

Predictive Analysis of Payment Delays

Chayan Sapra,

Department of Computer Science & Engineering,
Chandigarh University, Mohali

K Divya,

Department of Computer Science & Engineering,
Chandigarh University, Mohali

Kavita,

Uttaranchal University, Dehradun Email: kavita@ieee.org

Anil Kumar Dixit

Uttaranchal University, Dehradun Email: anildixit@uumail.in

Abstract—Predictive analytic could be a term that's largely applied to applied mathematics procedures. Statistics, machine learning, info approaches, and optimization techniques square measure all accustomed to produce this phrase. Its origin can be traced back to classical statistics. It forecasts the long run exploitation of each current and past knowledge. Prophetic analytic models will forecast future occurrences and therefore the behavior of variables. prophetic analytic strategies square measure accustomed to assign a score. The next score implies the next chance of an occurrence occurring, whereas a lower range shows a lesser chance of an occurrence occurring. These models use historical and transmutation knowledge patterns to supply solutions to a range of business and scientific issues. These models are helpful in assessing the risks and opportunities that every person faces. Using this method, I have developed a model by which I can predict the delay in actual payments by a client to any organization.

Index Terms— Predictive Analytics, Statistics, Machine Learning

I. INTRODUCTION

Predictive analytic, Extended Analysis Domain Branch is employed to predict future events. Analyzes current and his- torical knowledge and uses techniques from his- torical knowledge and uses techniques from statistics, data pro- cessing, machine learning, and computing to create predictions for the long run. This brings along data, technology, business modeling processes, and management to create predictions for the long run. With success applying prophetic analytic, firms will well use huge amounts of knowledge to get profit. It helps organizations become proactive at and absolutely predict trends and behaviors supported by knowledge. With the expan- sion of huge knowledge systems, it's grown up significantly. Suppose that the example of the E Retail company, ABC Inc. the corporate operates

Retail business worldwide and sells raw material to the big product companies and in real world there is delay in payments. I am trying to find the actual payment date using the previous data of the company and machine learning [1], [27].

II. LITERATURE REVIEW

Eric Siegel et al [1] Predictive analytic may be a term chiefly employed in applied mathematics and analytic tech- niques. It predicts the longer

term by analyzing current and historical knowledge. A long-term data is used to determine a score which is given by the machine. A high score means more accurate prediction and a lower score indicates a lower accuracy in the model. This paper also gives the use of predictive analysis.

Charles Nyce et al [2] Predictive analytic, in conjunction with most prognostic models and data processing techniques, admit more and more sophisticated applied mathematics strategies, together with statistical method techniques like advanced regression or time-series models. Such techniques modify organizations to work out trends and relationships that will not be without delay apparent, however still modify it to predict future events or behaviors.

M Nigrini et al [3] Forensic analytic is a branch of predictive analytics which uses the plethora of information of an individual or a group of peoples like an organization. By using the data, they can determine frauds, tax evasion, any criminal activities by studding the pattern of their banks and daily spending pattern and help Law.

M Schiff et al [4] Fraud: Insurance firms have long used data processing techniques to spot doubtless dishonest claims. Can be used in Security of a country. Determining medical fraud,

and catching any suspicious activity in the trading etc. V Dhar et al [5] Financial time-series prognostication is tough as a result of the inherently screeching nature of the domain. It's ordinarily best-known that the majority prognostication models do an awfully poor job in predicting future

returns.

B Kaminski et al [6] In this paper, the authors tend to consider successive call issues with uncertainty, described as call trees. Within the random model thought-about, the user typically has solely restricted data concerning various values of chances. There is a tendency to develop activity sensitivity analysis and, in this paper, we see how to tackle these problems.

J S Armstrong et al [7] "The Illusion of sure thing," proves that a multi variable analysis reduces the accuracy and cause a lot more confusion, and how to tackle this situation is written in this paper.

