LITERATURE REVIEW:
MODELS OF PREDICTIVE ANALYSIS AND ITS APPLICATIONS IN MANUFACTURING, FINANCE AND RETAIL SECTOR

PROJECT REPORT SUBMITTED FOR
FINAL ASSSESMENT OF COMP516
(RESEARCH METHODS IN COMPUTER SCIENCE)

BY
SHANKAR NARAYANAN VENKADASUBBU
OLUWOLE RUTH
SRI HARSHA VISHNUVAJHALA
SANTHOSH ARUNAGIRI
ONUOHA JOY
ADEGUNLE BUKOLA
IFEANYI CYNTHIA

UNIVERSITY OF
LIVERPOOL

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF LIVERPOOL
LIVERPOOL L69 3BX

TABLE OF CONTENTS

ABSTRACT

The study that follows examines efficient prediction models and how they are applied in three important sectors: manufacturing, retail, and finance. Decision trees, linear regression, logistic regression, and time series models are among the models that were considered. The paper delves deeply into how decision trees and time series models are implemented and found to be effective to predict customer retention, customer loyalty, improving supply chains, and financial performance in the mentioned industries. Correspondingly, traditional models, such as linear regression and logistic regression, are found to be still effective, in fact on par with current popular models, such as ANN, Neutral Networks.

1. INTRODUCTION

Predictive analysis cannot be discussed without first referring to Big Data and this is in fact a subtopic of Big Data. Organisations store and process large amount of data, take the bank for instance who has information on its customers ranging from their names, age, address, jobs, incomes, products, transaction etc. With such data a lot of information can be derived to make future forecast and plans thus giving organisations good opportunities and advantages[14].

Big Data is large in Volume, has good Value, Variety and a high speed of accumulation known as velocity that it becomes impossible to analyse such data using traditional method of analysis. The use of statistical tools and predictive models to extractive value from such data is what we know as Predictive Analysis.

The aim of this paper is to explore this emerging field in computer science, discussing some selected predictive models such Linear Regression, Logistic Regression, Decision Tree, Time Series, and their applications to areas like Health, Finance, Manufacturing and Retail. Some selected papers will be reviewed noting the challenges and important development in this field.

1.1. Predictive analysis

This is the use of statistical tools and predictive models to make future forecast and predictions from historical data. Organisations like insurance and marketing can use this technique to make informed business decisions to boost and improve sales and customer retention. The ability to look at past and present patterns and recognised the possibility of reoccurrence allows businesses to make productive use of resources to maximize profit, improves targeted marketing and advertising, cash flow prediction, fraud detection and predicting patient pattern and disease diagnosis.[14]

Predictive Models includes Linear Regression, Logistic Regression, Time Series, Decision Tree, Neural Network, and some of this will be discussed in detail in this paper.

1.2.   Importance of predictive Analysis
1.2.1. Informed Decision Making in Business:

Predictive analytics can offer knowledge to guide decision-making and give an advantage over competitors. Making important decisions is a crucial part of running a successful business. Any type of growth, including the introduction of a new product or service, involves assessing the risks involved with the possible results.[14]

1.2.2. Customer Profiling and Segmentation:

Predictive Analysis help us organize our data in a way that can be used to grow sales through targeted marketing. Opportunities are quickly identified resulting in successful customer engagement

1.2.3. Cash Flow Prediction:

With the use of predictive analysis, industries can build models to predict cash follow using historical and relevant data thus making forecast of future cash flow to help with business planning and decision making.

1.2.4.   Fraud Detection:

With the use of predictive models like the Logistic Regression organisation can predict fraud by feeding some customers historical and relevant data as variables into the model to output a binary variable of 1 if activity is fraudulent or 0 if it's not fraudulent.[13]

1.2.5. Predicting Patient Patterns and Disease Diagnosis:

We can understand and make diagnosis of Diseases using patient patterns in predictive analysis.

