Optimization of Supply Chain Workflow in Food Industry

MSc Research Project
Programme Name

Arpan Jhamb
Student ID: x20140771

School of Computing
National College of Ireland

Supervisor: Dr Majid Latifi

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Arpan Jhamb
Student ID:	x20140771
Programme:	Programme Name
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr Majid Latifi
Submission Due Date:	14/08/2021
Project Title:	Optimization of Supply Chain Workflow in Food Industry
Word Count:	7864
Page Count:	23

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:	
Date:	21st September 2021

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):	

Optimization of Supply Chain Workflow in Food Industry

Arpan Jhamb
x20140771

Abstract

Maintaining the freshness of perishable items like cheese is essential especially for restaurant business. This research is aimed at solving a problem of the restaurant and improving the workflow of the restaurant. Based on the historical sales data, the pizza sales for the next day is predicted, the Cheddar cheese used for making the pizza is calculated using forecasting methods Auto Regressive Integrated Moving Average (ARIMA) and TBATS. The performance of both the models are compared using different evaluation parameters where ARIMA outperformed TBATS and used to forecast the sales for the next day. An automated email system is created for sending the notification to the owner of the restaurant stating the quantity of cheese required for the next day. Based on the results obtained, this research can be deployed in the real world for solving the limited space constraints and providing timely information to the restaurant owner.

Contents

1 Introduction

With the ever increasing global population, the demand for food will certainly outstrip the availability. There is a strong demand for digitization, which employs new technologies such as Artificial Intelligence (AI), the Internet of Things (IoT), Machine Learning (ML). It appears to be the only realistic route out of the continuous cycle of production and waste. To assist the use of resources efficiently, food supply chain management is undergoing systematic adjustments from production and processing to sale.

The flow of products or services from raw materials to finished product delivery to the consumer is known as supply chain management. An effective supply chain instantly translates into rivalry across economic sectors by lowering manufacturing prices. An imbalanced logistics mode mix, high indirect costs, poor connection, inefficient networks, and a lack of technical expertise are all factors to consider for increase in prices.

Artificial intelligence in the f&b industry was estimated worth 3.07 billion dollars in 2020, according to research company Mordor Intelligence, and is anticipated to cross 29.94 billion dollars by 2026, with a CAGR of over 45.77 percent during the projection period of 2021-2026, Kazarian (2021)

Restaurants aren't the only ones who gain from revolution with AI, also benefiting are the suppliers. The ability to observe real-time demand benefits local owners involved in managing businesses like restaurant by allowing them to prepare for labour and materials, as well as reducing the need to accelerate. Suppliers save a lot of money on shipping expenses by minimizing the need to rush, which they may pass on to the restaurants.

The research study addresses the issue of managing the supply chain in a restaurant. The intention behind the research is to ensure that there is an optimum supply of raw materials required for the the restaurant to run optimally and at the same time ensure that the space can be utilized efficiently. This research will ensure that the pizza restaurant owner will have a timely information of the stock of the cheddar cheese used in making pizza. The idea is to allow the pizza owner to forecast the sales of the pizza using time series forecasting model (ARIMA) and TBATS algorithm. The algorithm having a better performance will be used to forecast the sale of the pizza, and the quantity of cheese needed for making the pizza, this information then will be passed on to the pizza owner through automatic mail delivery system.

Since restaurant industry is expanding their operations, with increase in number of cloud kitchens operating and with high cost of the rents associated with the cloud kitchens, there are various challenges including optimization of the inventory, maintaining the freshness of perishable items used for making the food, and ensure a continuous supply and maintain the inventory of raw materials.

This research will ensure that solution to these problems can be solved optimally and provide a real time solution to the problem faced by the restaurant owners. This work is based on real world problem and will solve the problem in restaurants where perishable raw materials are used. This will optimize the workflow in the restaurant.

1.1 Research Question & Research Objectives

To what Extent can Time Series Models Optimize the Workflow in the Restaurants.

The goal is to streamline the workflow in restaurants. Since, this is a unique problem with no precedent, tests will be conducted using a real-world situation. The comparison between two models will ensure better results in terms of accuracy in forecasting and better results.

- **Minimizing inventory:** With the high expenses of storage in urban areas, it's critical to keep inventory to a minimum so that space may be used to its full potential. Price for storage is increasing with the surge in rents. This research will assist in the reduction of inventory storage expenses.

- **Continuous supply of raw material:** The continuous supply of raw material ensures a seamless experience for the restaurant owners and the customers if the supply of cheese is regularised.

- **Preserving the freshness:** Since the raw material is perishable, this study will maintain the freshness of cheese and the pizza.

1.2 Document Structure

The research thesis is divided into different sections: Section 2 lays out the work done in the domain similar to the current study involving application of different algorithms. Section 3 gives an insight of the methodology used for implementation and steps involved based on the methodology. Section 4 lays out the design and the process flow of the research and assumptions are stated required for carrying out the research. Section 5 provides an insight of the experiments conducted and the results obtained through these experiments. The section 6 provides a comparison between the models and the results generated based on different evaluation parameters. In the last section 7 the results, conclusion and scope of Future Work is discussed.