Peter M Lee et al [8] Bayesian Statistics is the college of thought that mixes previous beliefs with the chance of a

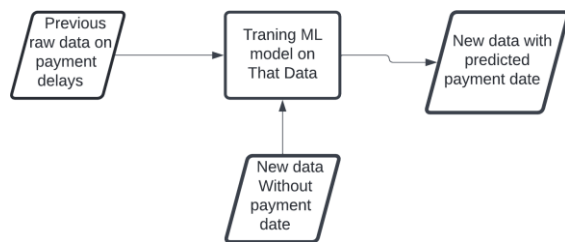


Fig. 1. 1: Flow of the process

hypothesis to hit posterior beliefs. The primary edition of Peter Lee's book appeared in 1989, however the topic has affected ever forward, with increasing stress on Monte Carlo based mostly techniques. This new fourth edition appearance at recent techniques like variation ways, theorem importance sampling, approximate theorem computation and Reversible Jump Mark off process Monte Carlo (RJMCMC), providing an elliptic account of the means during which the theorem approach to statistics develops in addition as however it contrasts with the traditional approach [24], [25]. The speculation is constructed step by step, and necessary notions like sufficiency area unit brought out of a discussion of the salient options of specific examples.

R Polikar et al [9] In matters of nice importance that have monetary, medical, social, or different implications, they regularly obtain a second

opinion before creating a call, typically a 3rd, and typically more. In doing so, they tend to weigh the individual opinions, and mix them through some thought method to succeed in a judicial decision that's pre-emptively the foremost abreast of one. the method of consulting "several experts" before creating a judicial decision is maybe habit to us; however, the in-depth advantages of such a method in machine-controlled higher cognitive process applications have solely recently been discovered by procedure intelligence service.

Jerome H. Friedman [10] Function estimation/approximation is viewed from the attitude of numerical improvement in the perform-house, instead of the parameter house. An affiliation is created between stage-wise additive expansions and steepest-descent reduction. Using gradient descent 'boosting' a para dime is developed which can support any fitting criterion . Connection between Huber M and boosting methods is mention in this paper.

C Cortes et al [11] The support-vector network may be a new learning machine for two-group classification issues. The vectors square measure is a nonlinear method and high dimension feature. All this is discussed in this research paper.

N S Altman et al [12] Non-parametric regression may be a set of techniques for estimating a regression line while not creating sturdy assumptions regarding the form of verity regression performed. These techniques square measure thus helpful for building and checking constant models, in addition as for information description [26]. Kernel and nearest-neighbor regression estimators square measure native versions of uni variate location estimators, then they'll promptly be introduced to starting students and consulting shoppers United Nations agency square measure accustomed to such summaries because the sample mean and median

Vaibhav Kumar et al [13] Predictive analytics is a term that is mostly applied to analytical and statistical procedures. ML, Data base approaches and other various techniques can be used to create this phrase. It forecasts the future using both current and past data. These models use previous and transferable data patterns to provide solutions to a variety of real-world problem mainly for business and scientific world.

J Lin et al [14] Numerous researches have considered using symbolic representative derived from real valued time series, this will help in the processing of bio informatics and text processing fields, as well as allowing the

streaming community to tackle previously "batch-only" problems.

Krishna Prakash Kalyanathaya et al [15] Engineers face significant problems and opportunities in adopting Artificial Intelligence as it pervades every aspect of our lives and businesses. Even though these models give more accurate results, ensembles such as Random Forest and Deep Learning algorithms exacerbate the problem in terms of explaining decision outcomes. We can no longer trust AI models' black box nature when we see the implications of their actions.

H. Naderpour et al [16] Ferrocement composites have lately become popular as an innovative approach for a variety of structural applications. This field for RC structures is not well defined due to a lack of substantial study in this method. The goal of this study was to use the GMDH method to determine the moment capacity of ferrocement members.

Hongren Shen et al [17] Business Analytics is one of the most essential components of a mix of technologies and skills which helps in analyze the organizational data to help in business plans and investment by predicting the future trends and market opportunities

Kasereka Henrys et al [18] Almost all of the organizations now days are leaning toward predictive analytic prediction techniques to determine future events which will help them to take critical decisions with which they can stay ahead of the competition. Many factors like ML model, Past data, Current scenario and market condition can help them to precisely predict the future events

Amarendra Mohanty et al [19] Predictive analytics refers to a set of methodologies that employ statistical and other empirical techniques to forecast future events based on past events. Predictive analytics can provide significant information to a supply chain company's management to help them make better decisions.

