Revolutionizing Demand Sales Forecasting: A Novel Approach through Ensemble of Statistical Time Series and Machine Learning Techniques

MSc Research Project
Data Analytics

Jyothirmai Myneni
Student ID: X21235325

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……..…………..Jyothirmai Myneni…………………………………………………………………

Student ID: …………………X21235325……………………………………………………………………..……

Programme: ……Data Analytics…………………………………… **Year:** .2022-2023.

Module: …………………………MSc Research Project………………….………

Supervisor: ………………………Vladimir Milosavljevic……………………………………….………

Submission Due Date: …………………………….14/08/2023……………………………….………

Project Title: Revolutionizing Demand Sales Forecasting: A Novel Approach through Ensemble of Statistical Time Series and Machine Learning Techniques

Word Count: ……………9157………………… **Page Count:**……25…

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………………Jyothirmai Myneni…………………………………………

Date: ………………………14/08/23……………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Revolutionizing Demand Sales Forecasting: A Novel Approach through Ensemble of Statistical Time Series and Machine Learning Techniques

Name: Jyothirmai Myneni
Student ID: x21235325

Abstract

Companies in today's fast-paced market have an immediate need to accurately foresee their future prospects due to the volatility of prices brought on by variables like inflation, economic conditions, and client demands. To get over this problem, we implemented a strategy for sales demand forecasting based on time series concepts. We test and contrast a variety of different Time-Series models to find the one that is most effective at forecasting future sales based on consumers' requirements and wants. Random Forest, Linear Regression, Decision Tree, Support Vector Machine (SVM), Gradient Boosting Regression, LSTM, and ANN are just few of the machine learning models we use in our research. ARIMA and Prophet are two examples of classic time series models. Measures of error like as Mean Squared Error, Mean Absolute Error, and Mean Absolute Percentage Error are used to evaluate the quality of each model. Our goal is to find the best model for Time-Series analysis, and we want to achieve so by conducting a thorough examination. Informed by such accurate forecasts of future business possibilities, companies may then make investment decisions that are in line with the needs of their target demographic. Our findings will allow businesses to anticipate and adapt to changing market conditions, allowing them to maximise their returns on investment.

1 Introduction

Demand forecasting is the method used to predict sales of a good or service in a certain market or industry. Predicting future demand is a crucial part of any successful company strategy. Demand forecasting is the practise of estimating the quantity of a good or service that customers are projected to buy over a specific time period, often in the near future. Organisations can't make informed choices regarding production, inventory, pricing, marketing strategies, or resource allocation without first doing demand forecasting. In order to predict what consumers will want in the future, businesses might use any number of methods. It is possible to utilise anything from simple qualitative methods based on expert opinion and market surveys to intricate quantitative methods employing statistics and mathematical models. Common approaches include things like time-series analysis, regression analysis, econometric models, and machine learning algorithms. With accurate forecasts of future demand, businesses can streamline their supply chains, save expenses, and ensure they never run out of stock or have too much inventory on hand. Furthermore, it allows them to match production capacity with expected demand, which improves resource utilisation and boosts happy customers. Real demand can still be impacted by external variables like economic conditions, changing client tastes, and unanticipated events, therefore it's important to remember that demand forecasting isn't foolproof.

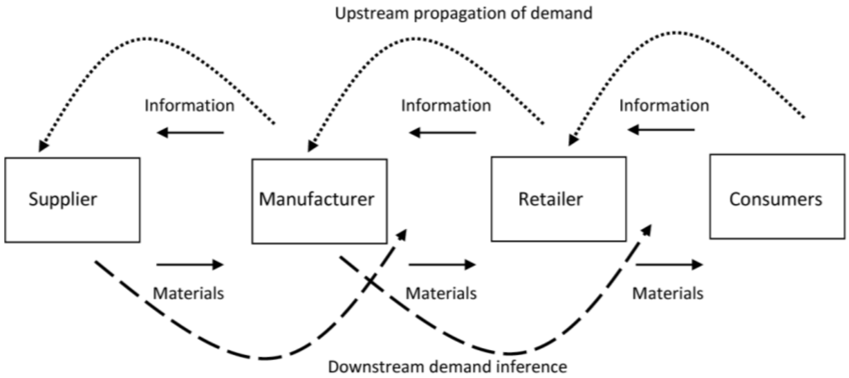


Figure 1: The process of Demand Forecasting in any supply chain industry

1.1 Aim

The study's overarching goal is to construct a robust and reliable demand forecasting model capable of predicting future customer demands across a wide range of goods and service categories. The study's goal is to help organisations with their planning by studying data from the past and using cutting-edge statistical and machine learning methods to draw conclusions. Supply chain optimisation, good inventory management, strategic marketing, and demand-driven production may all benefit from this. The ultimate goal is for companies to have access to the tools they need to make well-informed decisions, adapt their approaches to changing market conditions, and enhance their operational efficacy so that they may better meet the demands of their consumers.

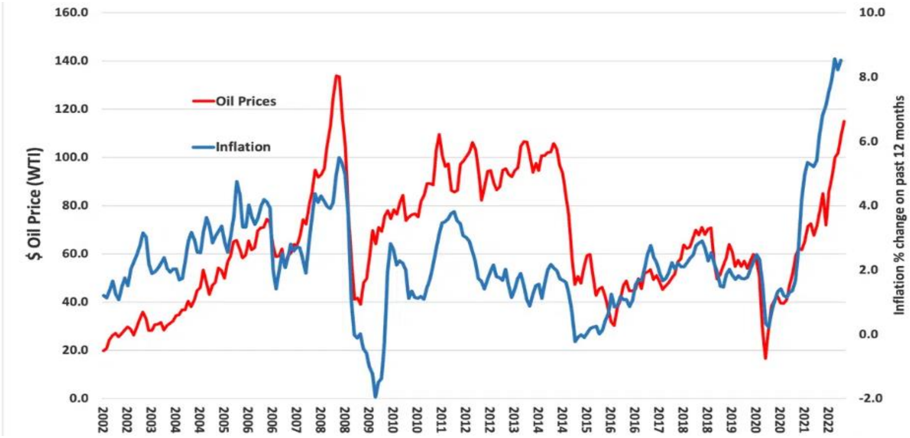


Figure 2: An example of time series data and forecasting using Oil prices and US Inflation Rate (www.economicshelp.org)

1.2 Motivation

There are a lot of compelling reasons that might lead to the creation of a universal and accurate demand forecasting module.

- Better Business Strategy Planning Firms may enhance their resource planning and utilisation with the help of accurate demand forecasts. Lower operational expenses and higher profits are the end outcome of optimising inventory levels, manufacturing schedules, and distribution networks.
- When firms are able to accurately predict demand, and hence store the proper quantity of things at the relevant times, they may increase consumer happiness. Reduced stockouts

and backorders lead to happier customers, who are more likely to remain loyal to the brand.

- Gaining a competitive edge in the market may be possible if businesses used reliable demand forecasting modules. They are quick to respond to market shifts, which helps them stay ahead of the competition and seize emerging opportunities.
- An integral part of efficient supply chain management is demand forecasting, which facilitates better coordination across the whole supply chain. Shorter lead times and fewer disruptions in the supply chain are the outcome of suppliers' improved capacity to anticipate the needs of enterprises.
- When businesses have access to reliable forecasts of consumer demand, they may make better use of their human and material resources. Because of this, resource utilisation is improved, leading to lower expenditures.
- Effective budgeting and financial planning are made possible with the information gleaned from demand forecasting. The accuracy of revenue projections, cash flow tracking, and investment choices are all improved by its use.
- Businesses may cut down on waste if they accurately predict future demand and produce only what is needed. This is an important step towards more sustainable and ecologically conscious company practises.
- Flexibility and adaptability: In a dynamic market, the ability to accurately predict future demand offers businesses the edge they need to swiftly respond to shifting consumer tastes, economic conditions, and industry trends.
- Decision-making Business leaders may reduce the risk of making financially disastrous mistakes or underestimating future demand by making data-driven decisions.
- Company Expansion In the Big Picture Strategic planning, both long- and short-term, relies heavily on accurate demand forecasts. It enables firms to find untapped opportunities, expand into untapped markets, and make well-informed decisions that lead to long-term success.

In today's fast-paced and competitive business world, every organization must have a general and trustworthy demand forecasting module because it not only enhances day-to-day operations but also aids in setting long-term objectives.

