IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207019 |

A Survey of Predictive Analytics Techniques, Their Advantages, Disadvantages, and Applications

Madhuri S. Desai, Dr. P. S. Deshpande,

PG Student, Dept. of CSE, Shreeyash College of Engineering and Technology, Dr. B. A. Tech. University, Chh. Sambhajinagar, India

HOD. Dept. of CSE, Shreeyash College of Engineering and Technology, Dr. B. A. Tech. University, Chh. Sambhajinagar, India

ABSTRACT: Predictive analytics is a term mainly used in statistical and analytics techniques. This term is drawn from statistics, machine learning, database techniques and optimization techniques. It has roots in classical statistics. It predicts the future by analysing current and historical data. The future events and behavior of variables can be predicted using the models of predictive analytics. A score is given by mostly predictive analytics models. A higher score indicates the higher likelihood of occurrence of an event and a lower score indicates the lower likelihood of occurrence of the event. Historical and transactional data patterns are exploited by these models to find out the solution for many business and science problems. These models are helpful in identifying the risk and opportunities for every individual customer, employee or manager of an organization. With the increase in attention towards decision support solutions, the predictive analytics models have dominated in this field. In this paper, we will present a review of process, techniques and applications of predictive analytics.

KEYWORDS: Predictive analytics, Predictive techniques, Machine Learning, Deep Learning, Statistical Methods, Application, Advantage, Sector, Benefit.

I. INTRODUCTION

Predictive analytics, a branch in the domain of advanced analytics, is used in predicting the future events. It analyses 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. It brings together the information technology, business modeling process, and management to make a prediction about the future.

II. PREDICTIVE ANALYTICS PROCESS

Predictive analytics involves several steps through which a data analyst can predict the future based on the current and historical data. This process of predictive analytics is represented in figure given below

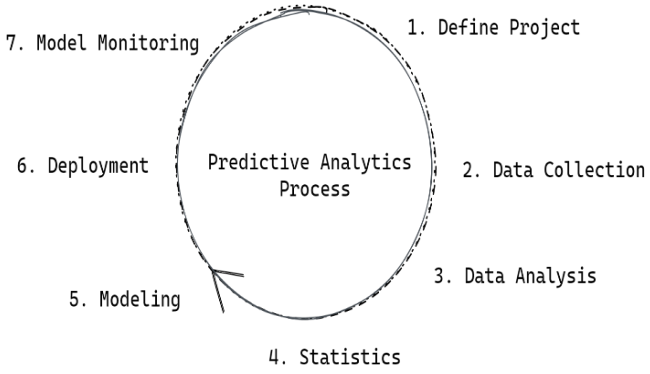

Fig. Predictive Analytics Process

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207019 |

II. PREDICTIVE ANALYTICS TECHNIQUES

The important techniques below which are used popularly in developing the predictive models.

1. Decision Tree

A decision tree is a classification model but it can be used in regression as well. It is a tree-like model which relates the decisions and their possible consequences. The consequences may be the outcome of events, cost of resources or utility. In its tree-like structure, each branch represents a choice between a number of alternatives and its every leaf represents a decision. Based on the categories of input variables, it partitions data into subsets. It helps the individuals in decision analysis. A typical model of the decision tree is represented in figure given below.

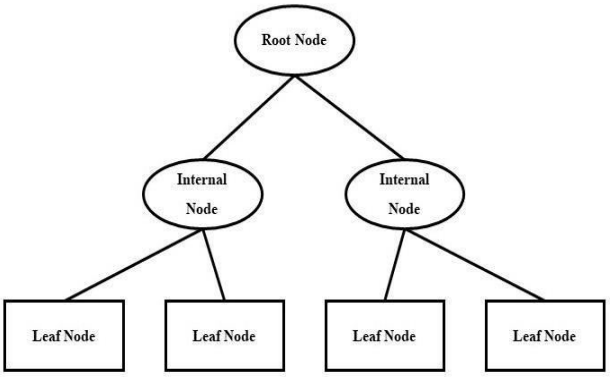

Fig. Decision Tree

A decision tree is represented in figure as a tree-like structure. It has the internal nodes labeled with the questions related to the decision. All the branches coming out from a mode are labeled with the possible answers to that question. The external nodes of the tree called the leaves, are labeled with the decision of the problem. This model has the property to handle the missing data and it is also useful in selecting the preliminary variables.

2. Regression Model

Regression is one of the most popular statistical technique which estimates the relationship between variables. It models the relationship between a dependent variable and one or more independent variables. Regression analysis is used for one of two purposes: predicting the value of the dependent variable when information about the independent variables is known or predicting the effect of an independent variable on the dependent variable. This modeled relation between dependent and independent relation is represented in figure given below.