III. METHODOLOGY

A. Implementation

Training sets can considerably influence the dependability of associate analytical models, therefore getting a coaching dataset for machine learning based on a primarily based analytic process is critical. Prophetic analytic performed here, follows the four stages as delineated in 3 machine-learning algorithms: NB, SVM, and KNN enforced by this strategy, every trained on a dataset of gnomonic array.

Predictive analytic involves many steps through which a data analyst will predict the long run and support the present and historical information.

1) Requirement Collection: In order to develop a prediction model, we have to understand the requirement of our prediction on which type of data we want to generate a prediction model. And in order to do so we need to collect various information like datatype, key columns, dependent columns, Outcome table etc. All of these falls under Requirement collection.

2) Data Collection: After understanding the requirement, we proceed to collect the data, in order to predict anything accurately we need large sums of raw data. For Example, if we want to predict the delay in payments for a particular client we need their payment delay history so we can train the machine to predict their delays.

3) Data Analysis and Massaging: In order to use the prediction model, we have to convert the raw unstructured data into structured data which suits fit according to algorithm. Once the entire data is out there within the structured form, its quality is then tested. [22]. There might be some missing value or some repeated data, or some data with a different format we need to handle all this for accurate result before implementing the ML model.

4) Statistics, Machine Learning: The prognostic analytic method employs several applied math and machine learning techniques. applied mathematics and multivariate analysis at the foremost vital techniques that are popularly utilized in analytic [22]. These all models are supported by applied math.

5) Predictive Modeling: This model supports applied math and machine learning techniques and thus the instance data-set. When the event, it's taken a look at the test data-set, that may be a part of the most collected data-set, to see the validity of the model and if in, the model is alleged to be matched. One it is all completed the model will generate accurate predictions.

6) Prediction and Monitoring: After applying the model on some test cases, we cross examine the results with the actual results with actual data and determine the accuracy of the ML model. If the result is not accurate we make some changes in the model or try different model and we do this until we get required accuracy to use in our actual prediction.

IV. PREDICATIVE ANALYTIC OPPORTUNITIES We have been using the

predictive analytic for many years but nowadays is the era of prophetic analytic due to the advancement of technologies and dependency on knowledge [23]. Almost all of the organisations want to use predictive analytic to boost their profits. Some of the uses are listed below: -

- 1) Growth within the volume and kinds of knowledge is the reason to use prophetic analytic to realize insights from large-sized knowledge.
- 2) As technology becomes cheaper and fates it is more affordable for organization to implement predictive analytic.
- 3) Many open source codes are available to learn the methods and knowledge and many tech companies provide the services to help organization to grow using ML

The ever-changing world and changing market has motivated many companies to use prophetic analytic

With the advent of simple to use and interactive computer code and its convenience, prophetic analytic isn't being restricted to the statisticians and mathematicians. It's being employed in an exceedingly full swing by business analysts and also the social control call method.

Most common use of predictive analysis is listed below:

- 1) Police work and Fraud: Can be useful in detecting fraud based on pattern of actual fraud committers and accused. Also, they can study the crime rate and circumstance to find the root cause of any crime in any area. Threats might also be detected by applying this idea [3], [31].
- 2) Reduction of Risk: It can be used to calculate the risk while issuing credit-card or loans based on salary and CIBIL score. [4], [29].
- 3) Selling Campaign Optimization: To determine the best-selling products and determining the advertisements impact on sales. It helps the companies to draw in and retain the foremost profitable customers [22], [28].
- 4) Operation Improvement: Getting advance the inventory for a store by Predicting the future sales based on seasons, trend, sale and previous sales [5], [30].

V. CATEGORIES OF PREDICTIVE ANALYTIC MODELS

In general modeling applies on the basis of historical data to train the machine on.

The categories are listed as below: -

- Predictive Models: This model is used to determine the future outcome based on previous events and circumstances [14], [33].
- Descriptive Models: This model is used to explain entity relation and primarily used to explain what is a model doing [14], [34].
- Decision Models: the choice models delineated the link between the information, the choice, and therefore the results of the forecast of a choice. so as to form a prediction on the result of a choice that involves several variables [14], [32].