1.3 Business Problem

The business problem in the field of forecasting is the need for an efficient and automated forecasting model that can quickly provide reliable and accurate predictions for different domains, such as supply chain or price forecasting. Current approaches need considerable time to create estimates, which might limit the planning team's ability to react swiftly to potential occurrences. The volatility of these systems might also need regular adjustments to the models used to make monthly forecasts. Bad choices might be further complicated by inaccurate forecasting patterns like flat lines. The challenge is in handling large amounts of data for forecasting while guaranteeing accurate, timely forecasts that are in line with actual data patterns. This is a substantial obstacle. Additionally, it is crucial to automatically correct any anomalies or outliers in the data in order to improve forecast accuracy. Most items employ traditional techniques of forecasting, which are limited in their ability to fully capitalise on patterns contained in the actual data points and make use of historical sales data. The best answer should take into consideration bias and volatility, and employ state-of-the-art technologies that make advantage of parallel processing to reduce the time required to accomplish the work. Intelligent preprocessing of data is required to enhance the prediction pattern and give more reliable and accurate forecasts. The difficulty that organisations face calls for a creative forecasting solution that can get them over these roadblocks and into a position where they can make rapid, informed choices.

1.4 Research Question

Based on the above discussions, the research question which is of great importance is given below,

RQ: How can the use of sophisticated machine learning algorithms and time-series models, when integrated together, increase the accuracy and consistency of demand forecasting across a variety of industries?

In this research, we first did a literature review of 15 research papers related to Time-series. In Chapter 3, we proposed the entire methodology upon which this research is based. In Chapter 4, we discuss the proposed implementation network by doing the comparative analysis of all the models we have used here and highlighting the most efficient model among them. In Chapter 5, we would the analysis of the results obtained due to this research and in Chapter 6, we would highlight the conclusions of this research.

2 Literature Review

To attain the objectives of this research on Demand Forecasting using Time-series in the most efficient way we can, here is a literature review of some renowned research papers published by many reputed authors and domain experts who have a lot of expertise in this area of Machine Learning. We have extracted the best out of 20 research papers or articles in this domain and have tried to implement the best model in forecasting the sales demand. Some related research articles which we would consider are as follows-:

(Yoo, T.W. and Oh, I.S., 2020) has proposed a method for forecasting the sales of agricultural products using the SLSTM(Seasonal LSTM)network model to stabilize supply and demand. This is partially a comparative study where they have tried to compare the SLSTM model with other classical time-series models like Auto_ARIMA,Prophet and standard LSTM with reference to three performance metrics like Mean Absolute Error(MAE),Root mean squared error(RMSE) and Normalization Mean Absolute Error(NMAE)where SLSTM has outperformed rest other classical models but it has also highlighted the limitations associated with the SLSTM model in the conclusion.

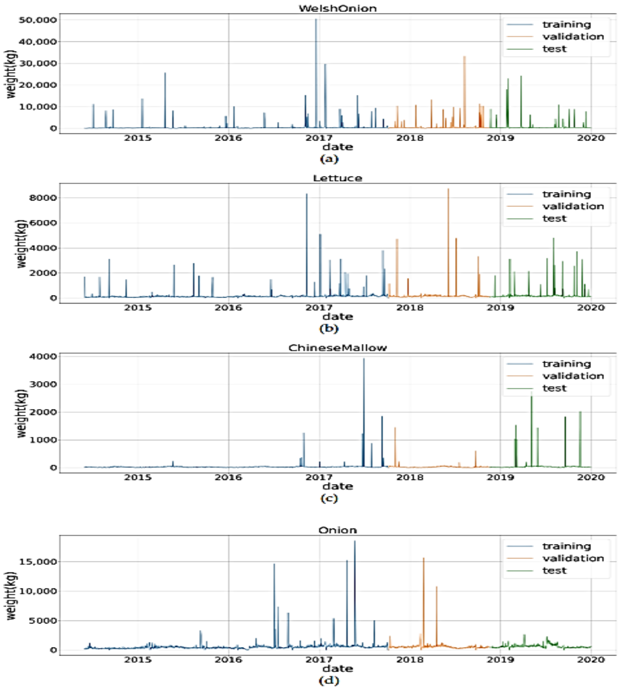


Figure 2: Sales volume of four agricultural items (Yoo, T.W. and Oh, I.S., 2020)

There are 2 significant problems (Arslankaya, S. and Vildan, Ö.Z., 2018) of a genuine production environment which are unknown demand and unbalanced production times.So this paper is aimed to determine the company's automobile sales using demand forecasting methods but as the company doesn't rely upon a single method to resolve this issue,therefore we have done a comparative analysis of Causal methods of Time-series and ANN to find the best demand estimation method where ANN gave the best results.

In order to predict the sales of furniture,(Ensafi, Y., Amin, S.H., Zhang, and Shah (2022) look at the past sales of a retail store. We first employed several traditional time-series forecasting approaches, like SARIMA and Triple Exponential Smoothing, before turning to more sophisticated techniques, like Prophet, LSTM (Long-Short Term Memory), and CNN, to determine the best forecasting modelsIn terms of several accuracy assessments, such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), we compared the performances of the traditional time-series models with those of the more sophisticated models. The results show us that all those neural network methods performed better than the classical forecasting methods and CNN models which were meant to work on image processing gave good results. However, the results of this paper also depend on the selected dataset and so it is required to finetune the parameters before applying it on any model.

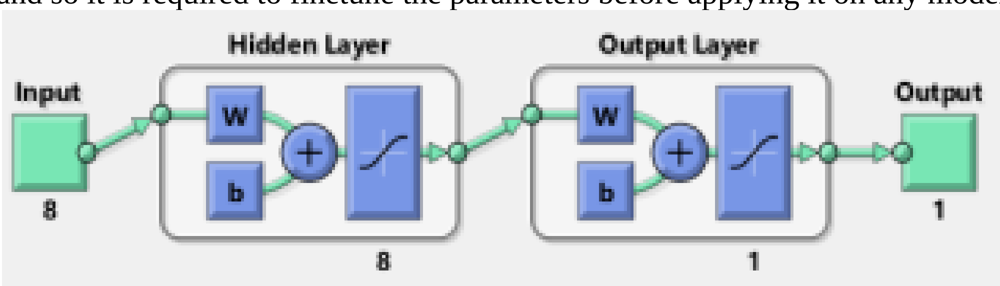


Figure 3: Representataion of ANN (Ensafi, Y., Amin, S.H., Zhang, G. and Shah, B., 2022)

(Ali,.G., Sayn, S., Van Woensel, T., and Fransoo, 2009) suggests utilizing Regression trees with specific features created using sales and promotion time-series from a European grocery shop.Data pooling, as we see, consistently enhances the model's performance.In order to determine the benefit the retailer can receive, we evaluate the accuracy performance of each

model to the benchmark technique of exponential smoothing of each store SKU time series with last like promotion modifications. Here, the system output quality is measured by Mean Absolute Error (MAE).
Due to digital revolution (Permatasari, C.I., Sutopo, W. and Hisjam, M., 2018), there has been a shift in reading hard copy of newspapers to reading electronic news which has a negative impact on printed newspaper demand where there is an inaccuracy of supply with respect to demand. We use the ARIMA Time-series model here in predicting the fluctuating demand and supply of printed newspapers.
(Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. and Seaman, B., 2019) proposes for the development of a globally trained LSTM network that exploits the non-linear demand relationships available in an E-Commerce product assortment hierarchy. Aside from the forecasting framework, they propose a systematic pre-processing framework i.e basically a Heuristic approach to overcome the challenges in E-Commerce business. We have used some variants of LSTM, hyper-parameter selection methods like Bayesian and SMAC and optimization learning algorithms like Adam and COCOB to achieve comparable results on the basis of various evaluation parameters like Mean mMAPE, Median Mmape.

