is neurons, which takes some input ,does some processing with them and gives one output. So these bunch of neurons when connected together forms a neural network.

©IJRASET: All Rights are Reserved

The different types of neural networks are as following:

1) Feed Forward Neural Network
2) Recurrent Neural Network
3) Convolutional Neural Network
4) Deconvutional Neural Network
5) Modular Neural Network

If we talk about the advantages of neural networkthen we can say that neural network performs parallel processing which means it can do multiple jobs at a time.Also the information is stored on entire network not on database.Neural networks have an excellent fault tolerance ability.

The neural networks ahev wide applications in the areas like natural languague processing,translation and language generation, stock market prediction ,route planning and optimization.

D. Bayesian Analysis

In Bayesian analysis ,on the basis of prior distribution prediction is encoded in probability distribution for the unknown parameters. This method assigns "Degree Of belief" means the probability for a process that can change as new information is collected rather than fixed old value .Bayesian is used to build a statistical model based on Baye's Theorem.

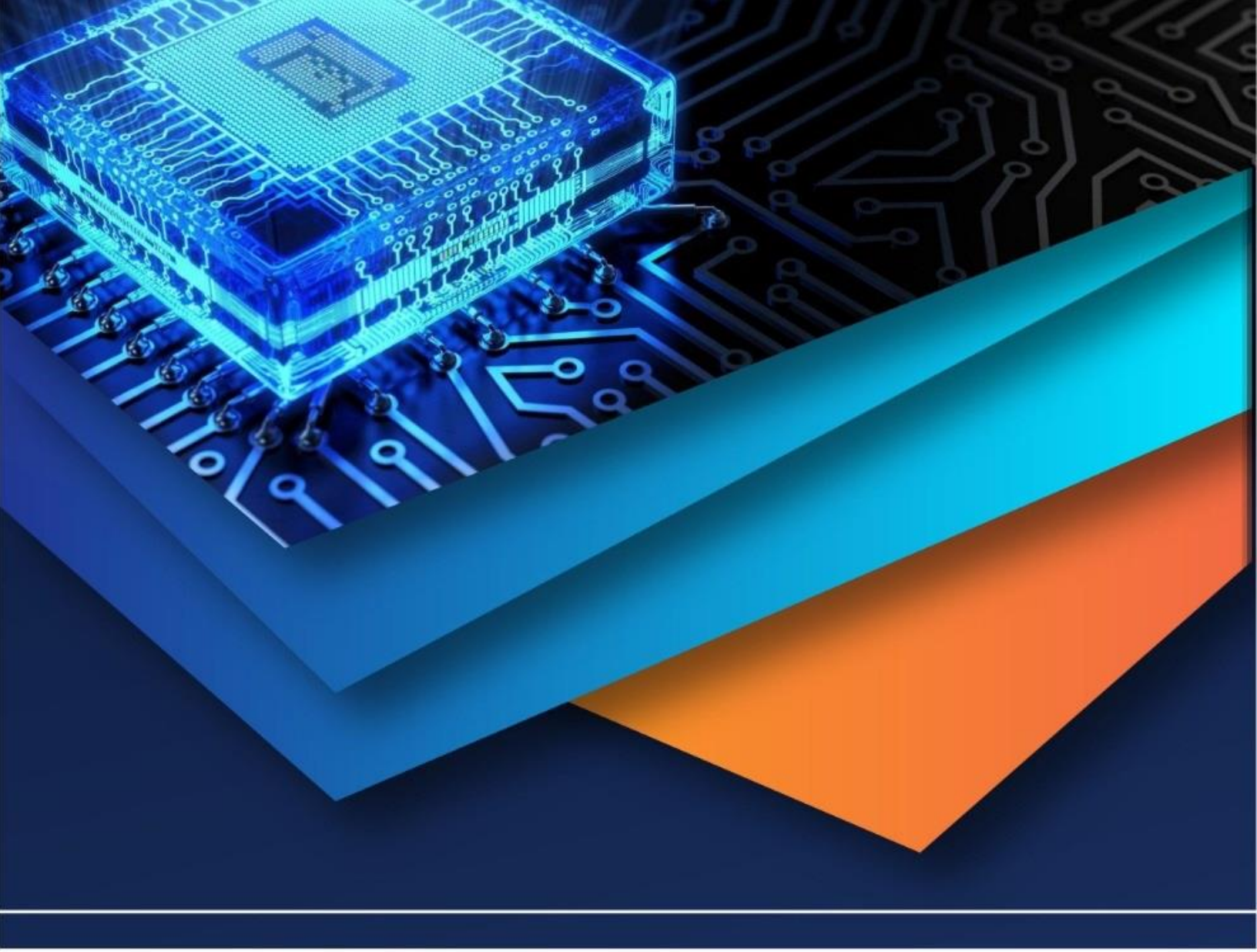
This technique is used in credit card fraud detection,spam filtering,medical diagnosis,patterns in customer datasets or marketing campaign performance,helps robots make decision.

IV. CONCLUSION

This paper gives am insight into various predictive analytics techniques which can use to predict out various future outcomes in different domains .The first thing you need to get started using predictive analytics is a problem to solve. What do you want to know about the future based on the past? What do you want to understand and predict? You'll also want to consider what will be done with the predictions. What decisions will be driven by the insights? What actions will be taken? Second, you'll need data. In today's world, there is a large amount of data. Transactional systems, data collected by sensors, third-party information, call center notes, web logs, etc. After that, the predictive model building begins.Anumber of easy-to-use software means more people can build analytical models. Someone in IT to ensure that you have the right analytics infrastructure for model building and deployment.

REFERENCES

[1] Deepti Aggarwal,Vikram Bali,Sonu Mittal. "An Insight into Machine Learning Techniques for Predictive Analysis and Feature Extraction",IJITEE,ISSN-2278-3075,Vol-8,Issue-9S,July2019
[2] Vaibhav Garg, M.L. Garg ." Predictive Analytics : A Review of Trends and Techniques ",IJCA,ISSN-0975-8887,Volume-182,July 2018.
[3] www.sas.com
[4] Muhhamad Razi A.,Kuriakose Athapilly, "A Comparative Predictive Analysis Neural Network, Non-Linear Regression,Classification and Regression Tree Model" Elsivier, Vol 29,Issue 1,65-74
[5] www.Investopedia.com



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)