1.3.   Predictive Analytical Models.

Predictive modelling is an arithmetic process used in predicting future occurrence by analysing patterns in a given set of input data. It is an important predictive analytics tool that tries or attempts to respond to the query, "What might possibly occur in the future?.[10]

Predictive modelling is widely used in businesses to improve the likelihood of predicting events, consumer behaviour, financial, economic, and market hazards, as well as to understand the relationships between various types of data. Due to the plethora of huge and unstructured data available, these predictive model tools are frequently applied via computer software applications.[1]

1.3.1. Linear regression Analysis:

This is a predictive model that uses a straight line to predict the value of a target variable based on given predictor variable(s). It is a correlation between the variables affecting the input and output.[10] To forecast the result of a particular event, this model just needs one predictor or feature. A real-life example is that of a company that has been recording an increase in sales from the beginning of the year, this company can determine or predict their future sales for the rest of the year by conducting a linear analysis of sales data using their monthly sales.[18]

1.3.2. Decision Tress:

Decision Trees are simple predictive models that are particularly useful in making decisions within a short time. It employs branch logic and a graphical representation to forecast all potential outcomes in light of certain circumstances.[12] There is an internal node which represents attributes and branch representing the outcome then the trees depict decision that was made after the computation of the attribute. For instance, with the use of a decision tree, management will be able to decide whether to invest in a new machine or upgrade the existing ones in a manufacturing industry, considering that upgrading the existing machines will save them money at the time while the new ones will cost more but will increase productivity. thus, Management will be able to make decisions that are in the best interests of the business while considering productivity and financial implications.[18]

1.3.3. Logistics Regression:

Also known as logit model is a type of statistical model that is commonly used in categorisation and predictive analytics; This model takes into consideration the likelihood that an event will occur based on a given collection of independent variables.[12] The range of the independent variable is 0 to 1. The odds, which are the likelihood of success divided by the probability of failure are typically subject to a logit transformation. It is used to find data anomalies that are indicative of fraud in fraud detection. It is also used in disease prediction in predicting the likelihood of diseases or illness in a given population.[18]

1.3.4. Time Series:

This prediction model examines data points gathered over a period of time to analyse how variables change over time. Data is gathered on an hourly, daily, weekly, monthly, annual, biannual, or yearly basis etc, and regularly recorded to demonstrate how data changes throughout the course of data points as well as the outcomes to verify the accuracy and consistency of the data acquired.[12] For time series analysis, many different data sets are needed. predictive time series modelling analysis can be used to predict changes in stock prices, industry forecasts, heart rate monitoring (EKG), quarterly sales, and brain monitoring (EEG).

2. AIM

The objective of this paper is to understand and compare widely used predictive models and further review important developments in predictive analysis in key industries namely manufacturing, retail, finance sectors.

3. OBJECTIVE

The goal of predictive analysis is to draw conclusions about potential future events from existing data by using traditional models like linear regression, logistic regression, decision tree and time series. These models aid them in planning their future and making crucial business decisions even though we have got many modern techniques. Our review mainly concentrates on usefulness of traditional models.

4. RETAIL

In the retail industry, predictive analysis can be used to better understand consumer behaviour, such as shopping habits. It allows the businesses the ability to examine and develop client suggestions based on their past behaviour as well as anticipate prospective actions that can improve customer retention. Retailers may focus on the most popular trends right now with the aid of these enormous datasets to improve their marketing choices.[15]

The gathering of data will be one of the major components of using a predictive analytics model to enhance a retail environment. These are typically gathered through participation in various surveys or coupon registrations. Although these days, cookies are used to make these data more broadly accessible to businesses.

4.1. Benefits of Predictive Analysis in Retail

Some of the key advantages of predictive analysis are;

1) It takes customer personalization to the next level.
2) New customer preferences and demands can be forecasted beforehand and thereby enabling the retailers to always stay ahead.
3) Companies are now able to generate new insights about customers by analysing their behaviour.[17]

Lee, S *et al.*[11] stated that, by employing logistic regression, their research concentrated on key factors from a wide range of products, textual qualities, etc. The decision criteria for judging helpfulness were generated by the decision tree approach based on the worth of the product data, review features, and textual characteristics. It was acknowledged that CRDT's prediction accuracy outperformed that of k-nearest neighbours. The key elements centre on the idea that more beneficial online user reviews may potentially offer buyers more value. Their study's findings advised managers to give customers better diagnostics to better comprehend customer feedback and increase their value.