2 Related Work

The transportation of blood in a hospital during emergency is a critical step, it is important to optimize the transportation of blood in a hospital. Abbasi et al. (2020) conducted a research where the machine learning models (MLP neural net- work, CART, Random Forest, and k-NN) were chosen to be trained for optimizing the supply chain in a hospital for transporting blood. The created machine learning models made daily choices on the quantity of units of blood to be ordered as well as those to be transshipped to other hospitals in the network. The simulation was performed for 18,500 days, and the outcomes of various machine learning models and those obtained by assuming availability of the optimizer were compared.

The results demonstrated that, when compared to existing policy, employing a trained neural network model, the average daily cost decreased by around 29%, while the precise optimal policy reduces the average daily cost by 37 %. The final analysis found that hospitals can employ such models as click-and-go software in their daily operations. The research's drawbacks were that it only tested machine learning models on a small collection of data in the case study. Another drawback of our approach was that if the hospital network or demand distributions changed, the machine learning models must be retrained.

The main requirement in a food delivery business is to provide hot food to the customers to ensure customer satisfaction which requires an optimum route for food delivery.

Wang et al. (2021) conducted a research to create a route for food delivery, using the XGBOOST algorithm. To create high-quality routes, the authors used an insertion-based heuristic with multiple sequencing criteria. The authors employed an acceleration approach based on geographical information to speed up the injection procedure, an adaptive selection-based technique was employed, which assures that computing time is saved while preserving solution consistency.

The study also found that utilizing XG BOOST decreased computing time by a significant amount.The study also discovered that combining machine learning technology with optimization algorithms resulted in efficient and effective solutions to optimization issues. The only disadvantage was that this technique was not suitable for dynamic environments.

With increase in population, there is a significant surge in energy requirements around the world especially in developing countries. If the energy requirements can be efficiently managed the resources can be planned optimally. Duc and Nananukul (2020) conducted a research on supply chain management optimization utilizing deep learning techniques to maximize the benefits of meeting energy demands while lowering the net cost of facility charges, extra expenses incurred as a result of unmet demand, as well as supply chain operational costs, a biomass supply chain model named Artificial neural network (ANN) was utilized as a starting point for finding the best solution. To apply optimization models, the IBM ILOG CPLEX was utilized where 1500 random samples were used in the study. CPLEX was also utilized to evaluate the optimization model's best solutions.

The results obtained were utilized as the starting solution of a "Warm start" function, and Relaxation Induced Neighbourhood Search (RINS), a heuristic method, was employed to improve it. This study created a system that blends a machine learning technique with an optimization challenge. As a result, the ANN model was found to be 98 percent accurate in predicting the location of the facility, and the RINS solution to be the best. One of the limitation of the research was the small data size, since the fixed cost associated with plants was huge, small number of plants were chosen.

Effective pricing prediction allows firms to anticipate the price hikes or cuts which will have an impact on customer demand. This ensures the businesses to accurately forecast. Şahinli (2020) conducted a research using Holt-Winters (HW) method for predicting the consumer potato prices using various different values for forecasting. The HW was found to have best accuracy for predicting consumer potato prices. The initial values were optimized using RMSE. The logarithmic consumer prices were used for the selection of ARIMA model. The Akaike information criterion (AIC) and the Bayes information criterion (BIC) were used for evaluation of the best ARIMA models. The mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute deviation were some of the parameters used to choose the best model for forecasting (MAD). When these criteria were compared, the ARIMA (1,1,2) model estimation was the best model in this study since the ARIMA (1,1,2) had lowest error of all the models. Wheat yield forecast is a critical challenge for predicting the demand Farmers can use an accurate prediction model to assist them select what to produce and when to grow it. Wheat production and maintenance play a crucial role. To feed the world's ever-increasing population, large-scale agricultural production, effective management, and distribution are critical. Devi et al. (2021) conducted a research to look at the stability and long-term viability of wheat production in Haryana. The researchers used wheat production data from several editions of the Haryana Statistical Abstract for the study period 1980–81 to 2018–19 in Haryana. ARIMA was used to predict the linear portion of a data series, whereas ANN was used to anticipate the nonlinear part or residuals. The HYBRID technique was then used to capture both linear and nonlinear patterns in one model, with the outcome displaying the accuracy measurements of the selected models using all three approaches (ARIMA, ANN, and Hybrid). The data from the first 34 years was used to create the model, and the data from the final four years was utilized to assess the correctness of the chosen models. With an average relative variation of 4.02 percent and a root mean square error (RMSE) of 581.05., the HYBRID method beat ARIMA and ANN in terms of predicting accuracy. The experiment indicated that a rising trend pattern was seen in the area, production, and yield of wheat crop in Haryana. In terms of area, output, and yield, all sub-periods show positive growth rates.

Restaurant predictive analytics has the power to impact restaurant's critical decisions. A restaurant sales prediction can tell when the optimum time is to establish a new site, how much stock to buy for the coming month and staff planning Zháo and Jayadi (2021) conducted a research to forecast the number of consumers and food demand using point-of-sale records from restaurant chain XYZ for about three years. The authors predicted the number of consumers for three months, and compared it to the actual number of customers during the same period.Using POS data, multiple regression algorithms and SVR algorithms were utilized as prediction algorithms. Seasonality and cook type had a strong correlation with menu demand estimates. The demand for water, Thai Tea, Tea, Potato, Rice, and Beef were highly correlated with the main menu element. Prediction graphs for all menus using a multiple regression model and an SVR model, both with MAPEs of 42 and 31.2 percent, respectively were observed.