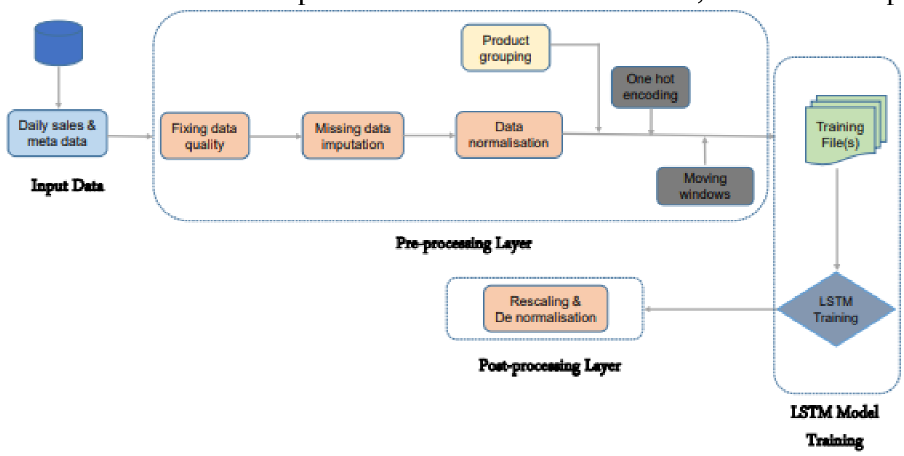


Figure 4: Proposed LSTM network for sales demand forecasting (Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. and Seaman, B., 2019)

For the successful business of any restaurant (Lasek, A., Cercone, N. and Saunders, J., 2016),it is necessary for them to forecast the demand of the customers beforehand so that it can improve its operations management, product management and in consequence reducing restaurant operation costs and increasing the quality of service of the food. So this paper has done a literature review of the best models from various renowned research papers and has selected the best model to forecast the demands of the customer.
(Ramos, P., Santos, N. and Rebelo, R., 2015) does a comparative study of state-space models and ARIMA models in forecasting the future retail sales as per the demand of the consumer of five different categories of products namely-Boots, Booties, Flats, Sandals and Shoes. The state-space models and ARIMA models are compared on the basis of evaluation parameters like RMSE,MAE and MAPE on both one-step and multi-step forecasts and their performance is quite similar in both cases.
In order to forecast sales, (Bohdan, M.P., 2019) took into account various key methodologies and case studies. It has also looked into the stacking method for assembling a regression ensemble from individual models. The results collected demonstrate that stacking approaches can be used to enhance the performance of predictive models for time-series forecasting of sales. (Gahirwal, M., 2013) suggests that forecasting approaches be developed in combination to increase forecasting accuracy. It has been compared to the Holt-Winter method and has outperformed it.

(Thiesing, F.M. and Vornberger, O., 1997), Neural Network techniques trained with back-propagation algorithm are applied to predict the future values of Time-series that consist of weekly demand on items in a supermarket. The performance of the networks is evaluated by comparing them to two prediction techniques used in supermarkets now. The comparison shows us that Neural Networks overperform the conventional techniques with regard to the prediction quantity.

(Pavlyshenko, B.M., 2022) outlines a deep learning strategy for predicting non-stationary Time-series with time trend corrective neural model that incorporates a subnetwork block for prediction weight for a time trend that is applied to a forecasted sales value. According to the findings, a trend correction block in the deep learning model can significantly increase forecasting accuracy for non-stationary sales with time patterns.

(Ford, J.B., Honeycutt, E. and Simintiras, A., 2003) paper proposes a combination of clustering methods along with selecting the appropriate model to improve accuracy and to cater for the drawbacks/challenges one finds while employing these time-series models in cloud enabling cloud service providers to better predict and respond to changes in demand and provide more reliable and efficient services to their customers.

(Abolghasemi, M., Hurley, J., Eshragh, A. and Fahimnia, B., 2020) proposes to develop and test a novel regime-switching approach to quantify systematic information/events and incorporate them into a baseline statistical model. The proposed model is validated empirically using sales and promotional data from two Australian companies. This analysis indicates that the proposed model can successfully improve the forecasting accuracy when compared to current industry practice which heavily relies on human judgement in all types of information/events.

(Merkuryeva, G., Valberga, A. and Smirnov, A., 2019) presents some state-of-the-art methods and key challenges in demand forecasting for the pharmaceutical industry where it has considered a real life example of pharmaceutical field and the results of the experimental analysis of the three forecasting scenarios show us that the symbolic regression-based forecasting model provides the best fitting curve to history demand data ,low error estimates across all scenarios and performed experiments and the ability to more accurately predict demand peak sales.

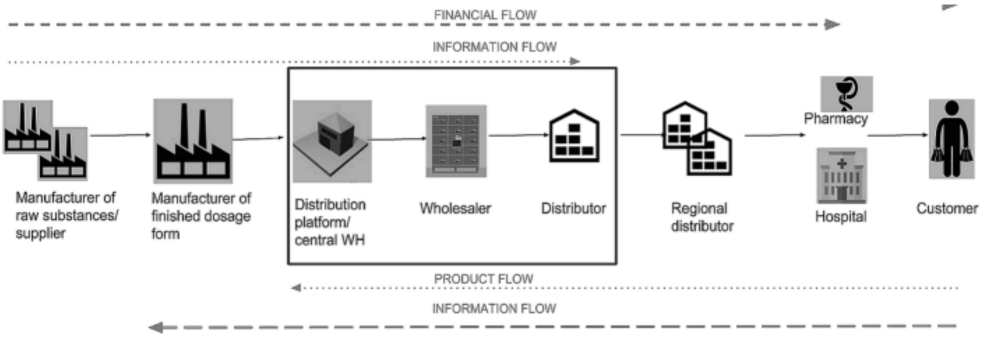


Figure 5: Supply chain of Pharmaceutical products (Merkuryeva, G., Valberga, A. and Smirnov, A., 2019)

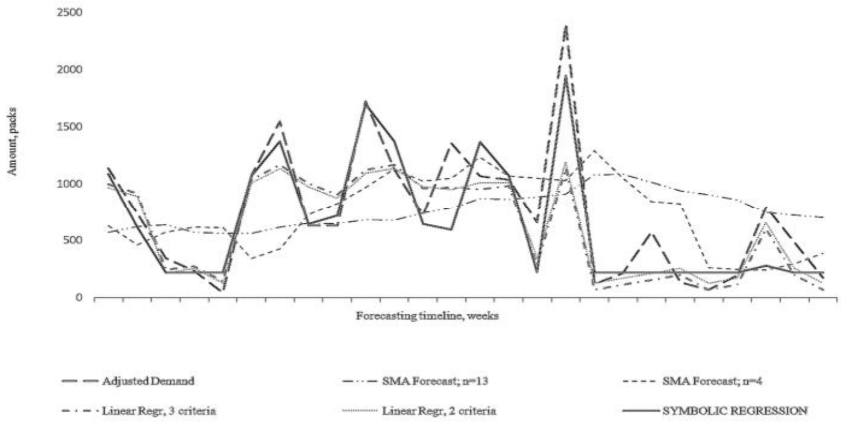


Figure 6: Forecasting timeline for Pharmaceutical supply chains (Merkuryeva, G., Valberga, A. and Smirnov, A., 2019)

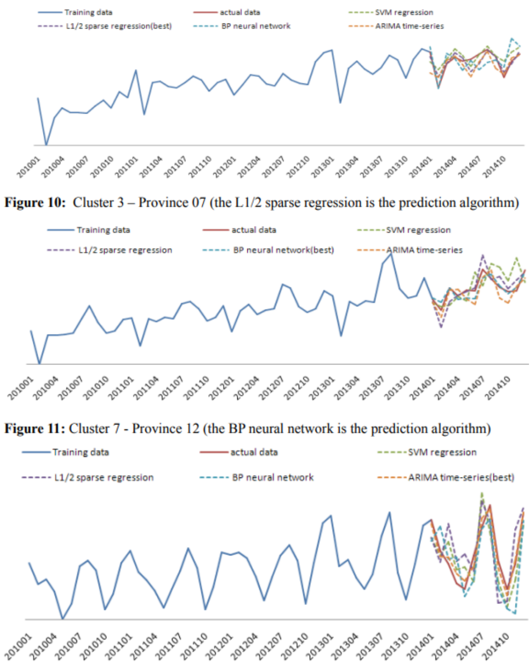


Figure 8: Model performance for Time-series and Regression

2.1 Research Findings

Modern methods, such as Seasonal LSTM, Artificial Neural Networks, and deep learning techniques, frequently outperform older models when it comes to accurately predicting sales demand. While ARIMA models are used to account for shifts in the demand for printed newspapers, regression trees with specified attributes have proven useful for forecasting the demand for SKUs during promotions. When qualitative and time-series forecasting are used together, the accuracy of sales projections for instant noodle products improves thanks to the incorporation of clustering algorithms into the decision-making process. Strategies that include shifting between multiple regimes may considerably increase the accuracy of sales forecasts during promotions. In general, the academic studies show that optimizing demand forecasting with advanced methods leads to better resource allocation, operational efficiency, and customer happiness for businesses of all types.