VI. PREDICTIVE ANALYTIC TECHNIQUES

All the prognostic analytic models square measure sorted into classification models and regression models. Classification Models predict the membership of values to an exact category whereas the regression models predict a range. Some of the popularly used techniques are listed below: -

A. Decision Tree

A decision tree may be a classification model; however, it is employed in regression additionally. It's a tree-like model that relates the selections and their \hat{N} consequences [6].

B. Time Series Analysis

Time series analysis could be an applied math technique that uses statistical information that is collected over a period of time at a specific interval. It combines the ancient information mining techniques and therefore the statement [15].

C. K-nearest neighbors (k-NN)

It is a non-parametric methodology used in classification and regression issues. During this methodology, the input contains the k nearest neighbour examples in a very feature area [12].

D. Bayesian Statistics

In this technique we evaluate parameters as random variable associate in using term "degree of belief" to determine the outcome [8].

E. Regression Model

Regression is one in all the foremost common applied math techniques that estimates the connection between variables. It evaluates connection between two variables specifically for freelance variables. It analyzes however the worth of variable changes on dynamic the values

of freelance variables in the sculptural relation [7].

VII. APPLICATION OF PREDICTIVE ANALYSIS

Their square measures several applications of prognostic analytic in an exceedingly sort of domain. From clinical call analysis to exchange prediction, wherever a wellness is foreseen, supported symptoms and come back on a stock, investment is calculable severally. I'll list out here below a number of the popular applications.

A. Financial and Banking services

In banking we can use the model to determine the future events to set the interest rates, loan rates. This can be used to determine rate of interests based on credit score, salary and type of job. For the money industries wherever cash is endowed in stocks or alternative assets, the prognostic analytic forecasts the come back on investments and helps in investment decision-making method [14].

B. Retail

By using this method one can determine which department of the store sells the most or which sector of a domain purchases which type of product and one can use the information to target specific groups of people for specific products. Asian paint is a living example for this method [24].

C. Other Sectors

The pharmaceutical sector uses prognostic analytic in drugs coming up with and up their supply chain of medication. The insurance sector uses prognostic analytic models to spot and predict the fraud claims filed by the shoppers. And using the medical history for a patient and their family medical history doctors can easily determine the major health risks for the patient [25].

VIII. CONCLUSION AND FUTURE SCOPE

These models can be used to predict the payments hence

helping an organization to maintain cash flow and cash reserve. And the same technique can be implemented in many fields. As it is used in many organizations now days to maintain cash flow. Companies uses predictive analysis to determine the actual payment date for their customer group and then they can use this information to make their payments. This Technique can also be used to determine the health of a person based on his physical activity, medical history and various

other factors. As the technology is becoming affordable many organization is adapting this methodology. Many companies have adapted this methodology to provide various services to different organizations. This analytics technique is very prominently useful for every part of the real word implementation and in future this analytical technique will be implement in many other fields and will help improve many fields of business, medical, banking, law and many aspects of human life.

REFERENCES

- [1] Hui Li, Hailing Zhang, Dan Liao, Xiaojie Zhu, Yueyue Dai, and Sahil Verma. 2022. A data sharing method for internet of drones based on federated learning. In Proceedings of the 5th International ACM Mobicom Workshop on Drone Assisted Wireless Communications for 5G and Beyond (DroneCom '22). Association for Computing Machinery, New York, NY, USA, 91–96. <https://doi.org/10.1145/3555661.3560870>
- [2] Mudit, K. Divya, S. K. Joshi and S. Verma, "Facial Expression Detection using Convolutional Neural Network," 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022, pp. 1-5, doi: 10.1109/ICCR56254.2022.9995993.
- [3] L. Singh, S. Khare, A. Parvez and S. Verma, "Research Paper on Path-finding Algorithm Visualizer," 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022, pp. 1-4, doi: 10.1109/ICCR56254.2022.9995925.
- [4] Zhihan Lv, Liang Qiao, Sahil Verma, and Kavita. 2021. AI-enabled IoT-Edge Data Analytics for Connected Living. ACM Trans. Internet Technol. 21, 4, Article 104 (November 2021), 20 pages. <https://doi.org/10.1145/3421510>
- [5] P. Rani, Kavita, S. Verma and G. N. Nguyen, "Mitigation of Black Hole and Gray Hole Attack Using Swarm Inspired Algorithm With Artificial Neural Network," in IEEE Access, vol. 8, pp. 121755-121764, 2020, doi: 10.1109/ACCESS.2020.3004692.B.
- [6] Kamin'ski, M. Jakubczyk, and P. Szufel, "A framework for sensitivity analysis of decision trees," Central European journal of operations research, vol. 26, pp. 135–159, 2018.
- [7] M. Kumar, P. Mukherjee, K. Verma, S. Verma and D. B. Rawat, "Improved Deep Convolutional Neural Network Based Malicious Node Detection and Energy-Efficient Data Transmission in Wireless Sensor Networks," in IEEE Transactions on Network Science and