The prediction results for the following 15 days after the training procedure for sample menus (two types of beverages and two types of food) indicate good results with a 95 percent confidence level. The MAPE obtained using the multiple regression method is 30.19 percent, 32.83 percent, 37.18 percent, and 43.86 percent, respectively, whereas the MAPE obtained using the SVR algorithm was 28.67 percent, 30.54 percent, 34.63 percent, and 40.39 percent.

Soybean yield forecast is a critical challenge for predicting the demand and supply cycle. Farmers can use an accurate production prediction model to assist them select what to produce and when to grow it. Soybean production and maintenance play a crucial role. To feed the world's ever-increasing population, large-scale agricultural production, effective management, and distribution are critical. Tanong et al. (2021) conducted a research study using monthly time series to predict Thailand's import demand of soybean meal for 10 years. The data were examined using an econometric technique that included time series stationary testing utilizing the ADF unit root and the Box Jenkins forecasting method or the SARIMA (p,d,q) model. The findings revealed that the import demand for soybean meal in Thailand (SBM) had non seasonal stationarity and seasonal stationarity at the level stage or I(d) equal to I(0) and first differencing order or I(D) equal to I(1), SARIMA(0,0,1)(0,1,1) was determined to be the best acceptable forecasting model after evaluating AC and SC criteria as well as autocorrelation between periods, and the import demand for soybean meal in Thailand in 2021 was projected to grow by 1.71 percent over the year 2020. According to this research, Thailand will be exhibiting indications of recovery from the Covid 19 epidemic by 2021. Customer and client satisfaction is enhanced through on-time deliveries, not only it helps in receiving items on time, but also makes customers satisfied. Since some of the food items are perishable, they must be supplied within a certain amount of time. Xing and Cai (2020) performed research into a food delivery route optimization program. Online fast food delivery is one of the fastest-growing industries in the planet. In the food delivery sector, the research was conducted out utilizing a deep reinforcement learning method. The study was focused on delivery time, client wellbeing, enhancing customer happiness, and significantly improving delivery efficiency.

The research compared the traditional tabu algorithm to deep reinforcement learning algorithm in order to find the best path for delivery in metro cities. The findings revealed that the reinforcement algorithm had a faster delivery time than the tabu search method and was a better fit for the task. Forecasting demand is critical in a restaurant business because it helps a company to set the right inventory levels, price its products appropriately, and determine how to grow or shrink its operations in the future. Poor forecasting may result in a loss of sales, exhausted inventories, dissatisfied consumers, and millions of dollars in income. Tanizaki et al. (2019) conducted a research to increase the accuracy of demand forecasting in a ubiquitous environment using Machine learning methods including Bayesian Linear Regression, Boosted Decision Tree Regression, and Decision Forest Regression, while statistical analysis methods include the Stepwise technique. The data was collected for 5 different restaurants for forecasting the results. Based on the yearly data, the number of customers were forecasted for the next year. The observations showed that prediction for Bayesian was highest for store A, The prediction rate utilizing Stepwise was highest for stores B and C. The prediction rate utilizing the Decision method was the highest for stores D and E. The forecasting rate was found to be low for boosting algorithm. Another observation made was that the higher the data usage rate, the higher the forecasting rate in Bayesian models. There was no noticable difference between Boosted and Decision. "The number of customers" and "the number of reserved customers" have a slight correlation. The number of reserved consumers is between 35 and 40 %, which was not insignificant. One of the drawback was that the forecasting accuracy was less.

Customer feedback have become increasingly important in affecting consumer decisions. Creating a review platform saves search costs and allows consumers to swiftly and easily sift information about a variety of eateries. Luo and Xu (2019) The research conducted used reviews from an online platform to test the restaurant performance on the basis of taste/food, experience, Sustainability value, and location. The majority of favorable ratings were about food/taste, whereas the majority of negative reviews are about value. The usage of SVM with FDO algorithm resulted in the highest prediction accuracy (79.59 % and precision rate 81.62 % when predicting the usefulness of restaurant reviews on Yelp in three U.S. cities. The analysis showed that the quality of the cuisine was the most important factor for customers, which is consistent with previous studies demonstrating that food is the most important factor in a restaurant's success. Positive internet reviews are more likely to be generated by good flavor and food quality. Customers tend to exhibit negative sentiments regarding value, contrary to prior studies, which indicated that restaurant ambience had the lowest but still good sentiment score, and customers likely to complain about service quality. The fact that this study included restaurant reviews from three major U.S. cities where living expenditures are high could be the cause.

When three algorithms for predicting review usefulness were compared, it was discovered that SVM with FDO outperformed two other algorithms, NB and a combination of NB and SVM for predicting online review usefulness. In comparison to NB, the SVM with FDO method raised the F1-score, recall, and precision metrics by 11.91 %, 13.31 %, and 10.23 %, respectively. The suggested helpfulness prediction approach was found to be least relevant to predict usefulness of restaurant reviews published in other languages because this study only looked at restaurant reviews written in English. Second, data was collected from only single source with ratings from restaurants in three U.S. cities, limiting the study sample.