2.2 Research Contributions

We suggest tapping into the capabilities of deep learning and attention processes to radically revamp the existing approach to demand forecasting. Inability to capture complex relationships and patterns in the data is a weakness of conventional time-series forecasting approaches. By introducing attention methods into deep learning architectures, we may prioritize important information in the time-series data and zero in on significant patterns that drive demand changes. As a result, we'll be able to realize our aims. Deep learning models incorporating attention processes, such as Long Short-Term Memory (LSTM) and Transformer networks, will be part of the study's design and training. The model will be able to better identify long-term relationships and non-linear patterns because to the attention approaches that allow it to assign different weights to different time steps in the input. To ensure the model's precision and scalability, we plan to employ extensive hyperparameter tuning and other optimization techniques. The proposed model will be compared to the results of traditional time-series forecasting methods in order to demonstrate the new model's superior performance in capturing complex demand patterns, accounting for seasonality, and dealing with anomalies. To improve forecasting accuracy and provide businesses with more reliable and actionable insights, a novel deep learning-based demand forecasting model with attention mechanisms is being developed. The model relies on attentional processes, thus this makes sense. When businesses have a better idea of how demand fluctuates, they can better manage their stock, allocate their resources, and promote their products. In the long run, this improves productivity and gives you an edge in dynamic marketplaces.

3 Methodology

In this section we will discuss on the different processes and the concepts that is taken into the account for the implementation of the demand forecasting of the time series products. In this scenario we are going to deal with the products or SKU's (Stock Keeping Units) which will have the history data based on which the forecasting model will be predicting the sales that is going to happen in the forecasting period that we are taking into the consideration. In the forecasting period the number of forecasts that needs to be generated is dependent on the number of the particular period on which the forecasting needs to be taken into account. In this whole process, the data that is taken into the account are some of the standard datasets like the shampoo sales, air passengers bookings etc. along with some of the real time datasets that we are going to consider. The real time datasets are masked and shared by a research group from Bengaluru, India. In most of the scenarios the forecast that the statistical model (predictive models with the machine learning algorithms) gives are taken as the main forecast. But in this research, along with the forecasts generated, we are even taking into the account the corrected forecasts that the demand planner like to have it as the main forecasts. The whole process that we will be discussing for the business is the CRISP-DM methodology.

3.1 Product Life Cycle

Mostly the SKU's follow a lot of different pattern depending upon the history of the sales pattern. The pattern which the original SKU follow is of primary importance. That pattern if any Model does not follow will be given more penalty than the accurate with which the model is predicting. The SKU's can be considered to be of many different types such as NPI which stands for the New Product Introduction, Seasonal which have the sales depending on the certain period of occasions, Normal which has the pattern in a regular demand pattern etc. The whole of the product life cycle can be put in according to the below figure.

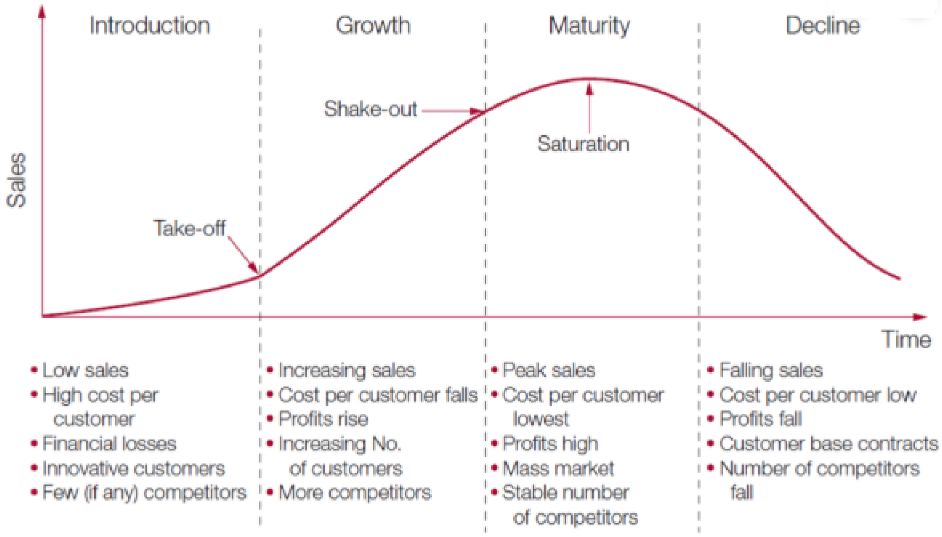


Figure 9: Product Life Cycle for any SKU in a general scenario (De Gooijer, Jan G., and Rob J. Hyndman, 2006)

In the above figure the product sales pattern is shown in which the life of the product is being categorized into 4 parts. The first part talks about the introduction of the product into the market which visualizes a low sales due to the competitor experience, high cost per consumer due to the initial immense high production cost. The product experiences a take-off in the sales pattern due to many advertising effects that the company takes into the account to enhance the sales pattern. This is seen with a sudden rise in the demand of the product which accounts into the increasing sales with a decrease in the cost per consumer. Since the profit directly depends on the amount of the sales and hence the increase in the value over in this phase. This phase marks with the attraction of the competitors views and hence the product faces a lot of competition to sustain its survival in the market. This phase is succeeded with the maturity of the product sales over the time that makes a peak in the value of the sales which marks as the maximum amount of the sales that product can achieve and marks with the saturation of the product in the market due to all the kinds of the steps that is taken for the survival of the same. Since, the volume of the sales is the maximum in its lifetime, the cost per consumer is the lowest. Profit is still in the high range with a stable competition from the market. This is seen with a decrease in the sales pattern that follows it. There can be many reasons such as the changes in the technology, new product with better sustainability etc. The can be another situation of making the sister SKU of the product to make some of the changes in the parent SKU and still make the product present in the market. Some of the reasons can be retaining the loyal customers who needs that particular product and the use of the same sustains. The number of the competitors decrease as evident due to the introduction of the other factors resulting in the decrease in the product sales.

3.2 CRISP DM

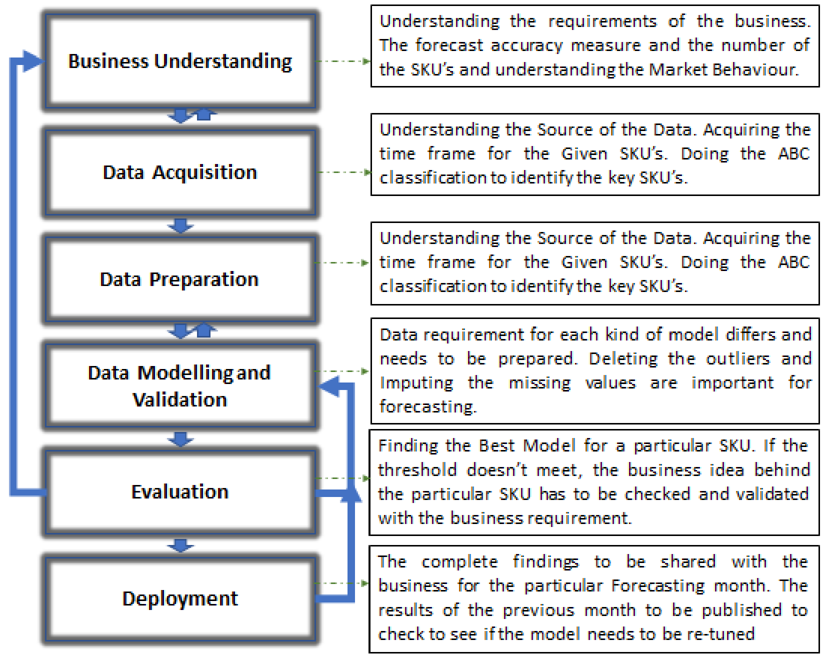


Figure 10: Process of the CRISP-DM (Generated through the Power point)

CRISP-DM is the methodology that is followed in order to carry out the demand forecasting within an organization. This methodology does not only limit itself to the supply chain domain rather to the process that needs a clear understanding of the analytic involvement in the solution. Demand Planners carry out the requirement and decision of the product sales and give the information up the supply chain. To fore see the amount of the sales that is going to come up in the future is one of the primary factor to be taken into the consideration. This factor is helped to analyzed by the use of the demand forecasting methods that helps the planners to plan out the streamlining of the demand curve and the pattern. This planning not only optimizes the inventory handling cost and the warehouse handling cost but also tries to take care for some of the unseen causes such as the demonetization, pandemic and epidemic outbreak etc. This helps not only to have a smooth connection between the demand and the supply of the products but also to take care of the inflation that can happen due to some of the uneven causes. In the above figure, the steps for the CRISP DM are described as below.