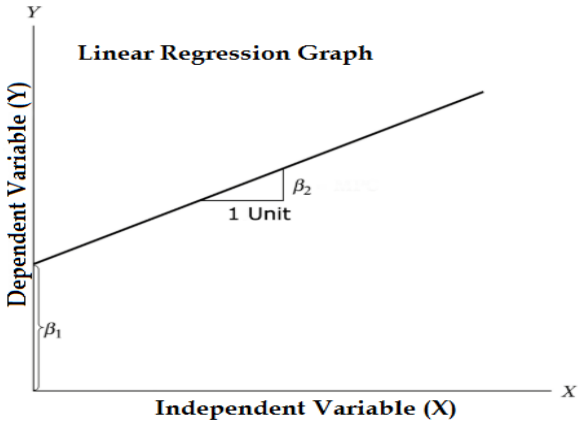

Fig. Regression Model

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207019 |

In the context of the continuous data, which is assumed to have a normal distribution, the regression model finds the key pattern in large datasets. It is used to find out the effect of specific factors influence the movement of a variable. In regression, the value of a response variable is predicted on the basis of a predictor variable. In this case, a function known as regression function is used with all the independent variables to map them with the dependent variables. In this technique, the variation of the dependent variable is characterized by the prediction of the regression function using a probability distribution. There are two types of regression models are used in predictive analytics, the linear regression model, and the logistic regression model.

3. Bayesian Statistics

This technique belongs to the statistics which takes parameters as random variables and use 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. On the other hand, the Bayes' theorem finds the probability of priori event given that posteriori has already occurred. It is represented in figure.

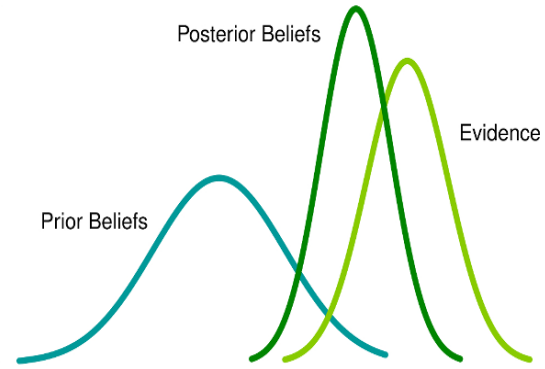

Fig. Bayesian Statistics

It uses a probabilistic graphical model which is called the Bayesian network which represents the conditional dependencies among the random variables. This concept may be applied to find out the causes with the result of those causes in hand. For example, it can be applied in finding the disease based on the symptoms.

1. Artificial Neural Network

Artificial neural network, a network of artificial neurons based on biological neurons, simulates the human nervous system capabilities of processing the input signals and producing the outputs [13]. This is a sophisticated model that is capable of modeling the extremely complex relations. The architecture of a general purpose artificial neural network is represented in figure.

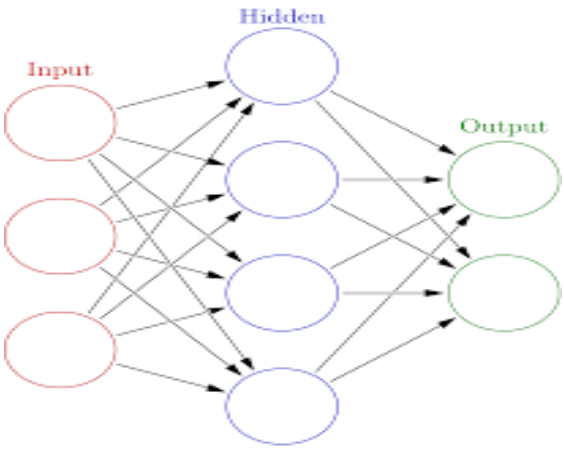

Fig. Artificial Neural Network

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207019 |

Artificial neural networks are used in predictive analytics application as a powerful tool for learning from the example datasets and make a prediction on the new data. Through the input layer of the network, an input pattern of the training data is applied for the processing and it is passed to the hidden layer which a vector of neurons. Various types of activation functions are used at neurons depending upon the requirement of output. The output of one neuron is transferred to the neurons of next layer. At the output layer, out is collected that may be the prediction on new data.

2. Support Vector Machine

It is supervised kind of machine learning technique popularly used in predictive analytics. With associative learning algorithms, it analyzes the data for classification and regression. 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. The example of separation by a support vector machine is represented in figure.