The electricity prices prediction plays a significant role in demand and supply activity. If the prices can be forecasted in time it can save significant costs to the government and the electricity providers. It will also benefit the consumers since they will have to pay lower costs. Taş (2018) conducted a research for predicting the hourly electricity price in Turkey. The research was conducted using hybrid models consisting of Seasonal Autoregressive Integrated Moving Average (SARIMA), TBATS and Neural Network. The research highlights that hybrid outperformed the individual models. One of the observations made was that TBAT was more practical than SARIMA model. SARIMA performed better for hourly electricity prices. It was also noticed that TBAT had the best performance among individual models, particularly for one-day forecasts where an unexpected behavior has happened. SARIMA-NARX NN, a hybrid model, provided the greatest forecast performance for both one-week and one-day forecasting. It was concluded from the time series plots of actual values and projected values that some approaches performed better in high-volatility time periods.

Food that is presently thrown can be used to feed hungry families. If food wastage can be predicted, the left over food can help feed hungry people and alleviate food insecurity today. It can also lower the production cost and effort involved. Mihirsen et al. (2020) , conducted a research in order to reduce the wastage in food industry and specifically in restaurants. The author used machine learning algorithms based on time series data, the Holt Winter's and STL algorithms for forecasting sales. The aim was to predict the dish sales over a period which can be used to predict the quantity of raw materials needed. The model was created for each meal using the STL decomposition and ETS time series forecasting methods, and then the model with the lowest RMSE was utilized to estimate the forecast for that meal.The time series assessment also resulted in in warehouse optimization.

The author also made a web application which had five features including login and sign up information, Stock Prediction, Add Dish, Dashboard and donation. The author suggested that this application can be utilized in any restaurant in the globe as long as the restaurant provides data on the meals being made as well as the raw ingredients needed for each dish.

Meller et al. (2018) conducted a research using linear quantile regression (LQR) and tree-based regression (TBR). The experiment used a real-world dataset to test the two models in inventory management problem. The observations made showed that LQRNV operated effectively in circumstances where the feature-demand relationship was linear and the extra noise was not overly large but when the feature-demand connection deviated from a linear pattern, LQR-performance suffered significantly, whereas TBR-performance was unaffected by changes in the relationship's shape. It was also discovered that heteroscedasticity favored TBR-NV and suggested TBR when feature-demand connections was unknown to be primarily linear. TBR-NV outperformed LQR-NV across the board for all of the items I have looked at. It was also discovered that TBR-NV produced superior outcomes even when very minimal data was available to train the models. There were some limitations to the research provided in this publication. It couldn't promise generalization which would hold true in all practically important situations.

Every organization benefits from a sales forecast to help them make better decisions.

It aids in the planning and design, planning budget costs, and managing the risks involved. Sales forecasting also aids organizations in properly estimating their expenses and income, allowing them to foresee their short- and long-term success. Gogolev and Ozhegov (2020) conducted a research to look for the ways on how asymmetric accuracy measures can be used to forecast daily retail sales. The author used different machine learning models to forecast the sales of a store based in Russia, the models used to forecast were linear regression, support vector machine, random forest, and gradient boosting models, the performance of models were evaluated using mean absolute error (MAE) and mean quantile loss functions (MQE)

The sales were forecasted using different models. The performance of RF and GB were found to be best among all the models with the lowest MAE and MQE.The study's key result provided better prediction accuracy when computed as the economic effect of implementing the forecast, which took into account the specifics of food retail.

The summary of the related work is discussed in the table below:

Table 1: Summary of Related Works

Related Works				
Author	Method	Specific Features	Advantages	Limitations
Abbasi et al. (2020)	MLP neural network, CART, Random Forest, and k-NN	Behavior of the transshipment optimization	the average daily cost decreased by around 29%	small collection of data for analysis.
Wang et al. (2021)	XGBOOST	Acceleration approach based on geographical information, an adaptive selection-based technique to save computational time	Time of computation decreased significantly, increase in efficiency and effectivity of the solutions	Technique was not suitable for dynamic environments.
Duc and Nananukul (2020)	Artificial neural network (ANN)	ANN had accuracy of to 98% and the solution from RINS is optimal.	accuracy of ANN was improved using warm start and RINS	sample size of data was less.

Şahinli (2020),	Holt-Winters method and ARIMA	-logarithmic consumer prices were used for the selection of ARIMA model. -The Akaike information criterion (AIC) and the Bayes information criterion (BIC) were used for evaluation of the best ARIMA models	Results generated by HWA had good accuracy for forecasting consumer potato prices	limited data for analysis
Devi et al. (2021)	ARIMA and ANN,	HYBRID method beat ARIMA and ANN in accuracy	rising trend pattern was seen in the area, production, and yield of wheat crop	
Zháo and Jayadi (2021)	multiple regression algorithms and SVR algorithms	demand for water, Thai Tea, Tea, Potato, Rice and beef were highly correlated with the main menu element	The menu demand correlation with seasonality and cook type	Accuracy of the SVM model was less.
Xing and Cai (2020)	Comparison between DRL and Tabu algorithm	delivery time for the reinforcement algorithm was less as compared to tabu search algorithm	reduction in delivery time and improvement in delivery efficiency	limited application on the real world problem.
Luo and Xu (2019)	Comparison between SVM, FDO and NB	quality of the cuisine was the most important factor for customers, positive internet reviews are more likely to be generated by good flavor and food quality	- Practical solution since the problem was based on actual reviews posted	Reviews were restricted to only English language. -Limited data sample.