3.2.1 Business Understanding

The first step is the understanding the business to know about the objective of the solution and create the hypothesis that is needed to solve the analytical problem and the requirement. In most of the scenarios there is a huge difference in the actual problem statement and the requirements. In this step the problem statement is formulated and the required information is gathered in order to solve the problem. This step includes the way the forecast accuracy needs to be calculated in order to check for the validity and the acceptability of the model. The primary step in the process is to check the acceptable accuracy that the client company wants to achieve through this measure since the accuracy in case of the forecasting model is a swinging value.

3.2.2 Data Acquisition

The second dataset consists of the data that needs to be input to the model and is one of the important factor in building the model for the doing the forecasting. The data can be of different types depending on the kind of data that is present. Some of the examples can be sales in data in which the data is only the sales data that the inventory is going to be having as a stockpile up and that is mostly taking into the account the inventory handling cost. Other can be the type of the data like whether the data is weekly, monthly, yearly etc. The model should be able to understand what kind of the data is input and based on the same the forecast should be done. Once the kind of the dataset is determined, the first phase clustering is done on the basis of the ABC classification. The sales of the last 3 months is taken as the reference and the volume of the sales percentage is calculated. The same is done for all the SKU's and then the arrangement is done on the basis of the same. The top 60% contributing SKU's is taken into the class A and the next 20% is taken into the Class B. The remaining is considered to be in the Class C. For each of the different classes mostly the method of the forecasting is taken to be a bit different depending on the complexity of the data and the model. Since the Class A is having more sales volume, more priority is given a better forecasting technique is to be given over for that class.

3.2.3 Data Preparation

Since the data is considered to be the sales data for the different SKU's the available data consists of only one column of the sales values and the data needs to be transformed to the particular format for the modelling to actually take place. In this step many pre-processing of the data is done to make the data stationarize. The stationarity of the data depends on the different approaches for the missing value treatment, outlier treatment etc. In all the scenario, not only one method is taken into the consideration rather many different approaches are taken into the scenario depending on which the data is fitting into. For the missing value treatment, the nearest neighbours are taken into the consideration and the same month or week of the previous two years or months respectively is taken into the consideration. All the values are then averaged to generate the value for the missing month or week values. The missing values that are considered over here are not only the blank or NaN values rather if the average sale is high for the moving 6 months window and the sale is zero. This assumption is taken into consideration and the correction of the result can be done once the statistical forecast is generated. For stationarity the data we use the Dicky Fuller Test and for determining the correlated Lags to be used for the data ACF (Auto Correlation Plots) are used. To determine the seasonality pattern, PACF (Partial Autocorrelation Plots) are used. The auto correlation plots are used to take a decision on the outliers of the data. The Outlier within the data are taken into the consideration by checking whether the value in a moving 6 month window is greater than the *mean +- 3* standard deviation* or *median +- 3* MAD (median absolute deviation).*

3.2.4 Data Modelling and Validation

Once the data is prepared for the modelling, the data is passed into the different statistical models or the machine learning models to generate the future forecast after doing a lot of validation for fixing the best of the models. In this process the whole of the training is done on the samples which is maintained along with the time i.e. the random sampling is not done rather the top of the previous some samples is taken and reserved for the training. For doing the validation, a kind of different approach is taken and 3 forecast is done in the validation sample that will try to figure out the stability of the models. The generalisation capability of the model is taken into the account using the hyper parameter optimisation. The hyperparameters are optimised using either the Grid Search or Random Search. Once the hyper parameter is optimised the model with the set of tuned hyper parameters is fitted on the training data and predicted on the 3 rolling validation period.

3.2.5 Evaluation

The evaluation of the model performances can take place in different format depending upon the requirement of the client company. In this scenario and in this particular research we will deal with two important aspects of the model behavior on the real time data. One is the accuracy with which the model forecasted the future values of the sales data and the pattern in which the forecast was predicted. Even if in some of the scenario the model performed with a better accuracy but the pattern followed is not correct, then that model didn't performed as per the expectations. The evaluation is done based on the fact that the model is checked on the accuracy based on the demand and operations of the company. If the company performs the production every week and the accuracy needs it be checked every week, then the weekly production is checked with the forecast that is generated weekly basis. The same process is taken into the account if the monthly, quarterly or even annually is the factor on which the company focuses its production.

3.2.6 Deployment

The forecasts that is generated needs to be checked by the demand planners to make the perfect analysis on whether the similar trend is there to place the order to the production unit. Once the order is placed, the production for the same SKU's happens. This is taken into the account the way the forecasts are generated.

4 Implementation

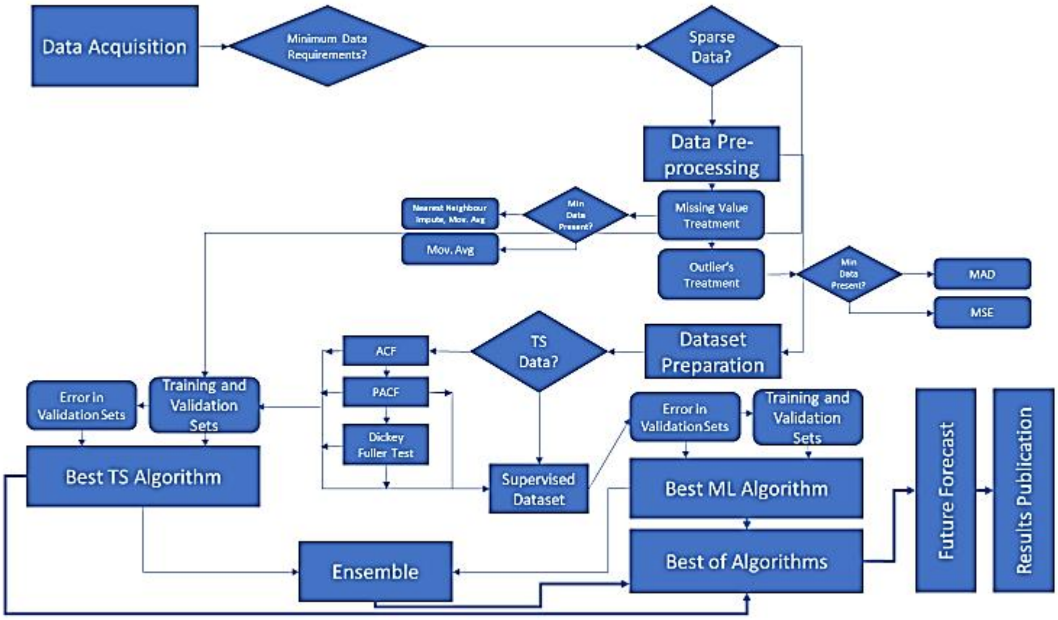


Figure 11: Algorithm Flow

Following is the steps of the flow chart for the forecasting the Sku's data:

Step 1: Data Gathering – Delivering the data to the modelling for the creation of a prediction necessitates the data to be in a specified format, which is accomplished through the first phase of the process, data entry. The DSA, which is shorthand for the data sharing agreement, should be followed in its full during the data preparation process. The data utilised must contain both the actuals and the SKU'S DATA for a given signal name. There are a few additional pieces of data for the SKU's DATA that must be included as well. These often represent the Sku domains of the SKU DATA in question. Signal Number and Location are two such instances. For the business understanding, and Modelling, o **Data Format**: The type of algorithm used in the forecasting process should dictate the data structure that will be used. The majority of the information should consist of the actuals, together with the specifications of the other elements. Due to the nature of the topic under investigation, Demand Forecasting, the bulk of the difficulty we confront is grounded in actual sales. The formulation of the time series problem must account for the external factors as a separate domain if they are to be integrated.

o **Business Knowledge**: Real forecasts can only be reported with confidence if one has a firm grasp of the business world. Essentially, the prediction in this effort is the statistical baseline projection supplemented by some subject skills and knowledge. This forecast being generated right now is the last one, and it includes everything that a good forecast would need to be accurate.