According to the study, the decision tree was chosen because it is more suited for analysing complicated linkages and articulating intricate webs of connections. This study develops a decision-tree-based prediction model down to 10 levels that offers great predictive power, enabling them to anticipate with high accuracy how customers are likely to judge a review, allowing consumers and others to successfully find beneficial online reviews.[15]

Rose Tinabo *et al.*[19] reported that customer retention can be characterised as their dedication, trust, and other characteristics that influence their willingness to make a purchase. This is a crucial component of every organisation since, depending on the cost of achievement, even a small increase can provide as much as 80% of profit. On the other hand, even losing one or two customers can have a significant impact on a company's revenue.

In their study, four potential models were analysed, and decision trees were suggested as the most efficient method. This was based on the broad attributes of the retail information, such as its scope and size, which were primarily used to describe consumer behaviour and help businesses decide how to keep customers. Additionally, it was said that the accuracy of the outcomes might be increased by employing organization-specific datasets, which are not accessible to the public for research.[4]

Hirche, M. *et al.*[7] studied the effect of temperature and holidays on the retail sales of various alcoholic beverages. The countries with the highest average temperatures are represented by their first sub-sample, while the counties with the lowest average temperatures are represented by their second sub-sample.

Time series modelling is used in this study as the foundation for forecasting and using specific approaches that would help marketers plan more effectively for both short-term and long-term sales campaigns. Distributors may predict the anticipated demand for their products based on numerous qualities by incorporating meteorological, regional, and holiday data.

Their results included Easter weekend which saw an increase in all categories, especially sparkling wine (+23.1%) and other wine (+18.2%), according to their findings. Additionally, it was mentioned that beer sales consistently rise over Memorial Day weekend (+15.4%). Beer and liquor sales both increased on July 4th by 22% and 23.7%, respectively

5. FINANCE

Predictive analytics is a growing area of interest in the finance sector. It provides suggestions as to how companies can improve their, stock performance, budgeting, reporting, sales, investment, cash flow management and gives insight to possible changes in stock behaviour as well as answers to all the future related investments. These can be used to predict future returns.[8]

Time-series analysis is a predictive analytics model. It analyses data using historical data. It focuses the data analysis on the assumption that the observation at a point of time depends on previous observations in time. It requires a prudent investigation of critical patterns and trends before further analysis is being carried out. It is important to know that there are different factors affecting decisions (time series data) at a point in time and this helps to draw conclusion on the right model to use for data analysis.

In the finance industry, time series data analysis, Logical regression and Decision tree Model are part of the models used to analyse a company's performance over a period of time. It helps company/investors to forecast market trends and behaviours, and this may result in gaining more profits than usual if successful as a result organization understand the reasons or causes of trends and the systematic pattern in each observation.[9]

A.Chakraborti *et al.*[5] states that without time series analysis it would be impossible to predict previous price movements and trends. In financial time series analysis, the long-term correlation between variables is observed and thus form the basis of forecasting the likelihood of an occurrence.

Muhammed Mubeen *et al.*[2] shows that the Logistic regression model can be used to forecast future stock performance and how the causes and trends in financial ratio can predict stock performance Using different financial accounting ratios as independent variables and good/bad stock performance as the dependent variable, they examined the trends in sales, causes of increase/decrease in Earnings Per share, trends in Debt-to-Equity ratio, return on equity and their findings indicated 88.7% accuracy in predicting good or bad performance level of stock. This shows that logistic regression model can help identify anomalies in forecast and also estimate the probability of a good stock or bad stock performance and this help increase the predicting power of companies/investors regarding stock price movement.

Roy Rada *et al.*[16] researched how decision tree model can be used on certain variables to make investment decisions. A decision tree algorithm was applied to different financial variables and the findings shows the accuracy of some variables used by decision tree are different from those expected. It was suggested that more financial variables (such as Total Assets on the statement of financial

position) should be used for stock evaluation as this gives a more accurate results as against the normal financial variables used for financial stock evaluation.