3 Methodology

There are a variety of data mining techniques that are used in predictive analytics. Most frequently used methodologies are CRISP-DM, SEMMA, and KDD. Since this research is based on solving a problem of a restaurant and is a business oriented problem, I have used Cross-Industry Standard Process for Data Mining (CRISPDM) for the research Europe (2021).

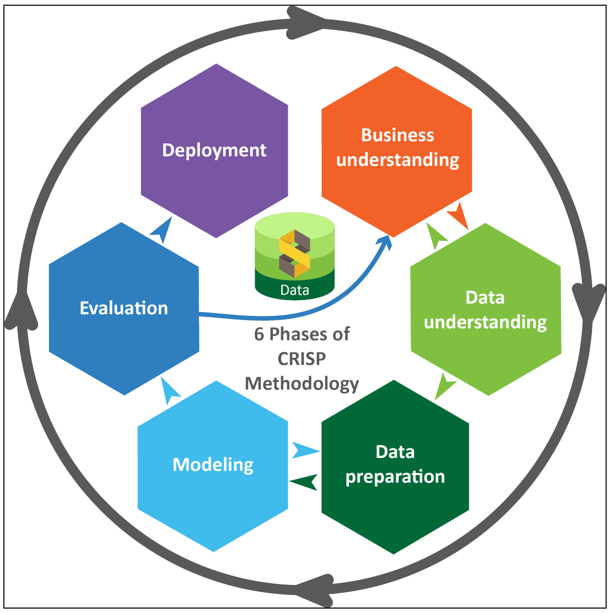

Figure 1: Crisp-DM Methodology

It comprises of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment are the processes involved.

- **Business Understanding**: The objective of any business owner is to predict the sales or demand, this is especially important in the food and beverage industry, where business owners constantly deal with perishable raw materials and unpredictable supplies. If the restaurant owner can accurately forecast the demand of the food, the task of serving customer becomes trouble free. Restaurants which have low margins and are extremely competitive, rely on strict cost-control to be profitable. In the research, an efficient kitchen management solution incorporated will allow planning the storage of perishable raw materials and trigger re-order reminders when inventory goes below specified thresholds. Eateries have benefited, if not been saved, as they change their business strategies in order to not just continue but thrive in this developing digital and cashless world, thanks to technology and creativity. The food and beverage industry has evolved to include everything from online ordering to self-checkouts and contactless payments to delivery and pick-up. The sector will have to adapt to the changing way of doing businesses for assisting businesses in reinventing themselves. Since everything is done digitally it becomes important to transmit information through digital platforms. To make sure everything happens digitally, I have created an auto email system being linked to the pizza owner's email address which automatically sends the email to the restaurant owner stating the quantity of cheese needed for next day. Since cheese is a perishable item, it becomes necessary to accurately predict the quantity of cheese required so that storage of cheese could be optimized and there is enough cheese available at all the times to cater all the customers. This allows restaurant owner to serve the customers without disruption which will help in generating more revenues and maintaining a good relationship with the customers. The aim of this research is the deployment of model in a real world, where the sales are forecasted based on the comparison between the TBATS and ARIMA, this comparison ensured us more accurate results where prediction was based on the model which performed better.

- **Data Understanding**: The understanding of data is crucial to have a clear understanding of what we are trying to achieve. The dataset for the research is of a fast food restaurant chain based in Italy which sells pizza and other food items which require cheese as its raw material. The dataset consists of time series data containing the historical daily sales of pizza for last 18 months. The dataset has two columns where the first column represents the date and second column contains the sales made of pizza on that particular date from the period 1st January 2017 till 3rd July 2018. Since the aim of this research is to forecast the cheese required, it is important to accurately predict the pizza sales accurately. To forecast the sales based on the uni-variate time series data, I have applied ARIMA and TBATS algorithms to predict the sales of pizza for the next day and consequently in calculating the cheese requirement for the next day.

- **Data Preparation**: The dataset was extensively examined and the missing values were identified. This time series data is uni variate since it only has one feature in the dataset for forecasting. Before applying the ARIMA, I checked for various conditions including:

  Check Stationarity: The first step is to look for stationarity.

  Difference: I have done different tests to check the stationarity of the model, the data was checked for stationarity after every differentiation until the data became stationary.

  Filtering out a validation sample: Model accuracy was checked using the validation sample.

  Selecting AR and MA terms: Based on previous values, an autoregression model predicted the current value. The model implied that the time series previous values had an impact on its current value which relates to auto-correlation in the values. However, the correlation is not evaluated between two variables in this case. I have used the Auto correlation (ACF) and Partialcorrelation (PACF) to decide if AR, MA, or both terms needs to be included.

  Build the model The next part is to build the model and to forecast the sales.

  The trend, seasonality, and noise of time series data have been analyzed and plotted to evaluate the trend, seasonality, and noise. The qqnorm plot, qqline plot, and Auto-Correlation function were used to check for linearity and non-stationarity in the data (ACF). For the TBATS model, the tbats() function was utilized, and for the ARIMA model, arima function was used. As mentioned in the literature study, the ARIMA model has a lot of variance. The auto.arima() function selects the optimal model for the data. The accuracy of the forecasting models will be calculated using error parameters once they have been implemented.

