A Comparative Analysis of Four Machine Learning Algorithms to Predict Product Sales for a Retail Store

By

Kenneth Ofoegbu

Department of Data Analytics

Dublin Business School

Supervised by

Dr. Amir Sajad Esmaeily

Declaration

I Kenneth Ofoegbu hereby declare that, except where specific references were made to other works, the content of this dissertation is original and submitted in whole for the award of master's in science in Dublin Business School and has never been submitted to any other institution for consideration. This dissertation covers only work completed by me without any other person or institution's cooperation and is in full compliance with the Dublin Business School's academic honesty policy.

Kenneth Ofoegbu

January 202

Acknowledgment

To Dr. Amir Sajad Esmaeily, my research supervisor, I want to express my sincere and warm gratitude for his patient support, for his excellent and positive recommendations in the planning and improvement of this research and his capacity to invest his time generously. Finally, I would like to thank my family for their support and motivation throughout the course.

Abstract

Controlling the retail market is the secret to sustainability in today's business world. Many business entities depend heavily on historical data and product demand projection of sales patterns. The accuracy of these projections has a significant effect on business. Data mining techniques are effective tools for retrieving concealed information from large datasets to increase predictions' precision and reliability. The systematic research and review of comprehensible machine learning classification models to boost product sales predictions are carried out in this work. The traditional statistical forecasting/predicting methods are challenging to cope with big data and accuracy in predicting product sales. However, these problems can be addressed through the use of different data mining and machine learning techniques. In this work, we briefly analyzed sales data and the prediction of product sales. Various techniques used in machine learning and data mining are discussed in the latter part of the research. For the performance evaluation, the best-suited classification model is proposed for the product sales type prediction. The findings are presented in terms of the reliability and accuracy of the different prediction algorithms used. The study shows that the best fit model is Random Forest, which produced the prediction's highest accuracy.

Table of content page

Declaration……………………………………………………………………………………………….2

Acknowledgment ……………………………………………………………………………………...3

Abstract……………………………………………………………………………………………………4

List of Figures ………………………...……………………...………………………………………….7

List of Tables ……………………………………………………………………………………………8

List of Algorithms……………………………………………………………………………………...9

Chapter 1……………………………………………………………………………………………...8

Introduction…………………………………………………………………………………...11

1.2 Research background …………………………………………………………..……………………12

1.3 Research aim and objectives…………………………………………………………………….…13

1.4 Research questions…………………………………………………………………………………...13

1.5 Why do forecasting ………………………………………………………………………………...14

1.6 Method of forecasting …………………………………………………………………………....16

1.7 Statistical sales forecasting methods……………………………………………………………18

1.8 AI sales forecasting methods……………………………………………………………………....19

1.9 Hybrid method of sales forecasting ……………………………………………………………20

1.10 Ensuring an accurate forecast……………………………………………………………..…..23

1.11 A successful sales forecast ……………………………………………………………………...24

1.12 Research structure……………………………………………………………………………...…26

Chapter 2…………………………………………………………………………………...……27

2.1 Literature review……27

Chapter 3……29

3.1 Methodology ……36

3.2 CRPIS-DM Methodology……36

3.2.1: The Generic CRISP-DM Reference model……37

3.3: Decision Tree ……41

3.4: Deep learning Artificial Neural Network (ANN)……43

3.5: Naïve Bayes ……44

3.6: Random Forest ……45

Chapter 4

4.1: Analysis and Results……48

4.2: Understanding the Data …… 48

4.3: Data Preparation ……50

4.4: Rescaling and Normalization ……50

4.5: Model Performance ……51

4.6: Parameters for Measuring Model Performance ……52

4.7: Comparison of the Model ……57

Chapter 5

5.1: Conclusion ……59

5.2: future Work ……59

5.3: Limitations ……60

References ………..61

List of Figures Pages

Figure 1: ……………………... Structure of the research ………………………………………………26

Figure 2: ……………………… CRISP-DM process reference model for data mining…………….. 38

Figure 3: ……………………... Original dataset ……………………………………………………… 40

Figure 4: ……………………….. Diagram of a decision tree ………………………………………….. 42

Figure 5: ………………………Basic unit in deep learning (perceptron) …………………………….. 44

Figure 6: ……………………….. ANN ……………………………………………………………………….. 44

Figure 7: …………………….. Overview of the CRISP-DM task and their outputs……………………. 47

Figure 8: ……………………………… Density of the features …………………………………………….. 49

Figure 9: …………………………… Heatmap of the features ……………………………………………… 49

Figure 10: ……………………………………… Data type ……………………………………………………….. 50

Figure 11: ………………………………. Rescaled data …………………………………………………… 51

Figure 12: ………………………………… Normalized data ………………………………………………….. 51

Figure 13: …………………… Decision tree confusion matrix …………………………………………. 53

Figure 14: ………………… Roc curve of Decision Tree …………………………………………………..53

Figure 15: …………………… Confusion metrics of deep learning ANN ………………………………... 54

Figure 16: ………………… Deep learning Roc curve …………………………………………………………54

Figure 17: …………………… Confusion metrics of Random forest ……………………………………….. 55

Figure 18: ……………………… Roc cure of Random forest……………………………………… ……………. 55

Figure 19: ……………………Confusion metrics of the Naïve Bayes classifier ………………………….. 56

Figure 20: ………………… Roc curve of Naïve Bayes ………………………………………………………….. 56

Figure 18: ………………….. ROC comparison for the models ……………………………………………… 58

List of Tables Pages

Table 1: …………………… The dataset for the machine learning models………………….……….. 48

Table 2: ……………………... Decision Tree classifier performance ……………………………………... 53

Table 3: ……………………… Deep learning ANN performance …………………………………………….. 54

Table 4: ………………………... Random forest classifier performance …………………………………….. 55

Table 5: …………………………... Naive Bayes performance ……………………………….…………... 56

Table 6: ………………………Comparison of the five Algorithms performance……………………….. 57

List of Algorithms

- Decision Tree Algorithm
- Deep Learning ANN
- Random Forest Classifier
- Naïve Bayes

Chapter 1

1.1 Introduction:

Customer satisfaction is a very vital parameter of business. Retailers always try to fulfill their demands as often and possible while continually looking to earn more profit with a calculated investment to increase sales. Therefore, product sales forecasting helps retailers to increase profit with absolute investment.

Forecasting or prediction refers to the process of trying to foretell the likelihood of future events from past and present data available by analysis of trends and patterns found within the data [1]. In the same light, sales prediction is the process of estimating future revenue for a firm or outlet by predicting the amount of revenue to be generated over the given period in review [2]. This period may differ from weeks to months or even years. For any business to be productive, it is expected that they make judicious use of resources available to them to make sound strategic decisions for the future to remain competitive and increase their revenue base [3]. Since it is known that all forecasts involve at least some level of uncertainty, businesses need to make estimates with the aim of minimizing this uncertainty. Therefore Organizations should not make predictions considering a few factors, but their predictions should comprise various variables such as raw-material requirements, optimal stock levels, borrowing conditions, and personnel conditions [3]. Furthermore, in order for any of these to be projected, it is crucial to predict the expected demand that will apply to the market and, appropriately, the company's potential sales [3]. Thus, Accurate forecasts allow for appropriate targeting of the company's activities (such as production, finance, R&D, purchasing, and marketing) and assist them in reaching their targets (Mentzer, et al. 1998).

1.2 Research background

A significant number of business problems require decisions to be made based on a large amount of historical data and human decision-makers have difficulty integrating these data; thus, their decisions are susceptible to several biases [4]. Machine learning algorithms and data mining have been employed in recent times to help organizations and businesses combat and tackle the problem of harvesting data and using the knowledge found within them to make predictions for sales. With a good data mining infrastructure in place and the right application of predictive algorithms, companies stand a better chance of making better-informed decisions. Therefore, Sales forecasting is deemed of paramount importance for companies with intentions of entering new markets or adding new services, products, or experiencing high growth. When guided and assisted with the right tools to perform the task of sales forecasting, organizations tend to exert control and portray an outstanding performance in the market, thereby ensuring minimal loss of revenue and increased profit [4]. It is assuring to note that the main reason a company does a forecast is to balance marketing resources and sales against supply capacity planning [4]. With the right implementation of forecasting, companies and corporations no doubt can address fundamental questions such as "can we drive demand with our current fusion of price, promotion, and marketing?", "are the resources currently at our deposal adequate to measure up to demand?", "Is there enough personal to match the volume of budgeted sales?"

[1]. In this regard, companies and organizations allocate and invest a reasonable amount of resources in both human and finance to obtain adequate and genuine prediction results.

1.3 Research aim and objectives

This study's principal objective will be to implement data mining techniques and machine learning algorithms to help identify the class of products that sell more for a business outlet to aid the business owners in making more informed decisions on inventory acquisition. This would be carried out by performing a comparative analysis of four different classification machine learning algorithms in line with the CRIPS_DM methodology to see which algorithms best perform on the available data. Data mining techniques and implementation shall be discussed in the latter part of the research. This paper will consider a variety of forecasting algorithms such as the Decision tree, deep learning, Naïve Bayes and Random Forest.

1.4 Research Questions

- What are the drawbacks of product sales prediction in past research works carried out in retail sales?
- What are the common and most essential factors or variables that influence retail sales prediction?
- What algorithms are best suited to handle product sales forecasting problems?

- How do data mining and machine learning algorithms help to make product sales or demand forecasts better than the conventional statistical approach?

1.5 Why do forecasting

Sales forecasting adds excellent value to the business because it deals with the future. For businesses in the private sector, the confidence level concerning their future investments is made strong when the predictions are accurate and can be trusted [5]. Similarly, for publicly traded companies, an accurate forecast will bestow strong credibility in the market and adds value to the business [5]. Also, firms that are into production employ forecasting to moderate and plan their production circle to ensure they meet up with demand, curb excess and waste [6].

In a scenario where there is inaccurate forecasting or disconnect within the firm or business, and forecasts are not communicated, the outcome is always adverse for the business as this will result in discrepancies and gaps within the organizations' goals and targets. It is worthy to note that when an accurate sales forecast is available for a business, it confers the following benefits.

- It guides the alignment of sales quotas and proceeds expectations.
- Helps to improve decision making about the future.
- It helps in reducing the time spent on planning territory coverage and establishing quotas.
- It helps to set benchmarks that will be used to access trends in the future.

In today's business world, the importance of forecasting and making future predictions for growth and profit maximization cannot be overstated as intelligent business decisions require both decision analysis and prediction [5]. Therefore, the key to making outstanding business decisions lies in business executives' ability to effectively make the right prediction on customer demands and future sales. Hence, a perfect forecast has a tremendous impact on earning a profit, storage, sales amount, and customer attraction [7,8]. Data mining comes in as helpful tools for analysts to extracting information, and hidden patterns from enormous datasets generated over time to enhance the accuracy and efficiency of predictions.