Engineering, vol. 9, no. 5, pp. 3272-3281, 1 Sept.-Oct. 2022, doi: 10.1109/TNSE.2021.3098011.

[8] Sahil Verma, Sonu Mittal, "Implementation and Comparative Analysis of Speed and Distance Beaconing Schemes for VANET", in *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 14, No. 12, December 2016, (pp. 362-370).

[9] Sahil Verma et al., "Ideas on collision detection in vehicles using WSN" *International Journal of Recent Research Aspects*, Vol. 3, Issue 3, September 2016, pp. 33-37.

[10] Kavita, Sandip Kumar Goyal, Sahil Verma, "A Fault Tolerant Approach For Load Balancing In Grid Environment" in *International Journal of Engineering Research & Technology*, Vol. 1 Issue 9, November- 2012, pp. 1-6.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273-297, 1995.

[12] N. S. Altman, "An introduction to kernel and nearest-neighbor non-parametric regression," *The American Statistician*, vol. 46, pp. 175-185, 1992.

[13] V. Kumar and M. L. Garg, "Predictive analytics: a review of trends and techniques," *International Journal of Computer Applications*, vol. 182, pp. 31-37, 2018.

[14] J. Lin, "A symbolic representation of time series, with implications for streaming algorithms," *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003.

[15] K. Kalyanathaya and Prakash, 2022.

[16] H. Naderpour, "A new proposed approach for moment capacity estimation of ferrocement members using Group Method of Data Handling," *International Journal*, vol. 23, pp. 382-391, 2020.

[17] Shen, B. A. Hongren, J. Tolson, and Mai, "Time to Update the Split-Sample Approach in Hydrological Model Calibration," *Water Resources Research*, vol. 58, pp. 2021-031 523, 2022.

[18] A. Mohanty and Ranjana.

[19] I. A. Shah, Q. Sial, N. Z. Jhanjhi, and L. Gaur, "The Role of the IoT and Digital Twin in the Healthcare Digitalization Process: IoT and Digital Twin in the Healthcare Digitalization Process," *Digital Twins and Healthcare: Trends, Techniques, and Challenges*, pp. 20-34, 2023.

[20] N. Z. Jhanjhi, S. N. Brohi, N. A. Malik, and M. Humayun, "Proposing a hybrid rpl protocol for rank and wormhole attack mitigation using machine learning," 2020 2nd International Conference on Computer and Information Sciences (ICCIS), pp. 1-6, 2020.

[21] K. Hussain, S. J. Hussain, N. Jhanjhi, and M. Humayun, "SYN Flood Attack Detection based on Bayes Estimator (SFADBE) For MANET," in 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp. 1-4.

[22] I. A. Shah, Q. Sial, N. Z. Jhanjhi, and L. Gaur, "Use Cases for Digital Twin," *Digital Twins and Healthcare: Trends, Techniques, and Challenges*, pp. 102-118, 2023.

[23] N. Conz, "Insurers shift to customer-focused predictive analytics technologies," *Insurance & Technology*, 2008.

[24] B. Vijayalakshmi, K. Ramar, and N. Jhanjhi, "An attention-based deep learning model for traffic flow prediction using spatiotemporal features towards sustainable smart city," *Int J Commun Syst*, vol. 34, 2021.