Step 2: Data Clustering – It will be quite time-consuming to get a prediction for all of the Sku's Data, since all of the approaches will need to be applied to it. This is why we recommend using SKU-based data clustering instead. Time and energy are going to be required for this. This is because the modelling process is not carried out on all of the Sku's Data. Instead, some sort of categorization is done so that a more sophisticated method may be used to generate the forecasts. Clustering as a whole may think about factors like price impact to find the best Sku's Data and put them in the A class, then the B class, as the next most important group. The signal data for turning on the green light depends on the time zone and place, and this is high on the list for the volume of Sku's Data at that precise moment. Sku's information and the number of cars per instance are used here, as is the length of time the green light is active. This is because the dataset is intended for use in predicting the number of green-light-following vehicles that will arrive at a given traffic light.

- **Data ABC Classification** – The grouping and classifying procedure entails the following actions: How to Calculate the Percentage of Each Volume • Designate the top 70% as Class A, the following 15% as Class B, and the remaining 30% as Class C. The volume for the most recent "n" period of the Sku's Data should be found and added in its entirety.
- **Classification of Information** Depending on the time span over which the data is collected and presented, different sorts of data will be offered for the predicting. The consistent time frame is the most significant element. If the time intervals are not equal, the data does not fit into a time series and needs to be translated into an equal time period. Minute-by-minute, hour-by-hour, second-by-second, weekly, monthly, and yearly statistics are just a few of the various formats available.

Step 3: Data Pre-Processing – Pre-processing, which involves cleansing the data in a variety of ways, is the first step in the forecasting process. In the process, this is the initial stage. Some of the most crucial intermediate phases are as follows.

- **Step 1**, we employ a technique known as null value imputation, which uses data from adjacent months to arrive at a conclusion. Instead of using the standard statistical technique of imputed by mean or median, this method uses data from the same time period collected in different time zones and imputes it in the same way.
- **Step 2** Outlier Removal elaborates on why it's important to check the mean and standard deviation before imputation in the context of dealing with outliers. These two numbers, however, are reduced to the extremes themselves and must be handled as such. As a result, this problem is sometimes dealt with by using the statistic known as MAD, which is shorthand for median absolute deviation. The time frame, or "window," can be thought of as the waiting times involved. Both of these meanings are correct.

- **Step 3**: Determining Lags Values necessary for formulating the dataset and correlating the most important features may be determined with the help of the ACF (Auto Correlation Factor) and PACF (Partial ACF). In order to predict what will happen next, the time series data uses the current month as a dependant variable on data from the previous month or from the past. This dependence on prior information for the occurrence of an event is the crowning glory of Time Series Data. There is considerable background noise, but the trend, seasonality, and level make up the bulk of the time series data. The time series data must be translated into a stationarized form before modelling can begin.
- **Step 4**: Constructing a Supervised Data Set Using the ACF and PACF factor in addition to the static data, the entire modelling procedure has been transformed into a supervised technique.
- **Step 5**: Data scaling is the fifth step in the process because it allows the data to be BIBO (Bound Input and Bound Output), which is utilised in machine learning regression models to approximate normality. Each algorithm has its own unique approach to scaling, thus having a pipeline to manage this is crucial.
- **Step 6:** Deep learning methods, machine learning with regression, and time series algorithms all require specially crafted training data. Deep Learning techniques can't be used because there isn't enough data.

Step 4: Modelling for Forecast development - The above-mentioned data preparation is now ideal for the development of forecasts. When creating a prediction, the user must specify how many time intervals should be utilised in addition to setting the hyperparameters. This accounts for the potential reduction in error brought on by bias and variation. Algorithms from both the Time Series and Machine Learning families are employed. In order to get a more accurate prediction, an ensemble is created.

- **Step 1**: Validation Sample Generation In order to evaluate the efficacy of the model, a training set and a test set are generated. Here, sample sets are often generated using the Hold Out approach, with either 60%, 70%, or 80% of the full dataset being held out. RMSE is employed in the tuning and error minimization processes. In addition to the above-mentioned metrics, you may also employ MAPE, MSE, etc. In order to ensure the behaviour and stability of the model, the mistake is then validated in the validation or testing sets. After reviewing the existing literature, we factor in these error measurements.
- **Step 2**: Time series algorithms such as ARIMA (Auto Regressive Integrated Moving Average), ARMA (Auto Regressive Moving Average), MA (Moving Average), WMA (Weighted Moving Average), ETS (Expectation to Saturation), Holts-Winter, Croston, Linear Trend, Naive Forecast, etc. The optimal model that yields the smallest validation error using the aforementioned data set.
- **Step 3**: Linear Regression, Decision Trees Regression, Support Vector Regressions, Passive Regression, etc. are all examples of Machine Learning-based Regression Algorithms. In the same way as in the preceding technique, the error is used to decide which algorithm to use. In creating the dataset, we used the ARIMA's p value as our confidence interval.
- **Step 4**: When the supplied training period is lengthy, DL algorithms such as LSTM-RNN and ANNs such as Multilayer Perceptrons are employed for forecast generation

in order to take into consideration the hidden features available in the data [Future Scope of Work and Computation of the system].

- **Step 5**: Group of the aforementioned algorithms: The weights for each of the aforementioned algorithms are produced using the mistakes from the Time Series, Machine Learning-based regression, and Deep Learning models created in the preceding steps:

$$w_{ts} = \frac{1}{error_{ts}} / \left(\frac{1}{error_{ts}} + \frac{1}{error_{ml}} + \frac{1}{error_{dl}}\right)$$
$$w_{ml} = \frac{1}{error_{ml}} / \left(\frac{1}{error_{ts}} + \frac{1}{error_{ml}} + \frac{1}{error_{dl}}\right)$$
$$w_{dl} = \frac{1}{error_{dl}} / \left(\frac{1}{error_{ts}} + \frac{1}{error_{ml}} + \frac{1}{error_{dl}}\right)$$

Thus, the above-mentioned weights aim to penalize the model that has the worst type performance over the validation period. After certain estimates have been verified, this is multiplied by the corresponding baseline statistical forecasts that have been prepared.

***Step 5:* Evaluation Methodology-** It includes the best model that is chosen from the bag of time series models, the best from the bag of all machine learning models, and the best from the bag of deep learning models. The ensemble methods are used to compare which model has the least error, and the forecasted values and graphs are generated for each model. Each industry has a unique statistic that must be taken into consideration when creating the FACC or predicting accuracy. The performance of the entire process is validated using this.

- $$Forecasting\ ACCuracy\ (FACC) = 1 - \frac{\sum_{k=1}^{n} |deviations_k|}{\sum_{k=1}^{n} |Actuals_k|}$$

The deviation in this case is the difference between the actual values and the predicted values that is greater than one standard deviation, or the absolute difference.

***Step 6:* Production Release for the Final Forecast -** The statistical baseline forecast that the forecasting engine produces, the Forecasted Values, must then be enhanced using particular techniques to get the substantially enhanced acceptable forecast. The following are some of the steps that were taken:

- ***Trend level Correction:*** Depending on the market intelligence, the given figures should also reflect the patterns from the preceding year or so.
- ***Seasonality Correction:*** It is necessary to modify the data for the high performing SKUs if the seasonality for a given product is subpar and does not follow the pattern from the previous year in terms of volume, price, etc.

5 Results

Here, the accuracies of Time-Series forecasting techniques and Regression-Based forecasting techniques and the ensemble model are compared. The forecasts for five months is used to compare the accuracies. The forecasts are generated sku wise. Different organizations follow different measures of accuracy like ‘t-1’ and ‘t-5’ which is the accuracy of the generated forecast after two and five months respectively from the current month. Here, the dataset consists of several skus’ whose names are masked following confidentiality. Here, different statistical measures are followed to compare the accuracy of ensemble method over the time series techniques and regression-based techniques.

5.1 Case 1 – Forecast accuracies for the month of January 2019 using actual data till May 2019

SKU-number	FACC (%)	MAPE
Sku-1	84	16
Sku-2	89	11
Sku-3	92	08
Sku-4	83	17
Sku-5	78	22

The forecast accuracies for five consecutive months are also compared to check the stability.