With the ever-constant evolution of technology, drastically affecting all aspects of human life and beyond. Business management has not escaped the trends of innovation and technological boosts that have occurred over the past decades resulting in a tremendous increment in business problems and relatively more enhanced solutions to tackle these organizational needs. In other to maintain competitiveness in the retail industry, which can be highly dynamic, business owners need and should be able to apply the right technological infrastructure to predict future sales and the precision of these forecasts is of immense importance [7]. For the retail market, the capability to sustain its inventory to ensure customer satisfaction guarantees a reduction in distribution expenses and stocking costs, thereby improving its profit [6]. Furthermore, the implementation of modern technological infrastructure to address business and organizational needs with regards to forecasting and investigating trends that concern the business's future places the management of the businesses in an advantageous position to make more informed decisions. Thus a business's administration can be a hugely challenging task that can go completely wrong if the executives in charge make erroneous decisions [8].

1.6 Methods of Forecasting

There are various methods in which an organization can perform and implement forecasting for sales. However, there are two basic forecasting methods: the bottom-up (BU) sales forecast or the top-down (TD) sales forecast approach. Should an organization decide to carry out the bottom-up sales forecast method, the organization starts by focusing on the projected amount of units it presumes it is likely to sell and multiply the presumed amount by the average cost pre-unit. This approach can also be cohesive into other metrics such as the number of sales reps available, the number of possible sales locations, and online interactions. The bottom-up sales concept aims to start with the smallest components of the forecast and build up from there. In the bottom-up sales forecast approach, dynamics in variables such as cost of an item or units and the number of reps and sales locations can be changed with relative ease. Thus, this methodology provides reasonably granular information.

In comparison, the top-down sales forecast methodology concentrates mainly on the market size (total addressable market-TAM). It estimates the possible percentage of the market that the business can capture by projecting the percentage of the total market share they plan on selling for the period in review. Furthermore, both these methods should be applied for sound and implementable sales forecasts by first starting with the bottom-up approach and then the top-down approach; this will ensure that the forecast is feasible.

However, sales forecasting in the retails industry can be a challenging task to undertake due to the highly volatile nature of the market with an ever-changing taste of customers, and sometimes the product life circle of goods is very short [9]. Besides, most retail businesses like fashion lines are significantly affected by seasonal factors such as weather and fashion trends factors alongside many other tricky variables such as marketing strategies, political climate, item features, and macroeconomic trends [9]. Therefore, making predictions for sales or demand in such a market will require the implementation of more sophisticated and versatile analytical tools and approaches. On the other hand, it is a well-known fact that the fashion industry's supply chain is indeed a long one comprising of cotton plants to fiber manufacturers, apparel factories, distributors, wholesalers, and retailers. In this regard, the infamous bullwhip effect [10] will have a dominating impact on the fashion supply chain. Because forecasting is a critical factor correlating with the presence and significance of the bullwhip effect, enhancing forecasting can help decrease the bullwhip effect, which directly enhances the supply chain's proficiency [9].

Within the past decade, quite an ample amount of research and studies have been carried out and detailed in literature as regards forecasting methods. Nonetheless, each forecasting technique has its curbs and downsides. For instance, when considering the traditional statistical approaches that depend hugely on the time series data's features, it is known to affect the forecast's accuracy significantly. In contrast, Artificial intelligence (AI) methodologies can perform better with regards to accuracy compared to the traditional statistical forecasting models. However, they typically require a much longer time to execute and require a higher degree of computational power [9]. Consequently, researchers have always recommended a combination of various techniques simultaneously to form a new "hybrid technique" to achieve a capable and yet efficient forecasting result.

1.7 Statistical Sales Forecasting Methods

In our dynamic and ever-changing world, statistical methods that tend to extrapolate past patterns into the future are considered and perceived to be myopic and slow to react to changes [11,12]. Furthermore, there may be little or no data from the past available for the current forecasting problem. Usually, past acquired data may come with noise, which implies that the data needs to be massaged to remove them before a statistical method can be applied to them [11]. This can turn out to be quite a laborious process when the amount of forecast that needs to be made is large [11]. Furthermore, Statistical techniques are less considered susceptible to biases and are known to use past data effectively. Thus, statistical prediction methods are dependable; when given the same data, they tend to deliver the same forecast results whether the series relates to costs or revenues and other related variables [12].

Conventionally, sales forecasting is achieved by implementing statistical techniques such as linear regression, moving average, weighted average, exponential smoothing (used when a trend is present but not linear), double exponential smoothing, and Bayesian analysis, to mention a few [12]. In Statistical time series analysis, tools such as ARIMA and SARIMA are widely used in sales forecasting. Given that these techniques have a closed-form expression for prediction, it is simple and can be implemented easily, and the results can be computed very quickly [9]. However, Despite being used popularly due to their simple nature and speed, statistical techniques are known to possess few challenges. First, choosing the right statistical techniques is never an easy undertaking as it entails an "expert" knowledge of the various techniques to be considered. Second, their performance does not usually promise outstanding results. Furthermore, statistical models tend to perform poorly compared to the more sophisticated and robust techniques like AI techniques. Third, sales are usually influenced by numerous factors such as trends and seasonality and exhibit a highly irregular pattern, suggesting that pure statistical methods may fail to achieve a desirable forecasting outcome [9].

1.8 AI Sales Forecasting Methods

As previously discussed, statistical models or techniques possesses some degree of deficiency in performing sales (or demand) forecasting. Hence to enhance forecasting accuracy, AI methods have been employed in recent times with the emergence of advanced computer technology. Nevertheless, AI models can efficiently derive "arbitrarily nonlinear" approximation functions directly from the data [9]. Prevalent models like the artificial neural network (ANN) models and fuzzy logic models are commonly used in literature as they are usually and quite often the first kind of models being employed for sales forecasting [9]. According to the journal paper written by Vahid et al. [13], they paid extensive attention to forecasting regarding export sales forecasting (ESF), where they pointed out that only but a few studies have been carried out on export sales forecasting in comparison to domestic sales prediction. Pointing out that export sales are an essential aspect for quite a vast number of firms, which is characterized by several unique uncertainties and challenges which generally would be very difficult for statistical models to handle. Vahid pointed out that Sales forecasting (especially in the form of exports) plays a critical role in linking internal decision-making and unmanageable external factors that can affect an organization. They stated that the distinguished nature of ESF could be traced to three principal factors. First, it is complex due to the involvement of country/regional executives, export agents and other third parties who all contribute to influencing the decisions. Secondly, the required data for the forecasting process can be challenging to obtain due to impediments associated with data gathering, which include availability issues, accessibility and permissions, and quality of the data. Lastly, various countries or regions may require unique forecasting models or different techniques for their forecasting model development, as in some markets, data availability issues and local expertise might restrict forecasting approaches. These factors contribute to making export sales a distinctive and challenging type of forecasting that cannot rely only on the traditional time series forecasting techniques to be analytically accurate and reliable [13].

1.9 Hybrid Method of Sales Forecasting

Hybrid forecasting methods typically are built on the premise that they can incorporate the strengths of different models to form a new forecasting system. As such, many of these models are believed to be more effective than mere statistical models and pure AI models.[9] Hybrid methods used in sales forecasting often combine different schemes, such as the fuzzy model, ELM and ANN with other methods, such as statistical models, the Gray Model (GM), to name a few. The Fuzzy Logic Based Hybrid method was initiated by Vroman et al. This model is drawn from a fuzzy-adaptive model that regulates the weighting factors of the exponential-smoothing statistical "Holt-Winter" forecasting process. They prove that the proposed fuzzy hybrid model is superior to the traditional Holt-Winter model. Subsequently, Thomassey et al. used a fuzzy logic concept in making predictions. Their new model enables automated learning of the effect of nonlinear explanatory variables. Their proposed forecasting method is based on multiple models such as fuzzy logic, neural networks, and evolutionary processes. They claim that the outcome is flexible in the processing of unknown data. Recently, Yesil et al. used a hybrid fuzzy model for fast fashion forecasting in their work. Specifically, they combined the fuzzy logic model and the statistical model for forecasting. In their hybrid method, they calculate the final forecast for weekly demand on the basis of the weighted average of the forecasts produced by multiple methods. They claim that their proposed technique accomplishes high precision.

In the hybrid neural network (NN) models, equally proposed by Vroman et al., a NN model with remedial seasonality coefficients is used for mean-term forecasting. They contend that their hybrid method can also be used for short and discontinuous time series forecasting. They report great results with their projected NN hybrid model and believe that the remarkable performance stems from the NN's ability to map the nonlinear relationship between data inputs and outputs. In addition, Thomassey and Happiette designed a hybrid neural clustering and classification scheme for the sales forecasting of new products. In comparison to the typical sales profile predictor, their model can improve the accuracy of mid-term forecasting. ANN can similarly be merged with other techniques such as Greymethod (GM) and Autoregressive Technique. Forthgoing Ni and Fan et al., in their proposed model, used a neural network to develop a multivariate error forecasting model. Their model introduces the concept of "influence factors" and divides the "impact factors into two distinct phases (long-term and short-term). Their experiment shows that a multivariate error forecasting model can generate strong predictive results for fashion retail sales forecasting problems. Aksoy et al., in their work, combines the fuzzy method and a neural network to form a new system called an adaptive-network-based fuzzy inference system. The proposed new system combines the advantages of both systems, namely the neural networks' learning capability and the generalization ability of the fuzzy logic technique and establishes a robust hybrid system.

Hybrid-based ELM methods. The Extreme Learning Machine (ELM) is easy to make predictions. While not ideal because of its unstable nature, its "fast speed makes it an excellent contender to be a component model for a more sophisticated hybrid model for forecasting [9]. Wong and Guo et al. suggest a new learning algorithm-based neural network to produce initial sales forecasts and then use a heuristic fine-tuning method to achieve a more reliable final sales forecast. Their learning algorithm combines an enhanced Harmony Search algorithm and an intense learning machine to boost network generalization efficiency. Also, Xia et al. looked at a forecasting model based on an extreme learning machine model with adaptive metrics. In their model, inputs can solve the problems of amplitude shift and trend determination, which in turn helps to reduce the impact of network over-installation. In addition to the types of hybrid methods discussed above, several other novel mixed forecasting methods are also mentioned in the sales forecasting literature. For example, Choi et al. employ a hybrid SARIMA wavelet transform (SW) method for forecasting fashion sales. Using real data and artificial data, they demonstrate that with relatively poor seasonality and a highly variable seasonality factor, their proposed SW method outperforms conventional statistical methods.

1.10 Ensuring an Accurate Forecast

To ensure an accurate forecast is to be made from a business perspective, the business or organization executives should be able to incorporate these five necessary steps into their forecast parameters.