- **Modeling**: Since the data contains the historical sales of the restaurant owner, Fattah et al. (2018) has demonstrated the features of ARIMA for forecasting the sales and Versluis (2016) had used TBATS for forecasting the sales of food and results were found to be quite accurate. These models were used to predict the sales of pizza, based on the sales, the model calculated the cheese required for the pizza. This information was then transmitted to the owner of restaurant through machine learning algorithm. After implementation of the forecasting models, the models were assessed on the basis of parameters such as Root Mean Standard Error (RMSE), Mean Absolute Errors (MAE), Mean Standard Error (MSE), Mean Error (ME), Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE). The models with the least accuracy were determined by these error parameters. The model with the best performance was chosen for calculating the sales of pizza and the mail stating the quantity of cheese required was then sent to the owner of the restaurant.

- **Evaluation**: Once the models were applied on the dataset, the model accuracy was measured on the basis on following parameters, the models were assessed on the basis of parameters such as Root Mean Standard Error (RMSE) which is defined as the square root of the average squared error. Mean Absolute Errors (MAE) which is the mean of the absolute error and calculated by dividing the total number of absolute errors by the demand and Mean Absolute Percentage Error (MAPE). It is calculated as the mean of percentage errors. Mean error (ME) was also used for evaluation, the calculated average of percentage errors by which a model's projections deviate from actual values of the quantity being anticipated is known as the mean percentage error (MPE). Based on the evaluation techniques, the model with the best performance was chosen for the calculating the sales of pizza and predicting the quantity of cheese required, the mail stating the quantity of cheese was then sent to the owner of the restaurant. The evaluation techniques provided the performance parameters of the model. In our case ARIMA model performed better than TBATS with the RMSE, MSE, ME, MAE, MPE and MAPE value of the model were found to be less for ARIMA. Once the model was found to be more accurate, I have used the prediction made by the model to calculate the cheese required for the next day. Once the quantity of cheese was known, the mail was then sent to the owner of the restaurant using the auto email system, the password was kept encrypted for privacy and security concerns.

- **Deployment**: The last stage in the CRISP-DM process is deployment of the model in the real world. This model after successful execution can be used in restaurants for predicting the raw material needed and can provide various additional benefits to the restaurant owners ranging from managing the inventory, optimum utilization of space and serving the customers on time. This problem can also be formulated in the businesses which require the inventory management. To run their companies efficiently, the business owners require inventory management to work efficiently, as a result, this technique can also assist them in automating the inventory management.

4 Design specification

The research is based on a real world problem where key focus is to solve the problem of a restaurant owner and assist him in reducing the workload and expenses. The owner will only spend on the raw materials when needed. Since cheddar cheese is an integral raw material used in making pizza and due to lower shelf life of cheese, it's crucial to preserve cheese, due to limited storage space in the refrigerator, the idea is to only order cheese based on the orders for next day.

There are few assumptions needed for successful implementation of the problem:

- The weighing scale is attached to the storage facility which can give the real time indication of the cheese in the storage facility.
- The quantity of cheddar cheese required for making a standard pizza is 50 grams.
- The historical sales data is extracted from the invoice machine.

The dependability of the results obtained is influenced by the research design. The figure 2 shows the flow design of the study and the process design of the research which is explained in the below paragraph.

As per the flow chart 2, In The first step I chose a dataset for the research which complimented the objectives of research which I was trying to achieve. In the next step I have cleaned the dataset, so that the data can be prepared for further transformations. In the next step, I have prepared the data for the modeling, before applying the model, I checked for the stationarity of the data using Augmented Dickey-Fuller Test(ADF) test and KPSS test. ACF and PACF were plotted to check the values of p, d and q. The time series models were decomposed to find out the nature of time series whether it's a multiplicative or additive series. After that TBATS and ARIMA models were implemented. The models performance was measured on the basis of evaluation parameters including RMSE, MSE, MAPE, MPE, and ME. Based on the performance parameters of the models, ARIMA was chosen for forecasting the quantity of pizza sold next day and based on the pizza sold, I found the quantity of cheese required for the next day. The novelty is the transmission of information through email system, which required making a connection with the server and generating the port number. The smtplib module is used for connecting to the client server to verifying login credentials and sending emails. When an email is sent via a secure connection (SSL), port 465 is used. I have used SSL for a secure connection. I have used starttls() function for connection to an SSL server, I have used this technique for improvement in the standard SMTP connection. This research based on the results will be deployed in the restaurants.

5 Implementation

5.1 Data Source & Preparation

The research is implemented using Jupyter-lab. The Python programming language is used for implementing the code. For predicting the quantity of pizza sale, the right set of dataset was found online. The dataset is taken from an open web portal. The dataset is sourced from Purkayastha (2021) in the form of .csv format, it was ensured that all ethical requirements are followed, After exploratory analysis of the dataset, it was found that the