6. MANUFACTURING SECTOR

Predictive analysis is used in the manufacturing sector to ensure effectiveness of production machines, maintain, and improve quality of products. The manufacturing industry deals with huge amounts of data which are generated from production devices and with the growing use of smart devices in various firms, huge volumes of data will continue to be generated.[3] The innovative use of generated data will make the industry more sustainable and profitable.[18]

A lot of manufacturing firms are however still struggling with processing huge volumes of data. The World Economic forum reported that in a 2021 study, surveying over 1300 manufacturing executives, it was revealed that just 39% had successfully scaled data-driven use cases beyond the production process of a single product and thus achieved a clearly positive business case.[18]

Predictive maintenance is one of the ways predictive analyses is deployed in the manufacturing sector.[3] The efficiency of a firm's production machines is a critical factor in managing cost, managing production levels, and keeping up with the market opportunities.[14] Predictive maintenance involves setting up systems that can allow early detection of machine failures based on historical data. Statistical inference methods are used to estimate the health status of machines, amongst which machine learning is widely used.[15]

Susto *et al.*[6] quoted that regression-based formulations are deployed to predict the Remaining Useful Life of an equipment and classification-based formulations for distinguishing between healthy and unhealthy conditions of systems. In 2015, they developed a machine learning methodology, called the multiplier classifier methodology for predictive maintenance to tackle the challenge of health factors associated with classification tools. The multiple classifier methodology allows for multiple logistic variables to be used at each process iteration.

In 2019, **Borucka** *et al.*[4] studied the application of logistic regression for production machinery efficiency and the subject of their research was a plastics manufacturing company. Possible explanatory variables were identified for the model which included shift, device, occurrence of failure (yes or no) and no production order (yes or no). These variables were subjected to statistical analysis and then compared to the dependent variable.

With the available data, logistic regression method was more suitable as predictor was a dichotomous variable and its values could be interpreted as the probability of an event occurring. The method made it possible to identify the causes influencing machine efficiency which in this case study was the shift system and the workload on individual machines.

Predictive analysis can also be deployed for maintaining quality standards and to aid effective planning in manufacturing.

In 2019, **Zalan Mahmood** *et al.*[20] created a decision-tree-based ontology model to assist in the allocation of newly received orders in a manufacturing network. An Italian based project management company (managing a network of 300 small manufacturing firms) was studied who had difficulty in scheduling newly received orders to firms who could deliver quality products on time.

They had found the decision tree model and ontology-based clustering to be more efficient and used WEKA software classifier algorithm to make a predictive decision tree and then mapped into the

ontology model. This model helped by accelerating planning and scheduling of tasks within the manufacturing network. Their model resulted in 60.4% accuracy using 8 suppliers as the sample size.

Classification models such as decision tree and Logistic regression models are predictive models used in the manufacturing industry to apply insights in taking better decisions and to aid proper planning. The classification models appear to be widely used based on the literature reviewed in this sector. Further studies can be done to improve on the traditional models, one of which involves the automation of some data processing phases.

7. OBSERVATION

It is evident that many sectors now use predictive analysis as a key component in order to make vital decisions and advances. For many situations, it has been demonstrated that conventional models like logistic regression and linear regression, when combined with the right variables and a small dataset, are still useful. Furthermore, a time series model has been effective at forecasting events from non-stationary data, while decision tree models have been shown to yield the best results in applications that required classifying the dataset with complex relationships.

8. CHALLENGES IN USING PREDICTIVE ANALYSIS MODELS

Implementing predictive analysis models can be tedious and time consuming. Most of the difficulties may arise from the quality of data, inadequate data, voluminous data, or the choice of the model itself.

Borucka *et al.*[4] noted that the lack of adequate and modern IT systems in some organizations to archive data from ongoing processes during production and the low quality of recorded documentation limits the use of mathematical tools.

Other areas such as anonymous customer reviews and information analysis, unstructured data from social media, low completion and quality of feedback surveys are also a big challenge for predictive analysis.