Accessibility to historical data: The availability of past data and trends should be readily available for examination. The basis of a good forecast starts by first breaking down the data by sales, sales period, quantity sold, amongst other relevant variables to be considered and then built into a "sales run rate," considered as the number of sales projected for the review period. To this effect, data mining tools are employed to guide and assist in the collection of relevant data that is to be used for the forecast. Also, the implementation of the CRISP-DM methodology is crucial at this stage, considering that it will help the analyst acquire data that corresponds to the business's demand and goals, thereby ensuring that the data collected for the forecast is relevant and reliable [14].

Include expected changes: This involves modifying the sales run rate by implementing the expected changes that may occur. This may include changes in prices due to competitions or other forces, incremental or reductional changes in customer base, promotions, and change in channels either additional or reductional as well as product change [9].

Anticipate market trends: This involves the projection of market events that are likely to affect feature sales, such as government policies and competition moves. In addition, some trends like seasonality, weather and region also affect the outcome of a forecast and should be considered when performing a forecast. This will ensure that the best features are used to carry out the forecast.

Monitor the competition: Keeping an eye on the market to ensure that their incentives are rivaled, and the tactics of emerging competitors are learned.

Incorporate the organization's business plan: Ensure that all business plans and projections are carried along with the forecast to maintain the organization's goal. The CRISP-DM methodology comes into play once again at this stage because no business or organization can proceed to carry out sales forecasts without first considering the business or organizational goals. Implementing the CRISP-DM methodology ensures that the forecast is guided by a well thought out plan to improve business proficiency and overall customer satisfaction [14].

1.11 A Successful Sales Forecast

The feasibility and precision of sales prediction relies on many aspects, including automation, accurate evidence, excellent analytical methods, and good operational teamwork. Superlatively, a good forecast should be data-driven, and predictive analytics tends to be more forward-looking than most subjective analytical approaches, and the ability for a forecast to be produced in real-time grants the leadership of an organization the ability to course-correct and make more informed decisions. Furthermore, a good forecast serves as a single source with multiple views giving greater insight into an organization's performance and aligning various business functions within the organization. Similarly, a good forecast serves as a base to provide more insights for more enhanced further predictions where the accuracy will gradually improve. Over time, organizations with good quality forecasting processes and mechanisms tend to perform better than their competitors because they have a clearer understanding of the business and market drivers and subsequently will have the capacity to manipulate the outcome of the sales period in review.

1.12 Research Structure

The following roadmap is used to establish a conceptual framework to execute this dissertation project.

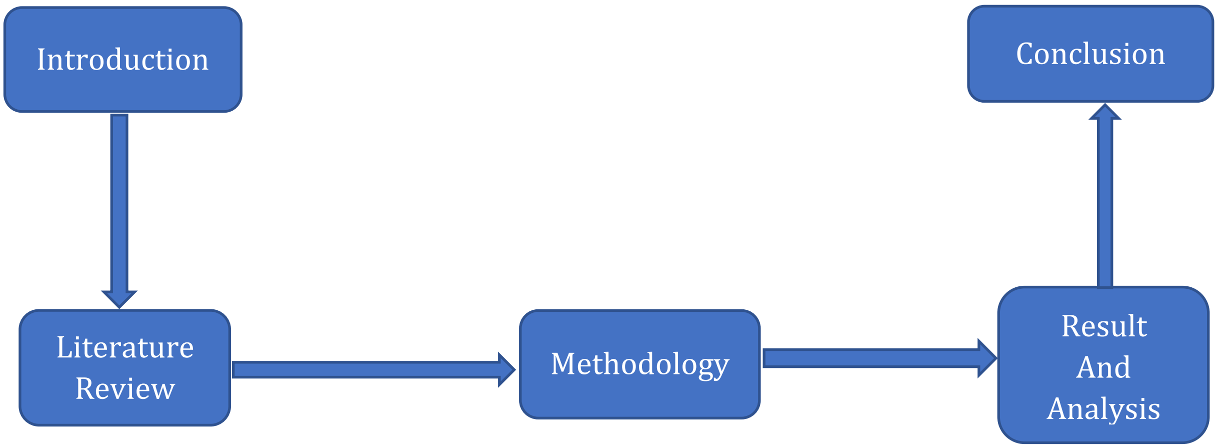


Figure 1: Structure of the research

Chapter 2

literature Review

The search for better models and techniques for data mining by companies and organizations is a continuous one as it aids them to remain competitive and generate higher revenue [5]. There have been lots of work in the area of product sales forecasting using machine learning and data mining. This section of the work goes through a brief survey of several related studies that have been carried out in sales forecasting and related forecasting problems. Many statistical models and techniques such as regression, (ARIMA)Auto-Regressive Integrated Moving Average, (ARMA) Auto-Regressive Moving Average, have been applied for creating multiple sales forecasting concepts. However, forecasting is a complex problem and is influenced by external as well as internal factors, and there are drawbacks to the statistical approach as told by Weigend et al. 1994. In their work, Krishna et al. 2018 [1] stated that "forecasting can be used to predict sales revenue at the product level, or at an individual business level, or at a company level. However, their work concentrated on product level sales forecasting. They explained that Future sales plan aids in optimal utilization of facility, scheduling, conveyance, and effective control of inventory. These, in turn, result in enhancement of clients' satisfaction and decrease in production cost. They explained that sales forecasts affect a company's marketing plan directly and that the marketing departments of businesses and companies are accountable for how their customers understand its services or products and compare it against its competitors by using their sales forecast to assess how marketing expenditure can increase sales and channel in more demand. Thus, it is crucial to develop effective sales forecasting models in order to make accurate and robust forecasting outcomes. However, they pointed out that in the business and economic environment, it is imperative to correctly predict various kinds of economic variables like the past economic performance, present global conditions, inflation rate, internal organizational changes, marketing strategies, seasonal factors, etc. so as to develop appropriate approaches. They considered a range of forecasting techniques such as Multiple regression, Polynomial regression, Ridge regression, and Lasso regression alongside a variety of boosting algorithms like AdaBoost, Gradient tree boosting to get the maximum accuracy.

Arif A I et al. 2019 [15] presents a new method-using machine learning to assist with accurate predictions in a comparative study. Their proposed method collects and analyzes the store's previous data by implementing various algorithms like the K-Nearest Neighbor (KNN), Gaussian Naïve Bayes, and Decision Tree Classifier to find out the best technique for forecasting demand. Their analysis of these algorithms found out that Gaussian Nave Bayes has the best accuracy for their historical data. In work done by S. Cheriyan et al. 2018. [5], they implemented various machine learning algorithms and data mining techniques with the perception of choosing the best approach to forecast sales with a high degree of precision. However, they pointed out that traditional forecast systems are complicated when dealing with big data and the accuracy of sales forecasting. Nonetheless, they explained that these issues could be surmounted by utilizing various data mining techniques.

Likewise, Chand N et al. 2017 [16] in their paper, suggest a new set-up methodology using the averaging approach, which not only gives priority to an algorithm that reliably retains good precision but also reduces the variance from actual sales. Their results found that the assembling of the time-series model results and the regression model gives a better result due to the nullification of over-forecasting and under-forecasting and put the forecast values close to the real in most situations and gives weightage to the model that has performed better previously. Also in the research paper written by Lin K. Y et al. 2016 [8], they developed a deep learning-based customer forecasting tool to improve the management quality for decision making. Their proposed approach extracted the relationship between the number of customers and the leading factors of the vast amount of historical data. A Deep Learning algorithm was used in executing sales analysis and produce steadfast predictions that would become administrative tools, especially with the building of a forecasting model based on customer behavior with weather as the prediction parameter. Their study built a deep learning-based customer forecasting tool to analyze Point of Sale (POS) data. Their proposed model allows business units to make the appropriate decisions that represent an improvement in the store's performance and development that enhances the customers' experience. Considering the lousy forecasting results of the standard support vector machine (SVM) for product sale series with the normal distribution noise. Wu Q et al. 2008 [17] proposed an SVM based on the Gaussian loss function. Then, a hybrid forecasting model for product sales and its parameter-choosing algorithm is presented in their study. Wu, however, pointed out that the standard SVM encounters several complexities in real-life application but new and improved SVMs have been developed to help solve the actual problems of SVM. Nonetheless, they further stated that the standard SVM adopting ε -insensitive loss function has good generalization capability in some applications but challenging to deal with the standard distribution noise parts of (time) series. Their work mainly focuses on establishing a new SVM that can deal with the Gaussian noise parts of series.

In their study, Jagielska I et al. 1993 [4] investigates the connectionist concept to sales forecasting. They built a neural network model for the prediction of sales levels, and the sales forecast produced was compared with the actual sales figures as well as with the estimates made by the marketing managers. The model they built was presented with the past sales data then the system trained on the data and learned the pattern from which it made its predictions. In their study, they showed that neural networks could be applied to effectiveness in solving problems that are complex and non-deterministic. The neural network they built was meant to predict sales figures for Tattersall, which is Australia's leading gaming operator, to help their marketing managers make sales forecast for lottery products. Furthermore, the work carried out by Kilimci Z H et al. 2019 [18] proposed the infusion of nine different time series methods, Support vector regression algorithm (SVR), and Deep Learning forecasting models to improve the demand forecasting process, which is one of the major issues concerning supply chain. To achieve a final decision of these models for their proposed system, they blended the algorithms with a novel integration strategy that is reminiscent of boosting ensemble strategy. Also, they introduced another novel approach in their work by adopting a boosting strategy to their demand forecasting model. In this way, their proposed system's final decision was based on the best algorithms with regards to gaining more weight. To this effect, their forecast was more reliable with respect to the changes in trends and seasonality behaviors. A sentimental analysis approach was used by Avinash K.S et al. 2020 [19] in their work to forecast the demand for a product for a particular period. Their study explained in detail how random forest and clustering algorithms with neural networks could be used to implement product demand forecasting to improve organizations or business profits. Regarding the work done by Mortensen S et al. 2019 [20], they develop a model that can forecast sales performance with a fair degree of accuracy. They experimented with a range of machine learning techniques, including binomial logit and various decision tree methods, including gradient and random forest boosting. In another study carried out by Saena W et al. [21], they applied machine learning methods for predicting the number of pharmaceutical sales. The efficacy of each model was compared using their errors. CNN-LSTM model produces better forecasting results for their study.