5.2 Case 2: Real Time Analysis with SAP – APO forecast

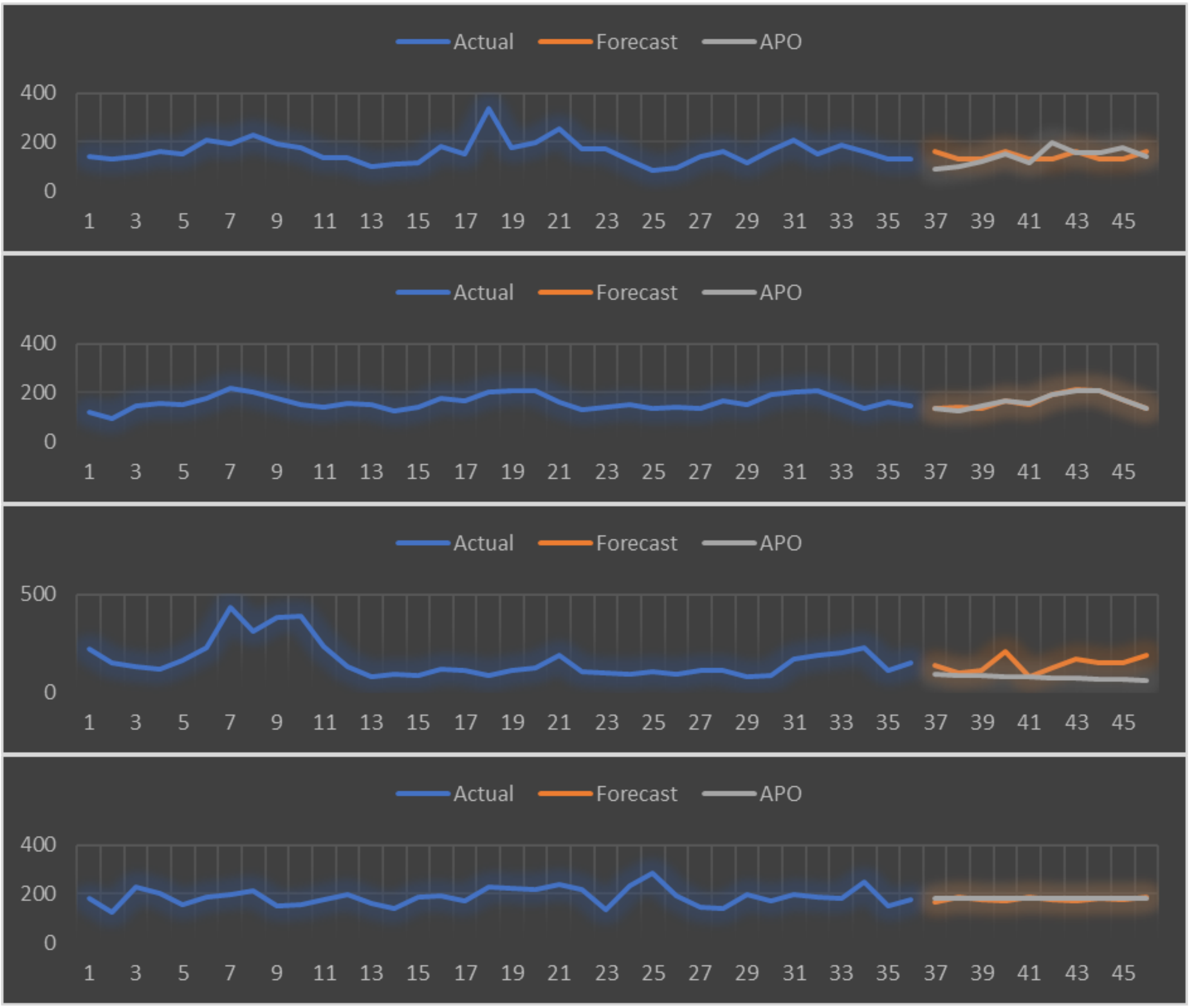


Figure 12: Forecast patterns comparison with SAP APO

In the above figure we can find that the forecast for SAP APO is straight in most of the cases while the forecasts generated by our model is following a pattern. The FACC comparison of our forecast with SAP APO for 1017 SKU's are as below,

Model	Jan-23	Feb-23	Mar-23	Apr-23	May-23
SAP APO	55%	87%	30%	64%	76%
Our Forecast	65%	89%	75%	89%	78%

It can be seen that our model which an ensembled based greedy selection model from different bag of algorithms takes care of the model stability and performs really good in the unseen scenario. Whereas SAP APO most of the time is not able to fit proper and gives a straight line. In all the months our model exceeds the SAP APO by +-10% FACC.

5.3 Case 3 – Accuracy Comparison using FACC measure for Best Models, Greedy Selected of our proposed Model and SAP APO

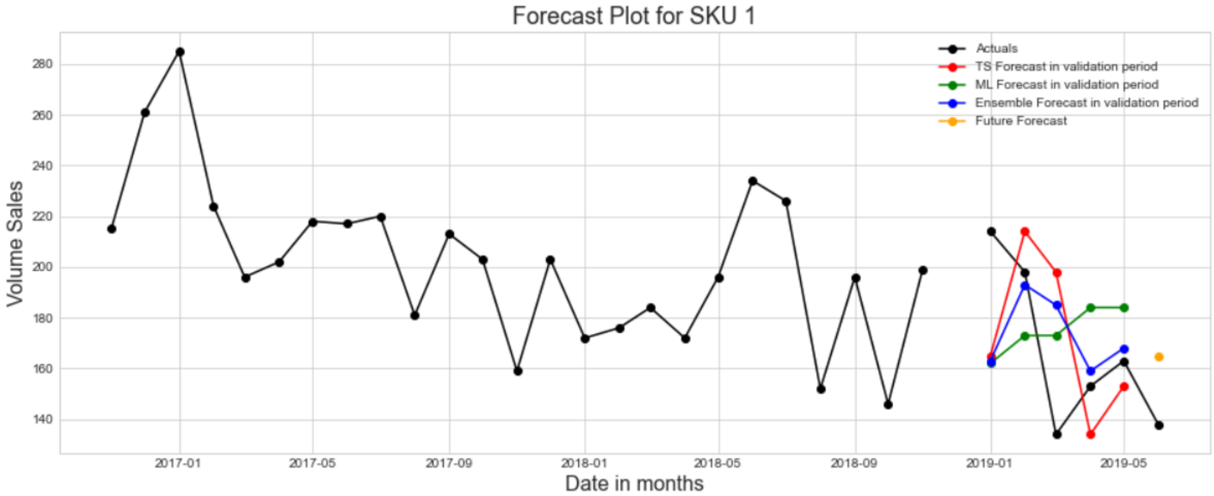
Month	SKU ID	Best Model FACC (%)	SAP FACC (%)	Ensemble FACC (%)
January 2019	Sku-1	77	76	76
February 2019		92	87	97
March 2019		52	71	62
April 2019		88	80	96
May 2019		94	87	97

In the above scenario for one of the major beverage company data, the performance is exceeding from the best of the models and the SAP Models. The ensemble based greedy selected models are best in almost all the months.