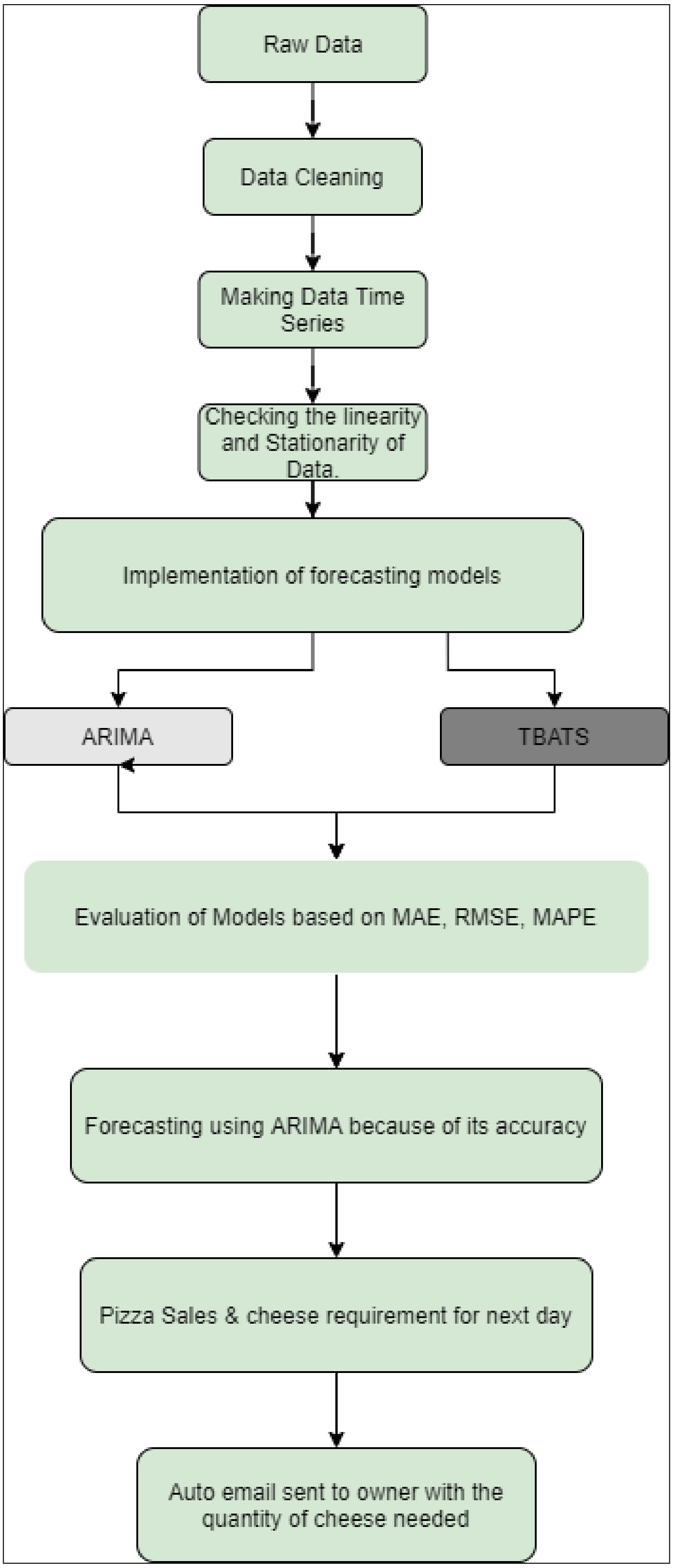

Figure 2: Data Flow Diagram of the Process

dataset contains the information of the historical daily sales of the restaurant which sells pizza. The dataset have two columns, the daily sales number of the year 2017-2018. The dataset contains 1096 values. The dataset had some dates for which sales were missing which were removed using excel. For cleaning and preparation, the data was loaded into Jupyter-lab, after the data was imported into a csv file, a time series object was created for further investigation.

5.2 Analysis, Modelling and Forecasting

There were different packages used for efficiently running the code, some of the packages used were ARIMA, plot_acf, plot_pacf, seasonal_ decompose, mean_absolute_error, math etc firstly these packages were installed and then imported in the system. After that I have plotted the time series model as shown in figure 3 to find out the trends and patterns in the data set. After analyzing the plot it was observed that the sales of pizza shot up in the month of April 2017, it was also observed that generally sales of pizza tends to be more in the month of April, October and November. The sales are minimum in the months between May and July.

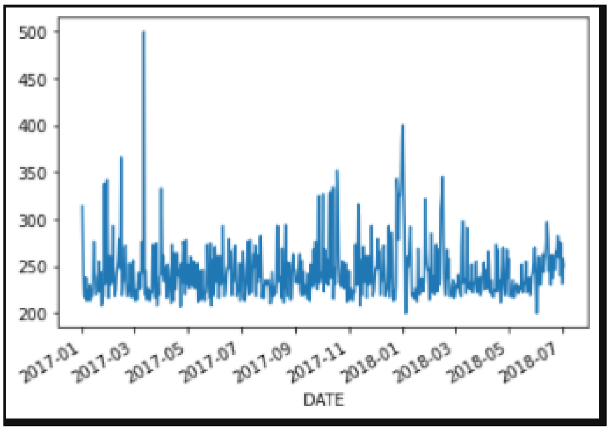

Figure 3: Time series Plot

The next step is to find out the nature of time series if it's additive or multiplicative so that every observation in the series could be written as a sum or a product of the components. I have looked at the components of time series model like trend, seasonality and residuals in time series to decide the nature of time series model. By looking at the seasonality and residual plot, where seasonality is captured better and more closely in multiplicative series than additive series, residual in multiplicative looks random, which gives us a clear indication of the series as shown in figure 4 based on the plot, I have chosen multiplicative time series for this series.