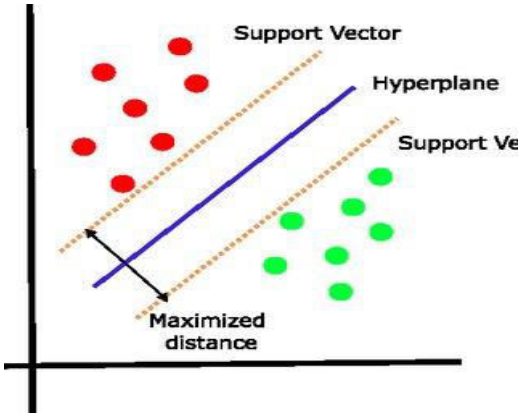

Fig. Support Vector Machine

III. ADVANTAGES AND DISADVANTAGES OF PREDICTIVE ANALYTICS TECHNIQUE

Sr No.	Technique	Advantages	Disadvantages
1.	**Decision tree**	• Easy to interpret and visualize • Handles categorical features and missing values • Fast training and prediction	• Can be biased towards majority class. • Prone to overfitting (greedy algorithm). • Not suitable for high-dimensional data.
2.	**Random Forest**	• Reduces overfitting and improves generalization • Handles high-dimensional data and missing values • Provides feature importance scores	• Computationally expensive • Can be difficult to interpret • May not perform well with very small datasets
3.	**Regression Analysis**	• Interpretable coefficients • Handles continuous outcomes. • flexible modeling options (linear, non-linear, regularization)	• Assumes linear relationships • Sensitive to outliers and multicollinearity • Can be overly simplistic for complex relationships

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207019 |

4.	**Bayesian Statistics**	• It allows for complex modeling and estimation of parameters. • Can be more robust to outliers and misspecification. • Allows for incorporation of prior knowledge or expertise into the analysis.	• Can be computationally intensive. • Validating Bayesian models can be challenging. • Can overfit the data if not regularized.
5.	**Artificial Neural Network**	• Can learn complex non-linear relationships • Flexible architecture design • High accuracy potential	• Difficult to interpret and visualiz • Requires large amounts of data and computational resources • Can suffer from overfitting and vanishing gradients
6.	**Gradient boosting**	• Its high predictive accuracy. • Can handle large datasets with many features. • Can handle missing values in the data.	• Can overfit the data if not regularized. • Can be computationally expensive. • Can be difficult to interpret the results in some cases.
7.	**Support Vector Machine**	• Robust to noise and outliers • Handles high-dimensional data and non-linear relationships. • Provides robust classification boundaries	• Can be computationally expensive • Not suitable for very large datasets • May not perform well with very imbalanced classes

Table 1: Advantages and disavantages of different predictive techniques

IV. BENEFITS OF PREDICTIVE ANALYTICS

1. **Scalability**
   Data science and data engineering tasks can be automated. Models can be trained, tested, and deployed across multiple enterprise applications in real-time. Extend data science skills in hybrid and multi-cloud systems.
2. **Simplicity**
   To handle the complete data science lifecycle, use a central platform. Develop and deploy consistent processes. Create unified data governance and security framework for the entire enterprise.
3. **Speed**
   Utilize pre-built applications and models that have already been trained. With state-of-the-art and open source technologies, data scientists and business teams may collaborate and streamline model construction.
4. **Fraud Detection**
   Predictive analytics analyses all network actions in real-time to spot anomalies that could signal fraud or other vulnerabilities.
5. **Operations Improvement**
   Businesses employ predictive analytics models to estimate inventory, manage resources, and run more efficiently.
6. **Risk Reduction**
   Predictive analytics is used to examine and determine the risk of future defaults in credit scores, insurance claims, and debt collections.
7. **Predictive Maintenance**
   Data is used by organizations to predict when routine equipment maintenance is needed and to plan it before a problem or malfunction occurs.

V. APPLICATION OF PREDICTIVE ANALYTICS

There are many applications of predictive analytics in a variety of domains. From clinical decision analysis to stock market prediction where a disease can be predicted based on symptoms and return on a stock, investment can be estimated respectively. We will list out here below some of the popular applications.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207019 |

1. Banking and Financial Services

In banking and financial industries, there is a large application of predictive analytics. In both the industries data and money is 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 minimizes the credit risk on which theses 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, the predictive analytics forecasts the return on investments and helps in investment decision making process.

2. Retail

The predictive analytics helps the retail industry in identify the customers and understanding what they need and what they want. By applying this technique, they predict the behavior of customers towards a product. The companies may fix prices and set special offers on the products after identifying the buying behavior of customers. It also helps the retail industry in predicting that how a particular product will be successful in a particular season. They may campaign their products and approach to customers with offers and prices fixed for individual customers. The predictive analytics also helps the retail industries in improving their supply-chain. They identify and predict the demand for a product in the specific area may improve their supply of products.