With inventory planning being one of the fundamentals of a retail fashion operation, an adequate forecasting system can undoubtedly go a long way to help businesses in fashion retail avoid understocking or overstocking in inventory planning. This further relates to other essential operations of the whole supply chain like production planning, pricing and achieving customer satisfaction. Na Liu et al. 2013 [9] presents a comprehensive review of the literature on fashion retail sales forecasting and equally explores the benefits and the downsides of various analytical techniques for fashion retail sales forecasting as well as examining the pertinent issues relating to real-world implementation of fashion retail sales forecasting algorithms. In another study, Dokuz Eylul universities et al. 2017 [3] investigates the forecasting performances of models such as state-space models and the ANFIS model, making them be the first to contribute to the literature on the study. Their study analyses forecasts of the sales volume of a retail furniture store in order to appraise the forecasting performance of five methods: namely, state-space models, the ARIMA model, the ARFIMA model, the ANN model, and the ANFIS model. Based on the accuracy observed for each of the forecasting techniques, they suggested that employees in various relevant departments can derive a general understanding of which forecasting approach conform best with which product sales and thereby identify opportunities to ameliorate the existing supply chain system. They aimed to determine which of the most employed blended techniques produce greater statistical enhancements in the relevant series' forecast accuracy. Their study concluded that it is needless to identify the best for individual models when seeking to improve the forecast accuracy achieved by combined forecasts. Furthermore, they claimed that a combination of very similar models affects the performance of forecasting positively. Arif M A I et al. 2020 [7] took a new machine learning approach that helps produce accurate predictions. Their work focuses on finding the best algorithm to estimate product demand with high precision and less error. They analyzed and used three popular algorithms, namely K-Nearest Neighbor, (KNN) Gaussian Nave Bayes and Decision Tree Classifier for forecasting and classification. Goodwin P et al. 2002 [11] review research on effective approaches designed to allow judgment and statistical methods to be integrated when short-term point forecasts are required. They pointed out that when forecasts are to be made, that two approaches to the integration of judgment and statistical methods are usually identified. Their paper identified that with regards to voluntary integration, the judgmental forecaster is provided with details of the statistical forecast and then decides how to use this in formulating their judgment. Thus the forecaster is free to ignore or accept the statistical forecast when making the judgmental forecast. However, voluntary integration entails applying judgmental adjustments to statistical forecasts. Nonetheless, it might also involve the forecaster to modify "prior" judgmental forecasts should there be a newly arrived statistical forecast. Whereas in mechanical integration, the integrated forecast is achieved by applying a statistical method to the judgmental forecast.

Gallagher C et al.'s 2015 [22] research paper considers Bayesian classifiers to evaluate whether they deliver higher performance for the task of sales forecasting for detecting deals that are at risk of being lost or deals with the possibility of being won, which may have been classified wrongly by Qualitative Sales Predictor (QSP). By combining both quantitative and qualitative sales attributes to improve performance further, they used a more sophisticated algorithm than those used in QSP. Their research claimed that Bayesian classifiers were the most useful and suitable learning algorithms for the application and data. More precisely, the TAN classifier had the maximum predictive accuracy in both testing and validation. The final result produced 90.6% accuracy, implying that the TAN algorithm drastically improved upon the QSP, which had an accuracy of 75.6% on the same test data. Although data processing methods such as machine learning, having been developed in the field of information science and are currently being employed in practical fields such as medical data mining. These data mining techniques, alongside big data, are an area of great interest. Researchers to date continue to discover knowledge and opinions that are ever beneficial to businesses and organizations from the large amounts of customer information data gathered by these businesses using these techniques. Among these, the deep learning approach is a relatively new method of machine learning; though commonly and widely used in image recognition, there have been little or no reports of its use in or instances of its induction to marketing in retail businesses [23]. Kaneko Y et al. 2016 [23] proposes a sales prediction model for retail stores as a way to introduce the deep learning approach into the field of marketing so that store managers can use it for marketing strategies. With demand forecasting being one of the main decision-making responsibilities of a business, business intelligence (BI) becomes inevitable for analytics and strong decision making for all sorts of businesses across all sectors. Thus, by analyzing past and current market data, businesses and organizations can effectively predict the future demand of goods and manufacture those goods that appear to have more demand in the future. In their work, Khan M.A et al. 2013[24] built a demand forecasting model structured on business intelligence and machine learning. Their model yielded a 92.38% accuracy when implemented in stores.

In 2015, Rajab S et al. [25] performed a comparison of prediction performance of two commonly used AI methods, namely Adaptive Neuro-fuzzy inference system (ANFIS) and Artificial neural network (ANN) for sales forecasting and stock market price prediction. The results they obtained indicate that both ANFIS and ANN are indeed effective tools that can be used for forecasting problems. Nevertheless, they pointed out that the learning period of ANFIS is briefer, suggesting that ANFIS attains the target earlier than ANN. Additionally, the accuracy they got for the ANFIS algorithm for the stock market dataset was greater than ANN, which goes to show that ANFIS is a superior forecasting model. Whereas for sales data, the ANN algorithm performed better. As noted by Islek I et al. 2015 [26] in their research paper, demand forecasting is building forecasting models to estimate the quantity of a product that customers will purchase in the future. Their work addresses the challenge of forecasting several product demands of distribution warehouses. They presented a suitable methodology for demand forecasting to overcome the limitations concerning product demand and distribution in warehouses while providing a high estimation accuracy. Their proposed methodology clusters similar warehouses according to their sales behavior and pattern by employing bipartite graph clustering. They also introduced a hybrid forecasting phase that combines the moving average model and Bayesian Network machine learning algorithms. Another intressing research was carried out by Chen T et al. 2018 [27], where they implemented trend alignment with dual-attention multi-task recurrent neural networks for sales prediction. They propose an innovative dual-attention based LSTM decoder to address the issue by labeling the prominent factors in sales time series into internal features and external features and then model them into two aspects with a multi-task based LSTM encoder. Their methodology allows the encoder to model various variables with different semantic meanings expansively. With regards to multi-source learning, Tsai K et al. 2020 [28] proposes a novel multi-source learning framework for sales prediction, allowing e-retailers to preserve their products smartly and adjust the price appropriately to maximize profit.

Chapter 3

3.1 Methodology

Various methods of data mining and machine learning have been implemented over the years to assist researchers and analysts in carrying out various prediction problems that affect our daily lives. Product sales prediction comes in as one of the tools that should be readily available for a business or organization to help them make more informed growth and expansion decisions. This paper will concentrate on implementing the CRPIS-DM as a guide to data mining to make a sound forecast that will correspond to business goals and objectives. This approach offers a framework that leads to more substantial and quicker outcomes. The CRISP-DM approach organizes research into six stages, explaining the process better and including a path map to be taken while preparing and carrying out the research.

3.2 CRPIS-DM Methodology

Data mining is an innovative method that involves a range of different skills and expertise. There is currently no standard structure for carrying out data mining projects, implying that a data mining project's success or failure is heavily contingent on the individual or team conducting it [29]. Data mining requires a standard methodology that can help to convert business or organizational problems into data mining tasks, propose suitable data transformations and data mining techniques, and provide means for assessing the feasibility of findings and recording the knowledge. The Cross Industry Standard Process for Data Mining (CRISP-DM) resolved some of these concerns by defining a process model that provides a context for the execution of data mining projects that are independent of both the industry and the technologies used. Analysts undertaking data mining tasks consider CRISP-DM methodology useful in several respects. For novices, the methodology provides guidance, helps to plan the project, and provides guidance on each task or stage of the process[29]. Also, seasoned analysts take advantage of checklists for each role and ensure that nothing substantial has been omitted. However, the most crucial function of the CRISP-DM methodology is to communicate and log outcomes. It helps to put together various tools and people of varied expertise and experiences to form an integrated and productive project.

3.2.1 The Generic CRISP-DM Reference Model

The CRISP-DM data mining reference model offers an outline of a data mining project's life cycle. It includes the stages, their respective tasks and results of a project. A data mining project's life cycle is split into six phases as seen in Figure 1. The order of phases is not very rigid. The arrows display only the most critical and regular dependency between the stages and each step's outcome, or which basic phase activity is to be done next. The outer circle in Figure 1 signifies the cyclical existence of data mining. Data mining is not over, even as the solution has been implemented. During the project and the approach deployed, the lessons learned will cause new, often more oriented business concerns, and subsequent data mining processes can benefit from previous experience. The CRISP-DM model is shown in the figure below.

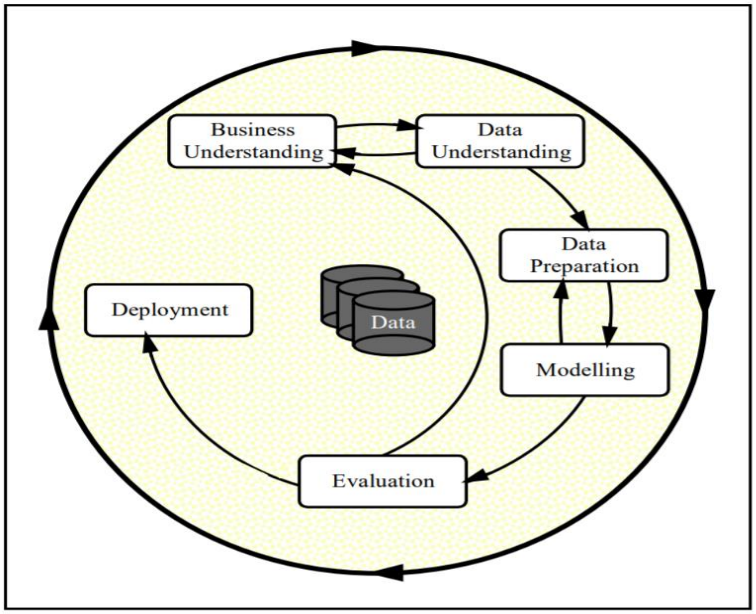


Figure 2: CRISP-DM Process reference Model for Data Mining

The various stages of the CRISP-DM model have been employed in the cause of this project, and the stages will be explained with respect to sales forecasting.

- Business Understanding

  This initial step emphasizes a company-led understanding of project priorities and aspirations and then translating this information into a problem description for data mining and a preliminary developed project plan. In this regard, this project's business problem is to perform a sales forecast for an off-license retail store to ascertain which of the products sells more by investigating their past sales data. This is necessary to enable the business managers to make more informed decisions on overstocking or understocking to ensure maximum profit and increase customer satisfaction. The purpose of this analysis is to compare the best machine learning algorithms to find the most effective prediction model for the two key product classes offered by the shop and the factors which influence them.

- ❖ Data Understanding

The process of data comprehension begins with the initial data collection and proceeds to become more familiar with the data to find out about issues of data consistency, uncover initial insight into the data, or detect interesting subsets in order to create conclusions about secret knowledge. This project made use of data that was collected from the sales of an off-license retail store. The dataset provides information of over 50 kinds of products sold at the shop for the month of January, classified into wine/champaign or spirits/liquor. The data collected consists of 16892 rows and 16 columns. The following are the columns of the dataset.