9. REFERENCES

1. A. Azhar, "What is predictive modeling?," TechRepublic, Oct. 26, 2022. https://www.techrepublic.com/article/what-is-predictive-modeling/ (accessed Dec. 11, 2022).
2. Ali, S. S.., Mubeen, M., Lal, I. ., & Hussain, A. . (2018). Prediction of Stock Performance by Using Logistic Regression Model: Evidence from Pakistan Stock Exchange (PSX). Asian Journal of Empirical Research, 8(7), 247–258. https://doi.org/10.18488/journal.1007/2018.8.7/1007.7.247.25
3. Allied Market Research. (n.d.). Manufacturing Predictive Analytics Market by Component (Software and Services), Deployment (Cloud and On-premise), Application (Demand Forecasting, Machinery Inspection & Maintenance, Product Development, Supply Chain Management, and Others), and End User (Semiconductor & Electronics, Energy & Power, Pharmaceutical, Automobile, Heavy Metal & Machine Manufacturing, and Others): Global Opportunity Analysis and Industry Forecast, 2019–2026.https://www.alliedmarketresearch.com/manufacturing-predictive-analytics-market [Accessed 10 Dec. 2022].World Economic Forum. (n.d.). The future of manufacturing is powered by data and analytics. Here's why. [online] Available at: https://www.weforum.org/agenda/2022/09/manufacturing-data-advanced-analytics/.

4. Borucka A and Grzelak M 2019 Application of Logistic Regression for Production Machinery Efficiency Evaluation Applied Sciences **9** 4770 Online: http://dx.doi.org/10.3390/app9224770

5. Chakraborti, A., Patriarca, M., Santhanam, M.S. (2007). Financial Time-series Analysis: a Brief Overview. In: Chatterjee, A., Chakrabarti, B.K. (eds) Econophysics of Markets and Business Networks. New Economic Windows. Springer, Milano. https://doi.org/10.1007/978-88-470-0665-2_4

6. Gian Susto, Andrea Schirru, Simone Pampuri, Sean McLoone & Alessandro Beghi (2015) Machine Learning for Predictive Maintenance: A Multiple Classifier Approach, IEEE Transactions on Industrial Informatics, 11:3, 812-820, DOI: 10.1109/TII.2014.2349359

7. Hirche, M., Haensch, J. and Lockshin, L. (2021), "Comparing the day temperature and holiday effects on retail sales of alcoholic beverages – a time-series analysis", International Journal of Wine Business Research, Vol. 33 No. 3, pp. 432-455. https://doi.org/10.1108/IJWBR-07-2020-0035.

8. Indriasari, Elisa & Lumban Gaol, Ford & Soeparno, Haryono. (2019). Application of Predictive Analytics at Financial Institutions: A Systematic Literature Review. 10.1109/IIAI-AAI.2019.00178.

9. Kirlić, Ajla & Hadžić, Muhedin. (2017). Big data and time series: A literature review paper. Univerzitetska misao - casopis za nauku, kulturu i umjetnost, Novi Pazar. 2017. 139-146. 10.5937/univmis1716139K.

10. Lawton, G., Carew, J. and Burns, E. (2018). What is predictive analysis. [online] tech target.com. Available at: tech target.com/search enterpriseai/definition/predictive-modelling.

11. Lee, S. and Choeh, J.Y. (2017), "Exploring the determinants of and predicting the helpfulness of online user reviews using decision trees", Management Decision, Vol. 55 No. 4, pp. 681-700. https://doi.org/10.1108/MD-06-2016-0398

12. M. Mitchell, "Selecting the Correct Predictive Modeling Technique," Medium, Aug. 26, 2019. https://towardsdatascience.com/selecting-the-correct-predictive-modeling-technique-ba459c370d59

13. M. R. Sumalatha and M. Prabha, "Mediclaim Fraud Detection and Management Using Predictive Analytics," *2019 International Conference on Computational Intelligence and Knowledge Economy* (2019): 517-522

14. Mishra, Nishchol, and Sanjay Silakari. "Predictive analytics: a survey, trends, applications, oppurtunities & challenges." International Journal of Computer Science and Information Technologies 3.3 (2012): 4434-4438

15. R. Barasch, "The power of retail analytics," Yodlee.com, Jan. 14, 2019. https://www.yodlee.com/data-analytics/big-data-retail-analytics

16. Rada, R., & Wimmer, H. (2017). Decision Trees and Financial Variables. International Journal of Decision Support System Technology (IJDSST), 9(1), 1-15. http://doi.org/10.4018/IJDSST.2017010101

17. "Retail Revolution: The Role of Big Data Analytics in Retail," Hitachi Solutions, Jan. 06, 2022. https://global.hitachi-solutions.com/blog/retail-analytics/

18. Roychowdhary, A. (2022). Predictive modelling techniques-A comprehensive guide. [online] Projectpro.io. Availableat:Projectpro.io/article/predictive-modelling-techniques/598.