[25] A. Hussain, "A Resource Efficient hybrid Proxy Mobile IPv6 extension for Next Generation IoT Networks," *IEEE Internet of Things Journal*.

[26] I. Batra, "Hybrid Logical Security Framework for Privacy Preservation in the Green Internet of Things," *Sustainability*, vol. 12, pp. 5542-5542, 2020.

[27] S. More, "Security Assured CNN-Based Model for Reconstruction of Medical Images on the Internet of Healthcare Things," *IEEE Access*, vol. 8, pp. 126 333-126 346, 2020.

[28] V. Singhal, "Artificial Intelligence Enabled Road Vehicle-Train Collision Risk Assessment Framework for Unmanned Railway Level Crossings," *IEEE Access*, vol. 8, pp. 113 790-113 806, 2020.

[29] M. Arora, "A Systematic Literature Review of Machine Learning Estimation Approaches in Scrum Projects," in *Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing*, vol. 1040. Springer, 2020.

[30] M. Kumar, P. Mukherjee, K. Verma, S. Verma, and D. B. Rawat, "Improved Deep Convolutional Neural Network based Malicious Node Detection and Energy-Efficient Data Transmission in Wireless Sensor Networks," *IEEE Transactions on Network Science and Engineering*.

[31] K. Srinivasan, L. Garg, D. Datta, A. A. Alaboudi, N. Z. Jhanjhi,

(Doi :
<https://doi.org/10.54216/JISIoT.010201>)

R. Agarwal, and A. G. Thomas, "Performance comparison of deep cnn models for detecting driver's distraction," *Materials & Continua*, vol. 68, no. 3, pp. 4109–4124, 2021.

[32] A. Almusaylim, Z. Jhanjhi, N. Z. Alhumam, and A, "Detection and mitigation of RPL rank and version number attacks in the internet of things: SRPL-RP," *Sensors*, vol. 20, no. 21, pp. 5997–5997, 2020.

[33] N. Kaur, D. Gupta, R. Singla, A. Bharadwaj, W. Kavita, and Mansoor, "Thermal Aware Routing Protocols in WBAN," 2021 4th International Conference on Signal Processing and Information Security (ICSPIS), pp. 80–83, 2021.

[34] Ibrahim M. EL-Hasnony, Intelligent Differential Evolution based Feature Selection with Deep Neural Network for Intrusion Detection in Wireless Sensor Networks, *Journal of Intelligent Systems and Internet of Things*, Vol. 0 , No. 2 , (2019) : 78-89 (Doi : <https://doi.org/10.54216/JISIoT.000204>)

[35] Hatem Abdul-Kader , Mustafa.Abd-El salam , Mona Mohamed, Hybrid Machine Learning Model for Rainfall Forecasting, *Journal of Intelligent Systems and Internet of Things*, Vol. 1 , No. 1 , (2020) : 5-12 (Doi : <https://doi.org/10.54216/JISIoT.010101>)

[36] Mona Mohamed, A comparative study on Internet of Things (IoT): Frameworks, Tools, Applications and Future directions, *Journal of Intelligent Systems and Internet of Things*, Vol. 1 , No. 1 , (2020) : 13-39 (Doi : <https://doi.org/10.54216/JISIoT.010102>)

[37] Mahmoud H.Alnamoly , Ahmed M. Alzohairy , Ibrahim M. El-Henawy, A survey on gel images analysis software tools, *Journal of Intelligent Systems and Internet of Things*, Vol. 1 , No. 1 , (2021) : 40-47 (Doi : <https://doi.org/10.54216/JISIoT.010103>)

[38] Abdel Nasser H. Zaied, Mahmoud Ismail and Salwa El- Sayed*, A Survey on Meta-heuristic Algorithms for Global Optimization Problems, *Journal of Intelligent Systems and Internet of Things*, Vol. 1 , No. 1 , (2020) : 48-60 (Doi : <https://doi.org/10.54216/JISIoT.010104>)

[39] K. Shankar, Improving the Security and Authentication of the Cloud with IoT using Hybrid Optimization Based Quantum Hash Function, *Journal of Intelligent Systems and Internet of Things*, Vol. 1 , No. 2 , (2020) : 61-71