- Year: showing the year of sales in review.
- Month: showing the month in review.
- Supplier: showing the supplier of the product to the store.
- Item code: showing the code for each item.
- Item Description: Describing which type of drink the item is.
- Item Type: showing which class the item belongs to wine/champaign or spirits/liquor.
- Damaged units: showing how many of the units got damaged.
- Retail transfers: showing how many items transferred from the store.
- Item weight: showing the weight of each item.
- Item visibility: position on store shelf.
- Quantity: showing the quantity sold.

- Discount: showing a discount percentage for each product.
- Returned: showing amount returned.
- Unit price: showing the price for each item.
- Payment method: indicating if the payment is made in cash or via online platforms.
- Total sales: total amount sold for the product.

❖ Data Preparation

The data preparation process includes all activities to create the final data (data fed from the original raw data in the modeling tool(s)). Data processing procedures can be completed several times and not in any specified order. These tasks include choosing tables, document and attribute, data cleaning, new attribute creation, and data conversion for modeling tooling. It basically entails the proper analysis of the data obtained to create the final dataset for the forecast. After initial data cleaning and preparation, the year, month, supplier and item code columns were dropped, and the item type converted to binomial. The figure below shows the dataset before and after initial cleaning and selection of the features.

1	YEAR	MONTH	SUPPLIER	ITEM CODE	ITEM DESC	ITEM TYPE	Damaged	RETAIL TRA	Item_Weig	Item_Visibi	quantity	discount	returned	Sales price	PaymentM	Total_Sales
2	2018	1	REPUBLIC N	100009	BOOTLEG F	3	0	0	14.5	0.08	99	1.96	11.1	98	0	9700.04
3	2018	1	INTERBALT	100012	PAPI P/GRI	3	0	0	17.35	0.05	70	1.84	10.5	92	0	6438.16
4	2018	1	ROYAL WIN	100080	KEDEM CRI	3	0	0	5.63	0.01	83	1.02	12	51	0	4231.98
5	2018	1	RELIABLE C	1001	SAM SMITH	2	0	0	18.6	0.1	67	1.42	9.7	71	0	4755.58
6	2018	1	ROYAL WIN	100200	GAMLA CA	3	0.08	0	13	0.04	58	0.5	10.2	25	0	1449.5
7	2018	1	DIONYSOS	100285	NAOUSSA I	3	0	0	6.405	0.05	75	1.74	9.5	87	1	6523.26
8	2018	1	DIONYSOS	100293	SANTORINI	3	0.92	0	20.2	0.09	97	1.84	10.4	92	0	8922.16
9	2018	1	KYSELA PEF	100641	CORTENOV	3	0.41	0	18.85	0.04	53	1.44	9.3	72	1	3814.56
10	2018	1	SANTA MA	100749	SANTA MA	3	0.33	0	7.21	0.1	69	1.96	10.5	98	0	6760.04
11	2018	1	REPUBLIC N	100803	CYT XPLOR	3	1.14	0	7.96	0.02	18	1.86	10	93	1	1672.14
12	2018	1	REPUBLIC N	100927	CRISTALINC	3	0.58	0	20.75	0.04	49	1.84	10.2	92	0	4506.16
13	2018	1	CLIPPER CI	10095	HEAVY SEA	2	9	58	17.5	0.09	84	0.56	10.2	28	1	2351.44
14	2018	1	CLIPPER CI	10096	HEAVY SEA	2	0	0	13.65	0.01	16	1.92	9.1	96	0	1534.08
15	2018	1	JIM BEAM I	10103	KNOB CREE	0	3.04	10	10.1	0.1	63	0.92	10.2	46	0	2897.08
16	2018	1	AMERICAN	101079	KSARA CUV	3	0	0	15.1	0.08	88	1.04	10.7	52	0	4574.96
17	2018	1	AMERICAN	101117	KSARA CAB	3	0.73	0	16.85	0.03	92	0.86	9.7	43	1	3955.14
18	2018	1	HEAVEN HI	10120	J W DANT E	0	2.2	0	19.5	0.04	76	1.76	9.3	88	1	6686.24
19	2018	1	BACCHUS I	10123	NOTEWOR	0	2.88	2	14.3	0.06	13	1.2	11	60	1	778.8
20	2018	1	BACCHUS I	10124	NOTEWOR	0	2.89	0	20.75	0.1	59	0.34	10.8	17	0	1002.66
21	2018	1	BACCHUS I	10125	NOTEWOR	0	2.21	1	9.695	0.02	36	0.9	10.7	45	0	1619.1
22	2018	1	MONSIEUR	101346	ALSACE WI	3	0	0	10.5	0.09	41	1.72	10.7	86	0	3524.28
23	2018	1	A VINTNER	101532	HATSUMA	3	1.36	0	17.6	0.08	49	1.16	8.8	58	1	2840.84
24	2018	1	A VINTNER	101567	HATSUMA	3	1.17	1	10.195	0.1	30	0.4	10.9	20	0	599.6
25	2018	1	WILLIAM H	101621	CH CALLAC	3	0.08	0	16.7	0.06	6	0.94	10.5	47	0	281.06
26	2018	1	ROYAL WIN	101664	RAMON CC	3	0.58	2	4.615	0.09	7	1.56	10	78	1	544.44
27	2018	1	ROYAL WIN	101672	KEDEM SHI	3	0	0	8.325	0.05	58	1.84	10.4	92	1	5334.16
28	2018	1	REPUBLIC N	101680	MANISCHE	3	0.34	0	8.395	0.07	64	1.98	9.2	99	1	6334.02
29	2018	1	ROYAL WIN	101753	BARKAN CL	3	0	0	16.2	0.01	79	0.58	12.6	29	0	2290.42
30	2018	1	LEGENDS L	10195	PEAK ORGA	2	0	0	6.46	0.07	74	0.94	9	47	1	3477.06

Figure 3: original dataset

- Modeling

  Different modeling techniques are chosen and implemented in this stage, and their parameters are tuned to optimal values. Typically, for the same data mining problem sort, there can be many techniques to tackle the problem, and special formats for such techniques are required. Data planning and modeling are closely intertwined and many times, when modeling one realizes issues with data or suggestions for creating new data. This project considers five different types of techniques for the prediction.

3.3 Decision tree

The decision tree algorithm is one of the family of algorithms under supervised learning that can be used in solving both regression and classification problems. It aims to build a training model that can simulate the target variable class by implementing a basic set of decision rules applied from the training results. The target variable of the decision tree can be either categorical or continuous. The decision tree is constructed from top to bottom and is called a greedy algorithm because it often makes a choice that appears to be the right at the moment. The decision tree algorithm begins by setting the root node that is achieved by calculating the entropy (H) and the information gain (G) of the attributes and then choosing the features with the maximum information gain as the root node. The root node is then divided by the selected attribute to create a subset of the results. This method will continue to be replicated on each subset of only attributes that have not been chosen before. The entropy is a measure of the variance of a random variable, while the gain of information is a measure of the shift of entropy. The higher the entropy, the more the information gain.

Entropy is expressed mathematically as

$$H(X) = \mathbb{E}_X[I(x)] = -\sum_{x \in \mathbb{X}} p(x) \log p(x)$$

Entropy using frequency table:

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

Entropy using a frequency table of two attributes:

$$E(T, X) = \sum_{c \in X} P(c) E(c)$$

The figure below shows how a decision tree looks like after it has been built.

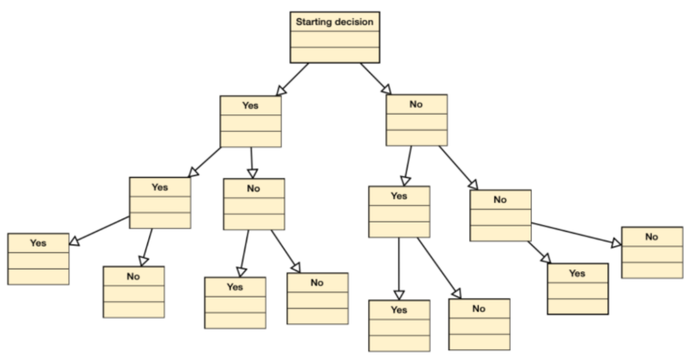


Figure 4: Diagram of a decision tree

3.4 Deep learning Artificial Neural Networks (ANN)

ANN systems are systems inspired by human brains that aim to mimic the way humans learn. It comprises of a large number of strongly interlinked computing components collaborating to solve a particular problem. Each node in one layer is linked to every other node in the next layer. A single layer neural network is referred to as a perceptron, which gives a single output. In a single observation where (X_1, X_2, X_3) are the inputs, these inputs will be multiplied by the connection weight (W_1, W_2, W_3), representing the strength of a particular node. The products are summed up and then passed to a transfer or activation function to generate a result and sent out as outputs [30]. A bias value is usually fixed and allows the activation function to be modified. In the ANN model, the algorithm needs to be trained to describe what is generated as output with regard to the input. In this respect, the difference between the real and the expected value is measured to give the error value called the cost function and is returned to the system to analyze and change the threshold and weight for the next input data. This process is repeated to achieve a smaller or lower cast function than the previous time. This process is called backward propagation and is performed continuously across the network until the error value is held to a minimum.

$$a = f\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

Diagrams 5 and 6 below show a single-layer network (perceptron) and an ANN structure.

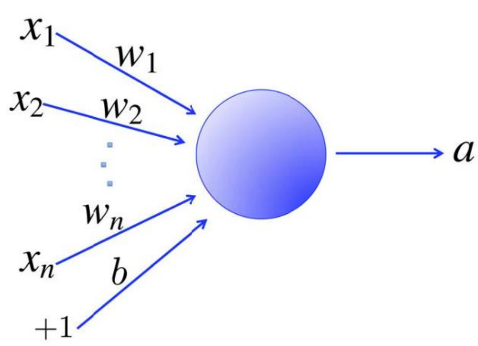


Figure 5: Basic unit in Deep Learning (perceptron)

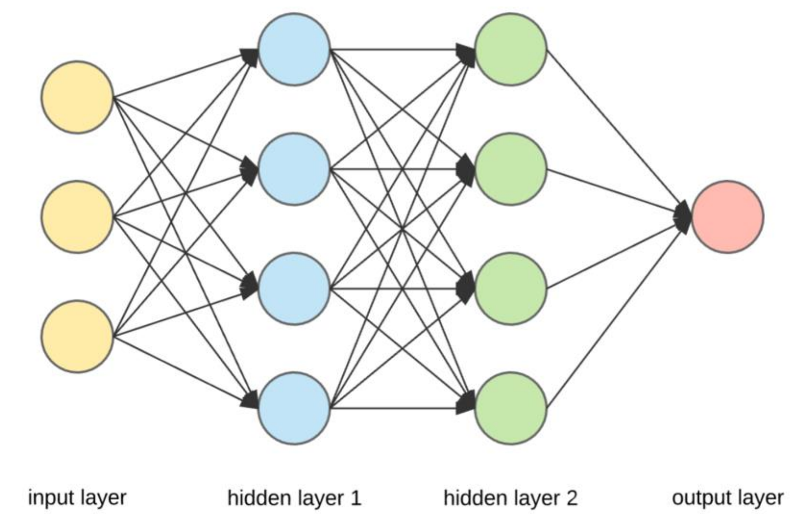


Figure 6: ANN

3.5 Naïve Bayes

The naïve Bayes algorithm is a classification algorithm based on Bayes' theorem. This algorithm makes a general naïve assumption that each feature makes an independent and equal contribution to the outcome, which may not be correct [30]. Bayes theorem finds the probability of an event occurring given the probability of another event that has already occurred. It is represented mathematically as P(B/A) = P(B/A)/p(B). the naïve Bayes algorithm tends to calculate the posterior probability (P(A/B)), which is also known as the conditional probability [30]. Despite this unrealistic presumption, the resultant classifier known as naive Bayes is amazingly effective in reality, sometimes competing with far more complex techniques [31]. Bayes' success in the presence of feature dependence can be explained as follows: optimal zero-one loss i.e., classification error, does not necessarily have to do with the quality of the fit for a probability distribution. Rather, an optimal classifier is achieved as long as both the actual and the estimated distributions agree on the most likely class [31].