3. 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 using 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 be best for their investment.

4. 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 the predictive analytics to forecast the future trend of the population at country level and state level. In enhancing the cybersecurity, the predictive analytics techniques are being used in full swing.

VI. RELATED WORKS

The table in the below section states the work done using various algorithms and modeling techniques in different areas such as education, finance-driven models, weather forecasting, medical research, neuroscience, agriculture, text-mining etc. The applied techniques enhance the simplicity, effectiveness and makes the system more robust.

Technique	Sector	Application	Benifit
Decision Trees	Finance, Healthcare, Marketing, Transportation	Credit risk assessment, disease diagnosis, customer segmentation, route optimization	Easy to interpret, handles missing values, and identifies important features
Regression Analysis	Finance, Economics, Marketing, Healthcare	Predicting stock prices, sales forecasting, risk analysis, medical research	Identifies relationships between variables, predicts continuous outcomes, and estimates confidence intervals

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207019 |

Random Forest	Finance, Healthcare, Marketing, Environmental Monitoring	Fraud detection, medical diagnosis, customer churn prediction, climate modeling	Reduces overfitting, improves accuracy, and handles high-dimensional data
Neural Networks	Technology, Healthcare, Finance, Marketing	Image recognition, natural language processing, fraud detection, recommendation systems	Handles complex relationships, learns from large datasets, and improves over time
Gradient Boosting	Finance, Marketing, Healthcare, Environmental Monitoring	Fraud detection, customer churn prediction, disease risk prediction, climate modeling	Handles large datasets, improves accuracy, and reduces overfitting
Support Vector Machines (SVM)	Text classification, Bioinformatics, Image recognition, Marketing	Spam detection, protein classification, image classification, customer segmentation	Handles high-dimensional data, robust to noise, and efficient in high-dimensional spaces

Table2: Different sectors, applications, and benefits of each prediction techniques

VII. CONCLUSION AND FUTURE SCOPE

This survey paper has provided a comprehensive review of predictive analysis techniques, including traditional statistical methods, machine learning algorithms, and deep learning approaches. The techniques discussed in this paper, including regression, decision trees, random forests, neural networks, support vector machines, gradient boosting have shown impressive results in various applications. We have explored their applications, advantages, disadvantages and benefits. Developing predictive models that provide interpretable results and explanations, enabling trust and transparency in AI decision-making. The future of predictive analysis techniques holds immense promise, with potential applications in various fields, including healthcare, finance, marketing, and environmental sustainability. This paper opens a scope of development of new models for the task of predictive analytics.

REFERENCES

[1] Anitha K , Bhoomika C, J Andrea Kagoo, Kruthika K, Aruna Mg, "Student Grade Prediction Using Multiclass Model", August 2022, Ijirt, Volume 9 Issue 3 , Issn: 2349-6002.
[2] Eric Siegel, 2016, “Predictive Analytics”, John Willey and Sons Ltd.
[3] Rustagi, Mitanshi; Goel, Neha Predictive Analytics: A study of its Advantages and Applications IARS' International Research Journal, vol. 12, núm. 1, 2022
[4] Raheja, S. and Jain, D. (2021) “AN EMPIRICAL STUDY OF SELECTED INDIAN COMPANIES’ TALENT MANAGEMENT”, Amity Journal of Professional Practices. Florida, USA, 1(1). doi: 10.55054/ ajpp. v1i1.461.
[5] John Edwards,” Predictive analytics: Transforming data into future insights | CIO”, article - 16 AUGUST 2019 15:30 IST.
[6] Peter M Lee, 2012, “Bayesian Statistics: An Introduction, 4th Edition”, John Willey and Sons Ltd.
[7] Charles Elkan, 2013, “Predictive analytics and data mining”, University of California, San Diego.
[8] B Kaminski, M Jakubczyk, P Szufel, 2018, “A framework for sensitivity analysis of decision trees”, Central European Journal of Operations Research, Vol26, Issue-1, Pages-135-159.
[9] Ben Hur et al, 2001, “Support Vector Clustering”, Journal of Machine Learning Research, Vol-2, Pages125-137.
[10] Kam Cheong Li, Hoi Kuan Lam, Sylvia Siu-Yin LamA Review of Learning Analytics in Educational Research.Technology in Education. Technology-Mediated Proactive Learning, 559 (2015), pp. 173-184
[11] Vaibhav Kumar Predictive Analytics: A Review of Trends and Techniques July 2018 International Journal of Computer Applications 182(1):31-37

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details