19. Tinabo, R. (2011). Decision Tree Technique for Customer Retention in Retail Sector. In: Hruschka, E.R., Watada, J., do Carmo Nicoletti, M. (eds) Integrated Computing Technology. INTECH 2011. Communications in Computer and Information Science, vol 165. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22247-4_11.

20. Zalan Mahmood Ayaz Khan, Salman Saeidlou & Mozafar Saadat (2019) Ontology-based decision tree model for prediction in a manufacturing network, Production & Manufacturing Research, 7:1, 335-349, DOI: 10.1080/21693277.2019.1621228

LITERATURE REVIEW:
MODELS OF PREDICTIVE ANALYSIS AND ITS APPLICATIONS IN MANUFACTURING, FINANCE AND RETAIL SECTOR

Accompanying Report

Introduction

Why Predictive Analysis?

After opting to move on with a literature review, our team investigated every potential subject that could influence numerous sectors right now. We made the decision to analyze papers on predictive analysis using conventional models in the retail, manufacturing, and financial industries after considering our options.

We picked traditional models such as logistic regression, linear regression and popular models like decision tree model, time series model and researched how these models are being used across manufacturing, finance, and retail sectors. To our realization, we saw insights where traditional models outperformed current widely used models like Neural network, ANN's in various sectors in some use cases. This was our trigger point to understand more about traditional models in predictive analysis and how they are implemented in key industries. We included decision tree model and time series model as they were the most used techniques for predicting events in all these industries.

Team and Work Delegation

The main objective of our project was to find insights and learn about how the four models are being used in the selected sectors. All our team members contributed equal efforts to this project. We made sure of this by splitting the workload equally between ourselves. For instance, we paired two members to analyze one sector. At the end of each week, we organized a meeting (both online and offline) to share the insights gained from a week's research. This helped us to track the workflow efficiently and ensure that there are no delays in the completion of the project. Once our presentation was completed on the 5th week, and after getting the much-needed feedback, we were able to fine tune our progress.

To make sure that everyone understood the entire project and that members weren't only limited to their sections, we also held several brainstorming sessions and asked various questions to one another. Each of us made an equal contribution to each section of the final literature review report, including organising the references and making sure the citations were accurate and in the right sequence.

Challenges

Determining whether to include analysis pertaining to the healthcare sector or not was one of the main issues we encountered. Here, the size of the pool of research publications at our disposal posed the biggest challenge. But more importantly, each model was outperforming itself in some use cases compared to other use cases. These factors made it very difficult for our team to provide any specific insights or talking points. Therefore, we made the decision to concentrate closely on the final three sections.

Our initial goal was to determine which among these predictive models works best in a certain sector. After compiling enough data, we realized that each predictive model has its own set of use cases and may not be used to address a particular issue. For instance, we discovered insights in the retail industry where a decision tree and time series model are employed to address different use case, one model was used to predict helpfulness of a product review and the latter one was used to predict whether temperature has impact on sales of alcoholic beverages in retail stores. This made it difficult for us to understand which models are best suited for a particular industry.

How we overcame this challenge?

To solve this issue, we chose to refocus our efforts on locating academic works in which these conventional models are contrasted with recently popular models like neural networks, Ann. We discovered that older models performed better and generated results with comparable accuracy in several specific use scenarios. This helped us to understand the context in which a given model can be used.

Conclusion and Future Scope

Following this review, the team was able to draw a conclusion with reference to use cases in which traditional models performed on par with or better than more modern, sophisticated models. These use cases were successfully analysed, and each one is described in the literature review report.

Depending on where or how they are employed, this literature review report demonstrates that traditional models do have a bright future. In order to boost the effectiveness and potential of these models, scenarios like those described in our report can be useful learning tools.