The premise of Naive Bayes is that all the features are conditionally independent provided the class label: p(x|y= c)= Y D i=1 p(xi |y= c). While this is typically false (because features are usually dependent), the resulting model is simple to suit and works remarkably well. We get p(x|y= c, θc)= Y D i=1 N (xi |μic, σic) for a Gaussian data, so we just need to estimate separate Gaussian parameters C × D, μic, σic. We get p(x|y= c, θc)= Y D i=1 Be(xi|θic) in the case of binary results, and only need to approximate a different Bernoulli parameters for C × D [32]

3.6 Random forest

At every step, decision trees entail the greedy collection of the best split point from the dataset. If they are not pruned, this makes decision trees vulnerable to high variance. By creating several trees with different samples of the training dataset and integrating their forecasts, this high variance can be harnessed and minimized. A disadvantage of bagging is that each tree is generated using the same greedy algorithm, which ensures that the same or very close split points are likely to be selected in each tree, making the multiple trees very similar (trees will be correlated). In essence, this makes their projections identical, mitigating the initially desired variation. We can cause decision trees to be different by restricting the features (rows) that a greedy algorithm can test at any split point when constructing a tree. This is referred to as the Random Forest algorithm [34].

The random forest classifier consists of a mixture of tree graders, in which each classification is generated with a random vector sampled independently of the induction vector, and each tree casts a unit vote to classify the input vector in the most common class. Random Forest is a computationally effective technique that can be used efficiently on large datasets. It has been used in recent research works and real-world applications in many fields. In most cases, bagging is more reliable than single classifiers. Nevertheless, it is sometimes much less accurate than boosting algorithms.

Furthermore, boosting can create ensembles that sometimes are less accurate than single classifiers. Generally, boosting can overfit noisy datasets, thus lessening their performance. While random forests are more robust than boosting with regards to noise, faster than bagging and boosting, their performance is most times as good as boosting and sometimes even better, and often do not overfit [35].

- Evaluation

  At this stage of the project, models that seem to be of acceptable quality from a data analysis point of view have been built. It is essential to review the models more closely and analyze the steps taken to build the models before moving to the final implementation phase to ensure they meet the business goals appropriately. This phase's main aim is to assess if there is a business problem that was not considered enough. A decision should be made at the conclusion of this point on the application of data mining findings. The four algorithms used in this work performed well on the data set, with all their predictions scoring above 50% in their accuracy. A closer look is taken into their individual performance to see how their scoring parameters and the is evaluated and compared so that the best algorithms can be recommended.

- Deployment

  The implementation stage can be as straightforward as producing a report or as complex as developing a repeatable method for data mining, based on the requirements. In certain instances, the customer will perform the implementation steps, not the data analyst. In all cases, it is necessary to consider in advance what measures are required to make successful use of the models made.

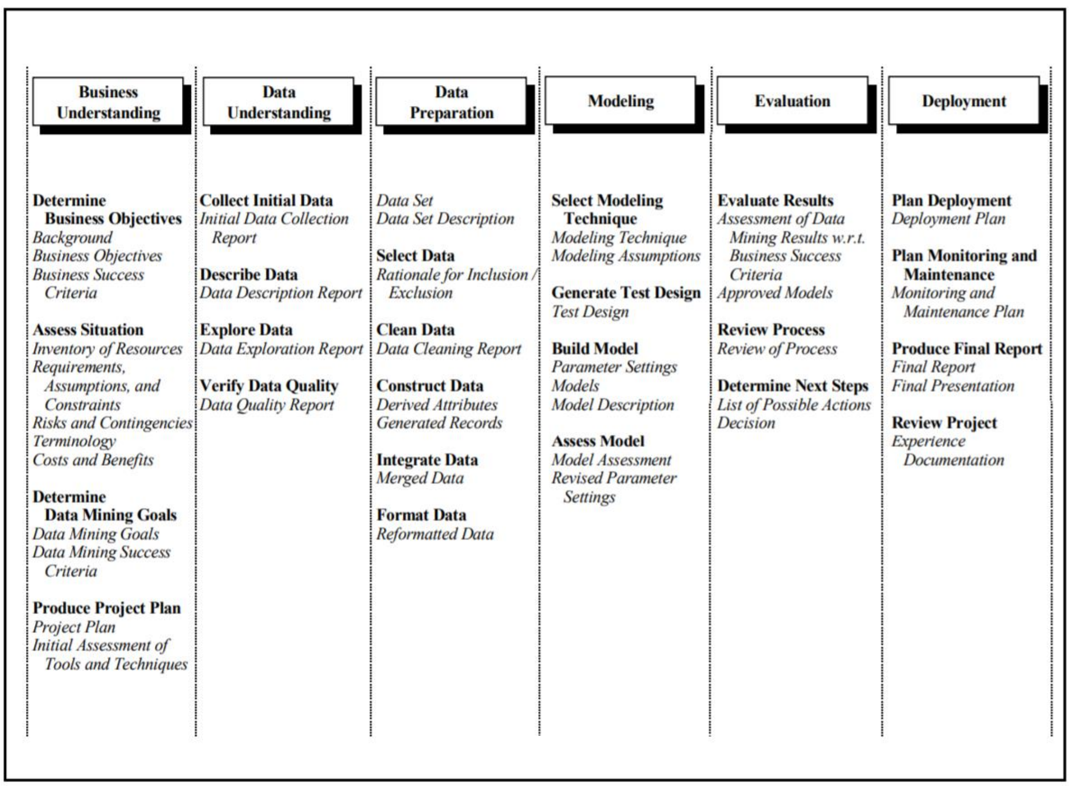


Figure 7: Overview of the CRISP-DM tasks and their outputs.

Chapter 4

4.1 Analysis and Results

In order to determine the most effective model for each of the two groups of items, the results of this research will be reviewed by following the road map below.

- Exploratory Data Analysis
- Model results

4.2 Understanding the Data.

After data collection, an exploratory data analysis (EDA) is carried out on the dataset to evaluate and classify the data's key features by means of visualizations, then data cleaning and preparation is carried out before the models are implemented.

The table below shows the first ten rows of the dataset.

Row No.	47	sales	quantity	description	return	promo	returned	product type
1	42940	66788.596	218	5	2	4	1	0
2	42941	36555.247	100	12	6	4	1	0
3	42942	31633.545	47	5	172	16	1	1
4	36643	23619.720	82	10	641	18	2	1
5	36639	23486.889	48	9	29	17	2	1
6	42943	16196.534	15	12	6	17	1	1
7	42944	14574.647	130	20	9	9	1	1
8	42945	14404.609	25	5	2	5	1	1
9	42946	13380.880	39	7	52	10	1	0
10	42947	12620.585	9	15	3	7	1	0

Table 1: the dataset for the machine learning models

The peaks of the density plot help to demonstrate where the values are clustered over the interval. The advantage of Density plots over Histograms is that they are effective at evaluating the distribution form and they are not influenced by the amount of bins used. Figure 8 shows the density of the attributes.

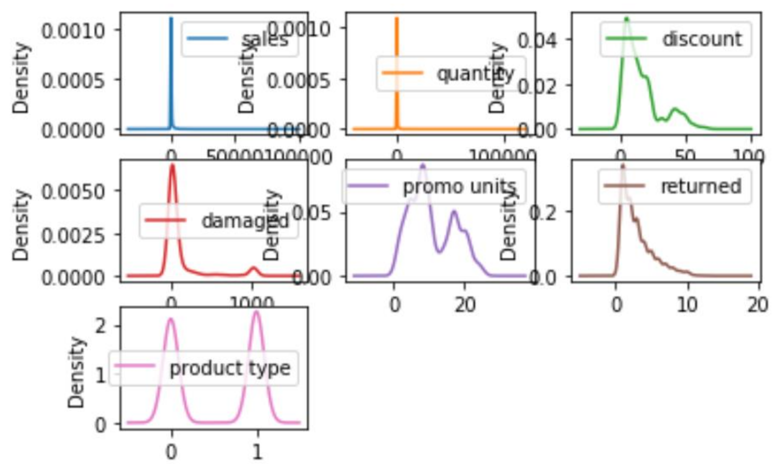


Figure 8: density of the features

To further understand the data, we take a closer look o see the correlations between the features. This is done to ensure that the correlation between them is not too high and that they are suitable for machine learning algorithms and avoid overfitting or underfitting the models. Figure 9 shows the heatmap of the features in the dataset.