5.4 Accuracy Comparison using MAPE measure.

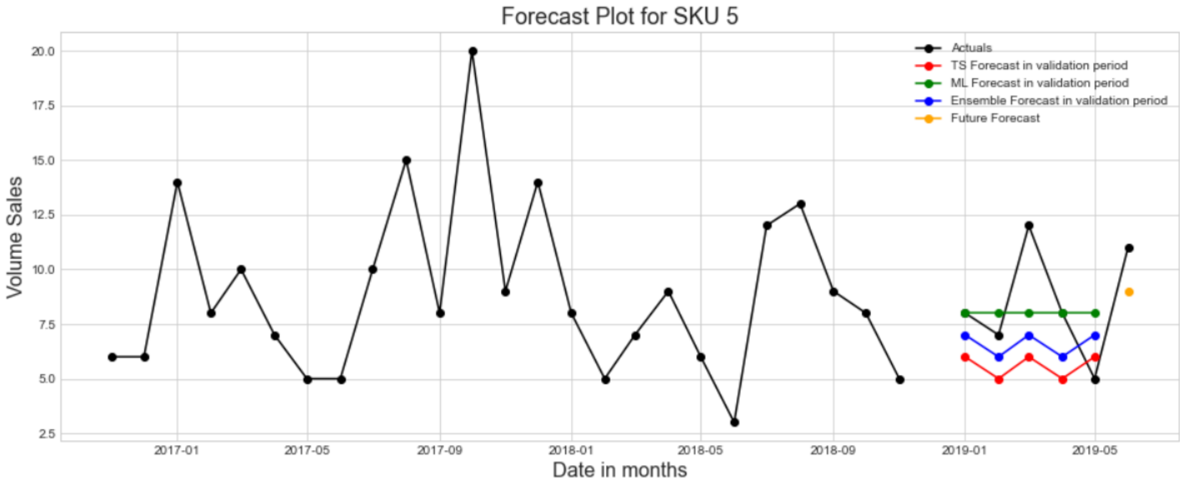
Month	SKU ID	TS MAPE	ML MAPE	Ensemble MAPE
January 2019	Sku-1	23	24	24
February 2019		8	13	3
March 2019		48	29	38
April 2019		12	20	4
May 2019		6	13	3

5.5 Plot for sku 1



The above is the plot of the curve for different modules and the best selected modules for the sku 1.

5.6 Plot for sku 2



The above is the plot of the curve for different modules and the best selected modules for the sku 2.

5.7 Discussion

- **Model Selected** For each of the SKU's are saved with the hyper-parameters and the results of different preprocessing techniques.
- The **Weights Assigned** to each of the model whether TS, Reg-ML or DL are passed as a parameter values in the result output for each of the SKU.
- **Business Domain Knowledge** of giving inputs in terms of hyper-parameters values (Ex: Best p,d,q for ARIMA etc.) or in pre-processing steps (Enrichment in a particular month) helps in the better model performance by improving the weightage of those particular months
- **Clustering of the Algorithms** tries to find the best algorithm from the bag of Similar performing Algorithms and hence **decrease the Computational Time** and **Improve the Accuracy by Ensembling** in the Rolling Validation Window

- The whole process takes care of arriving at a Trade-Off by decreasing the error by reducing the Bias (Profiling/Clustering) and Variance (Rolling Window) using the Greedy Selection of Algorithms

- The whole forecasting method is tested in a major,

  - Automobile (FD) data and the proposed solution is found to be working really fine against the SAP APO from Jan-23 to May-23

  - Beverage Company, The model performance was compared with against the best of the model, SAP APO and Ensembled based techniques proposed in this paper. The performance was checked from Jan-19 to May-19 and apart from March-19 when there was an uneven dip, the model outperformed in all the months.

	Our Solution	SAP APO & IBP	Microsoft Dynamics 365	Llamasoft Demand Guru
Preprocessing	Automated outlier correction & missing data imputation by ML algos'.	Manual outlier correction and preprocessing	Manual outlier correction and preprocessing	Manual outlier correction and preprocessing
Provision for external features	Domain specific external features as well as macroeconomic features	No external features. For short term forecasts distributors withdrawals data can be used in IBP	No provision for considering external features	Macroeconomic features from 550000 time series datasets in own database
Product Classification	ABC/XYZ	ABC/XYZ	Not Available	Not Available
Algorithms	ML, TS and DL based algorithms available	APO – TS only IBP – TS, ML	Statistical TS algorithms available	Several ML algorithms available
Deployment	Cloud	Desktop App. Excel plugin for IBP	Cloud	Cloud
Demand Classification	based on Trend, Seasonality, Maturation, Sparsity	Not Available	Not Available	based on Seasonality and Lifecycle
User Defined Model Parameters	tweak the forecast based on user defined model and parameters	tweak the forecast based on user defined model and parameters	tweak the forecast based on user defined model and parameters	Not Available
Post Processing	Users can manually adjust forecast	Users can manually adjust forecast	Users can manually adjust forecast	Not Available
What-If Analysis	Not Available	APO – N/A , IBP - based on manual change-point inputs	Not Available	can quantify effect of changing values of external factors
Visualization	Simple	Complex	Simple	Complex
Pricing	TBD	Not Available	Not Available	Not Available

Figure 13: A comparison of our model with different market leaders.

6 Conclusion

Forecasting provides relevant and reliable information about the potential future events and their consequences for the organization. From this information might be useful for the organization to take steps for reducing the complications and uncertainty of the future. Hence it is very important to provide accurate forecasts. From the results it is clear that ensemble of the forecast from the times series forecasting technique and machine learning forecasting techniques gives a better as it modifies the bias and variance thus bringing the forecast close to the actual sales. Python platform is used to develop the models. Here, almost most of the techniques in the Time Series and Machine Learning are taken into the consideration, so that one data which might not work fine in one of the algorithms can be used in the other methods. In the future work, profiling and clustering of the skus based on their characteristics will be done and external factors that influences the sales will be taken into consideration and check how the accuracy of the forecasts improves. Some of the USPs are as listed out below:

- The **Weights Assigned is a penalty factor** for the worst performing algorithms and reward for the better performing Algorithms.

- With the **<u>Amount of Data Volume Increase over the time</u>**, the DL algorithms weights will increase, and this will try to capture the harmonics within the data.
- The **<u>Feedback System in the 3 Rolling Forecast</u>** tries to check for the stable model over the validation sample data.
- Measure to check the forecast of each of the top contributing SKU's to see the variation and the forecast for each month/time-period not only in the forecast period but in the Validation period as well.
- **<u>Greedy Search</u>** works not only finding the best of TS, ML-Reg or DL rather between the Best-of-TS, Best-of-Ml-Reg, Best-of-DL along with the Ensemble in the 3 Rolling Validation Forecast.

References

Yoo, T.W. and Oh, I.S., 2020. Time series forecasting of agricultural products' sales volumes based on seasonal long short-term memory. Applied sciences, 10(22), p.8169

Arslankaya, S. and Vildan, Ö.Z., 2018. Time series analysis of sales quantity in an automotive company and estimation by Artificial Neural Networks. Sakarya University Journal of Science, 22(5), pp.1482-1492

Ensafi, Y., Amin, S.H., Zhang, G. and Shah, B., 2022. Time-series forecasting of seasonal items sales using machine learning–A comparative analysis. International Journal of Information Management Data Insights, 2(1), p.100058

Ali, Ö.G., Sayın, S., Van Woensel, T. and Fransoo, J., 2009. SKU demand forecasting in the presence of promotions. Expert Systems with Applications, 36(10), pp.12340-12348.

Permatasari, C.I., Sutopo, W. and Hisjam, M., 2018, February. Sales forecasting newspaper with ARIMA: A case study. In AIP Conference Proceedings (Vol. 1931, No. 1). AIP Publishing

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. and Seaman, B., 2019. Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26 (pp. 462-474). Springer International Publishing

Lasek, A., Cercone, N. and Saunders, J., 2016. Restaurant sales and customer demand forecasting: Literature survey and categorization of methods. Smart City 360°: First EAI International Summit, Smart City 360°, Bratislava, Slovakia and Toronto, Canada, October 13-16, 2015. Revised Selected Papers 1, pp.479-491.

Ramos, P., Santos, N. and Rebelo, R., 2015. Performance of state space and ARIMA models for consumer retail sales forecasting. Robotics and computer-integrated manufacturing, 34, pp.151-163

Bohdan, M.P., 2019. Machine-learning models for sales time series forecasting. Data, 4(15), pp.1-11

Gahirwal, M., 2013. Inter time series sales forecasting. arXiv preprint arXiv:1303.0117

Thiesing, F.M. and Vornberger, O., 1997, June. Sales forecasting using neural networks. In Proceedings of International Conference on Neural Networks (ICNN'97) (Vol. 4, pp. 2125-2128)IEEE

Pavlyshenko, B.M., 2022. Forecasting of non-stationary sales time series using deep learning. arXiv preprint arXiv:2205.11636

Ford, J.B., Honeycutt, E. and Simintiras, A., 2003. Sales management: A global perspective. Routledge

Abolghasemi, M., Hurley, J., Eshragh, A. and Fahimnia, B., 2020. Demand forecasting in the presence of systematic events: Cases in capturing sales promotions. International Journal of Production Economics, 230, p.107892

Merkuryeva, G., Valberga, A. and Smirnov, A., 2019. Demand forecasting in pharmaceutical supply chains: A case study. Procedia Computer Science, 149, pp.3-10