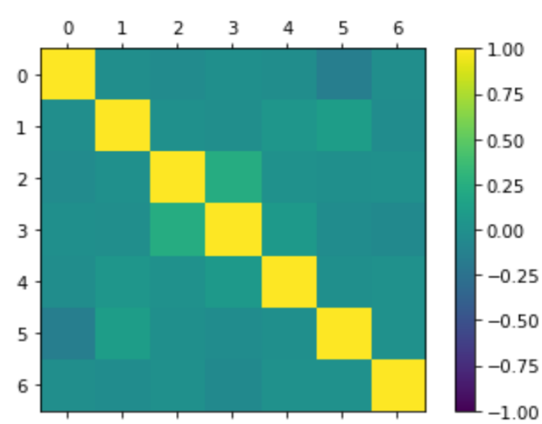


Figure 9: Heat map of the features

4.3 Data Preparation

This stage entails the steps taken to ensure that data is suitable for machine learning. It begins with viewing the data types and checking if there are missing values, then investigating the data type. Figure 10 shows the data type of the features.

```
sales           float64
quantity          int64
discount          int64
damaged           int64
promo units       int64
returned          int64
product type      int64
dtype: object
```

Figure 10: data type

4.4 Rescaling and Normalization

Data rescaling is an essential aspect of data processing prior to the implementation of machine learning algorithms. Data can include attributes with a mixture of scales for different quantities, and this variance affects the results of machine learning algorithms. Thus, rescaling transforms all the features of the dataset into the same scale. Normalization is a scaling method in which values are moved and rescaled such that they end up between 0 and 1. It is also called Min-Max scaling. the normalization formula is given as

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Figure 11 and figure 12 show the first five rows of the rescaled and normalized data respectively.

```
[[5.450e-03 1.907e-03 2.500e-01 0.000e+00]
 [2.500e-03 9.533e-03 2.500e-01 0.000e+00]
 [1.175e-03 3.260e-01 1.250e+00 0.000e+00]
 [2.050e-03 1.220e+00 1.417e+00 1.667e-01]
 [1.200e-03 5.338e-02 1.333e+00 1.667e-01]]
```

Figure 11: rescaled data

```
[[1. 1. 1. 0.]
 [1. 1. 1. 0.]
 [1. 1. 1. 0.]
 [1. 1. 1. 1.]
 [1. 1. 1. 1.]]
```

Figure:12 Normalized data

4.5 Model Performance

Classification algorithms efficiency is generally based on Classification Precision, Accuracy and Confusion Matrix in each class, which indicates the number of predictions of each class that can be compared to each class's instances. The first step is to select the best features for the model. This is achieved by performing an automated feature selection using the univariate method of select kbest technique with f-score as the scoring parameter.

4.6 Parameters for Measuring Models Performance

The following parameters are considered when measuring the efficiency of a classification model.

- Precision: precision is a measure of a classifier's exactness. It is described as the ratio of true positives to the sum of true and false positives for each class.

$$precision = \frac{TP}{TP + FP}$$

- Recall: This is an indicator of the completeness of the classifier; it is the ability of the classifier to detect all positive instances accurately.

$$Recall = \frac{TP}{TP + FN}$$

- F1 score: This is the harmonic mean between precision and recall.

$$F1 = \frac{2 \times recall \times precission}{recall + precision}$$

- Specificity: This is the percentage of the negative predictions that were predicted correctly. $Specificity = \frac{TN}{TN+FP}$

- Sensitivity: This is the percentage of the parties predicted correctly.

$$Sensitivity = \frac{TP}{TP + FN}$$

- Accuracy: This is the overall performance of the model in predicting the classes.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

The performance and confusion matrix of the five algorithms used for this project is shown in the tables below.

Decision Tree Classifier

Measure	Value
Accuracy	88%
F1 score	90%
Recall	100%
Precision	82.%
Roc	96%

Table 2: Decision Tree Classifier Performance

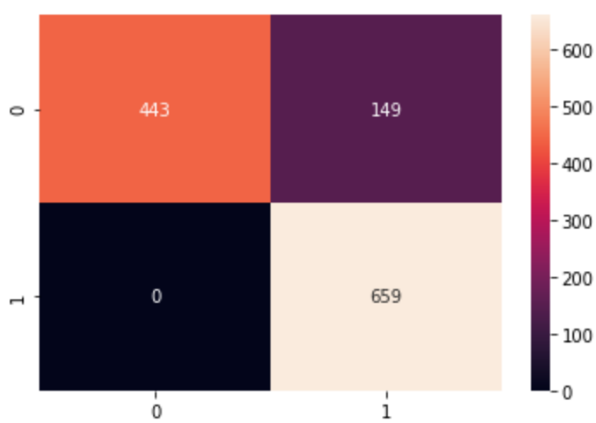


Figure 13: Decision tree confusion matrix

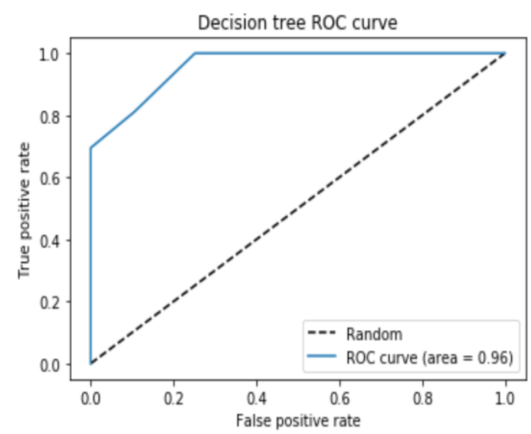


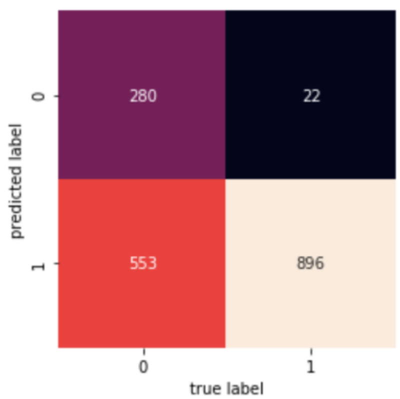
figure 14: Roc curve of decision tree

The decision tree algorithm ran on a split test and training data of 25 and 75 percent, the accuracy produced was 88% with an F1_score of 90% and the recall for the algorithm is 100% and Roc is 96%. This indicates that the algorithm performed well on the data.

Deep learning ANN

Measure	Value
Accuracy	67%
F1 score	76%
Recall	98%
Precision	62%
Roc	78%

Table3: Deep Learning ANN Performance



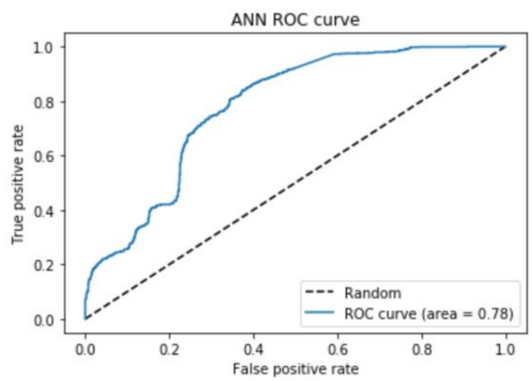


Figure15: confusion metrics of Deep learning ANN figure 16: Deep learning Roc curve

Deep learning ANN performed its classification with an accuracy of 64% and an f1 score of 67%. The recall is at 79%, with precision at 65% and the Roc 78%

Random forest classifier

Measure	Value
Accuracy	98%
F1 score	98%
Recall	98%
Precision	98%
Roc	97%

Table4: Random forest classifier performance

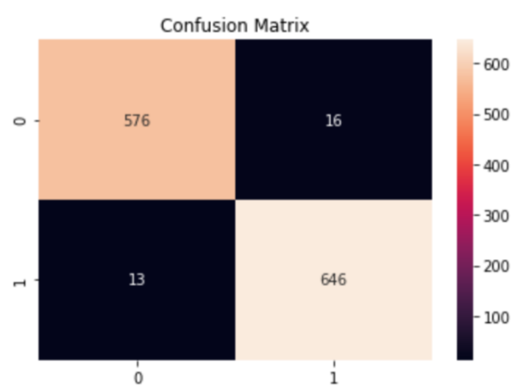


Figure 17: Confusion matrix of Random forest

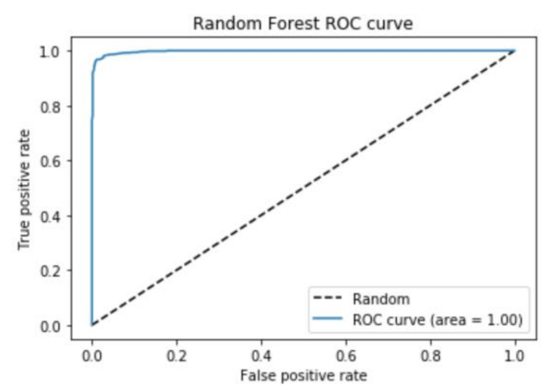


figure 18: Roc curve of Random forest

The Random forest classifier had an accuracy score of 98% with a precision of 97%. The recall for the model is at 98%, with an F1 score of 98% and roc of 100%.

Naïve Bayes classifier

Measure	Value
Accuracy	55%
F1 score	70%
Recall	97%
precision	54%
Roc	71%

Table 5: Naïve Bayes performance

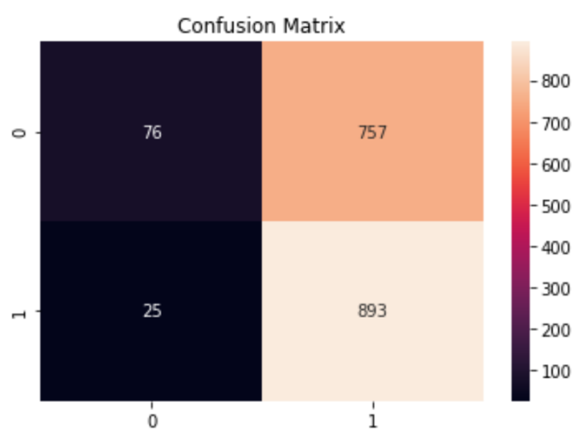


Figure 19: confusion matrix of the Naïve Bayes

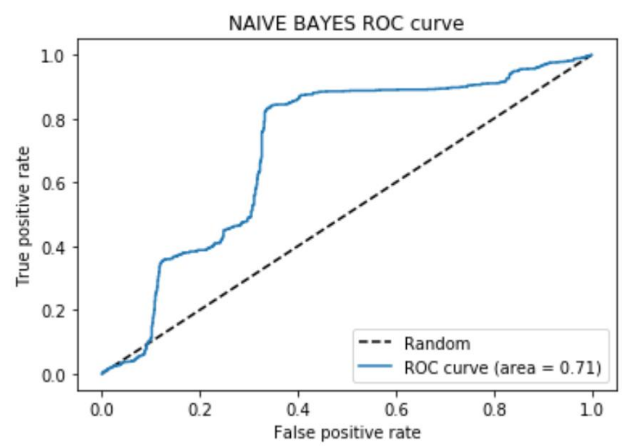


figure: 20 Roc Naïve Bayes curve

The Naïve Bayes classifier had an accuracy of 55% with an F1- score of 70%. The recall for the algorithm is at 97% and Roc of 71%.

4.7 Comparison of the Models

Models	Performance Summary of Algorithms				
	Accuracy	Precision	Recall	F1-Score	ROC
Decision Tree Classifier	88%	82.1%	100%	90%	96%
Deep Learning ANN	67%	62%	98%	76%	78%
Random Forest Classifier	98%	98%	98%	98%	97%
Naïve Bayes	55%	54%	97%	70%	71%

Table 6: comparison of the five algorithms performance

The table in figure 7 shows the performance of all four algorithms. The Random forest algorithm performed better when compared to others. It has an accuracy of 98% and an f1-score of 98%, with 98% recall. The precision for the algorithm is 98% and Roc is 97. The decision tree algorithm followed with an accuracy of 88% and an F1-score of 90% with a recall of 100%. The decision tree algorithms' precision is 82%. The figure below shows the ROC comparison of the models.

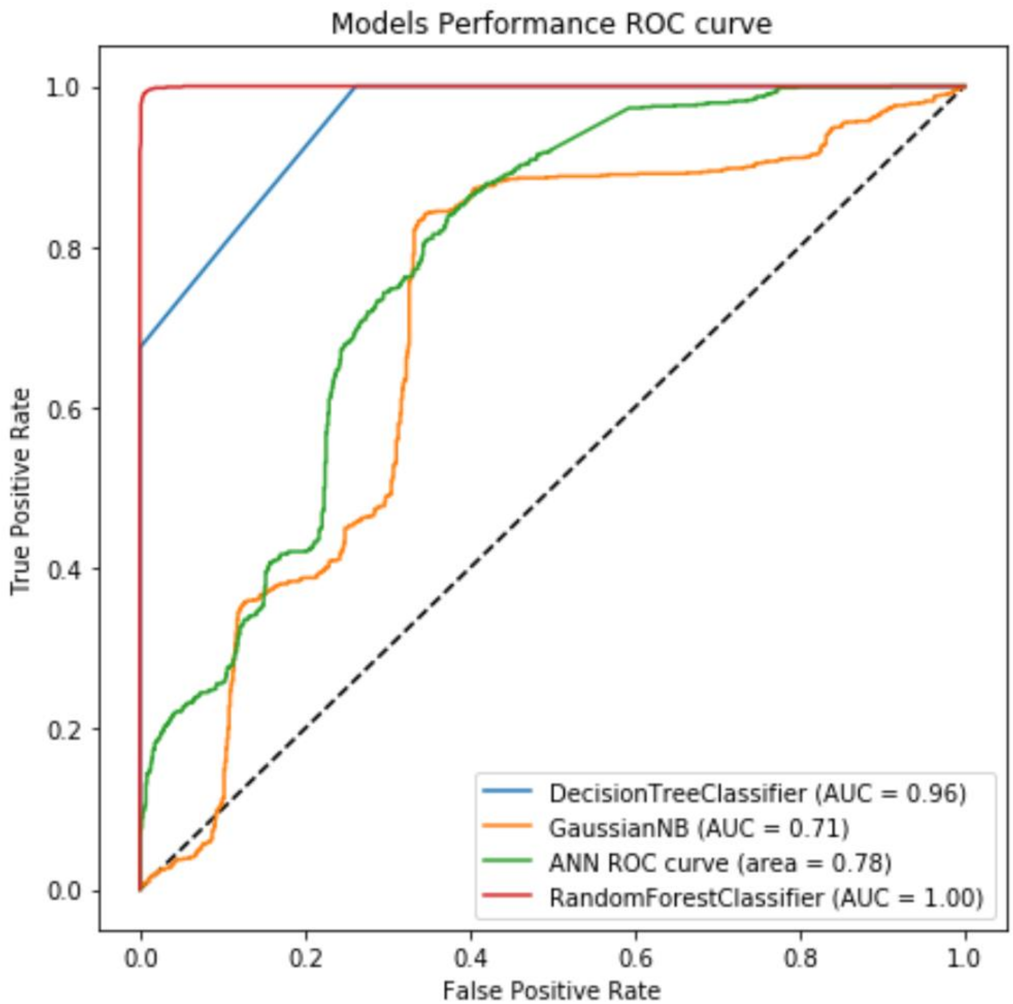


Figure 21: ROC comparison for the five models

Chapter 5

5.1 Conclusion

The research carried out in this project has concluded that product sales prediction systems are needed to manage massive volumes of data by businesses and that company decision-making is focused on data processing technologies' speed and precision. In this study, machine learning methods highlighted can provide an efficient mechanism in data tuning and decision-making. For businesses to be competent, it is necessary that they equip themselves with specialized techniques in order to take account of diverse forms of consumer behavior by making predictions for their product sales and demand to increases profit. Approximately 16,000 records of data instances were collected for this research for the initial algorithm comparison. However, because the implementation time was immense and the handling of such an extensive collection of data was complicated, some of the records were discarded during the analysis and data processing process phase. Simultaneously, the fields and attributes used in this analysis were not appropriate for further analysis. It was a huge obstacle that was faced during the study. However, we carefully weighed our work by introducing effective ML techniques that would work for the problem.

5.2 future work

The work performed in this research suggests the use of four machine learning algorithms to forecast product sales: Decision tree, ANN deep learning, Random forest, and Naive Bayes. Future studies in this area would integrate more complex deep learning models and hybrid algorithms to make predictions and forecasts more accurate. The introduction of more up-to-date models will boost the efficiency for larger datasets and support enterprises and cooperation in decision-making and overall improve customer loyalty and retention.

5.3 limitations

Data acquisition was a major limitation to this project work as the lockdown in virtually all cities around the world due to the covid-19 pandemic hindered business from opening and thus remained closed for business. Another constraint for this research was the computer's computational power to run the phyton codes. The computer used for this work has a 64bit processor and took more time to run the codes and RapidMiner simulations.

References

[1] Krishna A, Akhilesh V, Aich A and Hegde C 2018 Sales-forecasting of Retail Stores using Machine Learning Techniques *Proc. 2018 3rd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2018* 160–6

[2] Majhi R, Panda G, Majhi B, Panigrahi S K and Mishra M K 2009 Forecasting of retail sales data using differential evolution *2009 World Congr. Nat. Biol. Inspired Comput. NABIC 2009 - Proc.* 1343–8

[3] Universitesi D E 2017 COMPARATIVE STUDY ON RETAIL SALES FORECASTING BETWEEN SINGLE AND COMBINATION METHODS **18** 803–32

[4] Jagielska I and Jacob A 1993 A neural network model for sales forecasting *ANNES 1993 - 1st New Zeal. Int. Two-Stream Conf. Artif. Neural Networks Expert Syst.* 284–7

[5] Cheriyan S, Ibrahim S, Mohanan S and Treesa S 2019 Intelligent Sales Prediction Using Machine Learning Techniques *Proc. - 2018 Int. Conf. Comput. Electron. Commun. Eng. iCCECE 2018* 53–8

[6] Wang J and Liu L 2019 A Selection of Advanced Technologies for Demand Forecasting in the Retail Industry 317–20

[7] Arif M A I, Sany S I, Nahin F I and Rabby A S A 2020 Comparison Study: Product Demand Forecasting with Machine Learning for Shop 171–6

[8] Lin K Y and Tsai J J P 2016 A deep learning-based customer forecasting tool *Proc. - 2016 IEEE 2nd Int. Conf. Multimed. Big Data, BigMM 2016* 198–205

[9] Liu N, Ren S, Choi T, Hui C and Ng S 2013 Sales Forecasting for Fashion Retailing Service Industry : A Review **2013**

[10] Padmanabhan V 2014 Comments on " Information Distortion in a Supply Chain : The Bullwhip Effect " Information distortion in a supply chain : The bullwhip effect

[11] Goodwin P 2002 Integrating management judgment and statistical methods to improve short-term forecasts **30** 127–35

[12] Armstrong J S and Collopy F 1998 Integration of statistical methods and judgment for time series forecasting : principles from empirical research 269–93

[13] Sohrabpour V, Oghazi P, Toorajipour R and Nazarpour A 2020 Technological Forecasting & Social Change Export sales forecasting using artificial intelligence *Technol. Forecast. Soc. Chang.* 120480

[14] Anon CRISP-DM The New Blueprint for Data Mining Colin Shearer (Fall 2000) .pdf

[15] Arif A I, Sany S I, Nahin F I, Shahariar A K M and Rabby A 2019 Comparison Study : Product Demand Forecasting with Machine Learning for Shop

[16] Chand N, Adhikari D, Garg R, Datt S, Das L, Deshpande S and Misra A 2017 Ensemble methodology for demand forecasting

[17] Wu Q, Yan H Sen and Yang H B 2008 A forecasting model based support vector machine and particle swarm optimization *Proc. - 2008 Work. Power Electron. Intell. Transp. Syst. PEITS 2008* 218–22

[18] Kilimci Z H, Akyuz A O, Uysal M, Akyokus S, Uysal M O, Atak Bulbul B, Ekmis M A and Silva T C 2019 An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain *Complexity* **2019**

[19] Vkdupd D, Lq D, Frp J J R J, Frp P J, Ru L, Vrxufhv I and Wkhfrpsdq R I 2020 $ylqdvk .xpdu 6kdupd **17** 341–5

[20] Mortensen S, Christison M, Li B, Zhu A and Venkatesan R 2019 Predicting and Defining

B2B Sales Success with Machine Learning

[21] Saena W Predicting Drug Sale Quantity Using Machine Learning

[22] Gallagher C, Madden M G and D'Arcy B 2016 A Bayesian classification approach to improving performance for a real-world sales forecasting application *Proc. - 2015 IEEE 14th Int. Conf. Mach. Learn. Appl. ICMLA 2015* 475–80

[23] Kaneko Y and Yada K 2016 A Deep Learning Approach for the Prediction of Retail Store Sales *IEEE Int. Conf. Data Min. Work. ICDMW* **0** 531–7

[24] Khan M A and Saqib S 2020 Effective Demand Forecasting Model Using Business Intelligence Empowered With Machine Learning 116013–23

[25] Rajab S and Sharma V 2015 Performance Evaluation of ANN and Neuro- 749–54

[26] Islek I and Oguducu S G 2015 A retail demand forecasting model based on data mining techniques *IEEE Int. Symp. Ind. Electron.* **2015-Septe** 55–60

[27] Chen T, Yin H, Chen H, Wu L, Wang H, Zhou X and Li X 2018 TADA : Trend Alignment with Dual-Attention Multi-Task Recurrent Neural Networks for Sales Prediction 49–58

[28] Tsai K, Wang Y, Kuo H, Tsai J, Chang C, Hung H and Shuai H Multi-Source Learning for Sales Prediction

[29] Wirth R CRISP-DM : Towards a Standard Process Model for Data Mining

[30] Deshpande B and Deshpande B 2020 Data Mining : Modelling Classification

[31] Lee E P F, Lee E P F, Lozeille J, Soldán P, Daire S E, Dyke J M and Wright T G 2001 An empirical study of the naive Bayes classifier *Phys. Chem. Chem. Phys.* **3** 4863–9

[32] Sulzmann J N, Fürnkranz J and Hüllermeier E 2007 On pairwise naive bayes classifiers *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **4701 LNAI** 371–81

[33] Brownlee J 2020 How to Develop a Light Gradient Boosted Machine (LightGBM) Ensemble

[34] Brownlee J 2020 How to Implement Random Forest From Scratch in Python

[35] Oshiro T M, Perez P S and Baranauskas J A 2012 How many trees in a random forest? *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **7376 LNAI** 154–68