The next step is to check for another property of a time series model which is called stationarity where values remain constant with time. The characteristics doesn't change with time. There are different ways of checking the stationarity of the model. The most common ones are Augmented Dickey Fuller test (ADF Test) and Kwiatkowski-Phillips-Schmidt-Shin – KPSS test (trend stationary). I have used ADF test and KPSS test to check the stationarity of our model.

ADF TEST: The null hypothesis is where the time series is non-stationary and has a unit root, the null hypothesis is rejected if the P-Value in the test is below the significance

Figure 4: Multiplicative Series

level (0.05). In our model as shown in figure 5, the null hypothesis is rejected, which confirms that our model is stationary.

```
Results of Dickey-Fuller Test:
Test Statistic                -1.258263e+01
p-value                        1.895704e-23
#Lags Used                     1.000000e+00
Number of Observations Used    5.450000e+02
Critical Value (1%)           -3.442406e+00
Critical Value (5%)           -2.866858e+00
Critical Value (10%)          -2.569602e+00
dtype: float64
```

Figure 5: ADF Test

KPSS TEST: is a stationarity test that evaluates if a series is stationary around the mean or not.

- The null hypothesis is that the series will remain stationary.

Alternative hypothesis is that the series is non-stationary.

The p-value for this test should be larger than 0.05 if a series is stationary. To put it another way, I want the null hypothesis to be rejected. In our case the null hypothesis is not rejected as shown in figure 6 which confirms that our data is stationary.

```
Results of KPSS Test:
Test Statistic            0.137911
p-value                   0.100000
Lags Used                 9.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

Figure 6: KPSS Test

The next step is to check for the correlation between the past and current values. Three terms define an ARIMA model: p, d, and q where, The AR term's order is p.

The MA term's order is q. The number of differencing necessary to make the time series stationary is denoted by the letter d. THe ADF and KPSS test confirms that the data is stationary so there is no differentiation required. PACF is the correlation between the

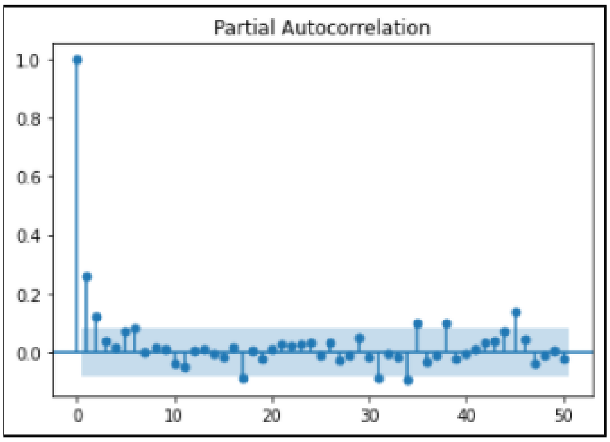

Figure 7: PACF Plot

series and its lag. As a result, PACF encapsulates the pure correlation between a lag and a series. The plot as shown in figure 7 shows that there are two lines that crosses the significance line. Therefore, the value of p is 2. The number of MA terms is shown

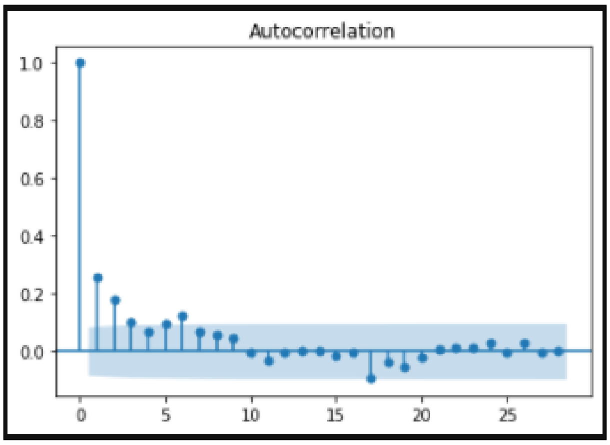

Figure 8: ACF Plot

by an ACF plot. Technically, the inaccuracy of the delayed forecast is referred to as an MA term. The ACF indicates the number of MA terms necessary to eliminate any auto correlation from the stationarized series. It is clear from the figure 8 that two line crosses the significant line.

After analysing the nature of the dataset, The TBATS and ARIMA models were chosen for implementation on the dataset. The reason of choosing TBATS model is because of its advantages over other algorithms Karabiber and Xydis (2019) used TBATS for forecasting the electricity price and the author highlighted the various advantages of TBATS including The TBATS model can handle with complex characteristics such as non-integer seasonality, non-nested seasonality, and large-period seasonality since it has no seasonality constraints, allowing for complete, long-term forecasts. After implementation of models, the performance of both models were evaluated to choose the best model for prediction. The ARIMA model was also implemented using ARIMA() function where the data was split in the ratio of 90 and 10 for training and testing respectively.

Once both the methods were implemented, the models performance were evaluated on the basis of MSE, RMSE and MAPE, ME, and MPE. The performance of each model is discussed in the section 6.

Based on the performance parameters, the ARIMA model being more accurate, was used to forecast the sales of pizza for next day. Based on our assumption that each standard pizza sold requires 50 grams of cheese, I have calculated the cheese requirement for the pizza. To make sure everything happens digitally, I have created an auto email system being linked to the pizza owner's email address which automatically sends an email to the restaurant owner stating the quantity of cheese needed for next day. This email system allows the information to be passed on to the owner of the restaurant in timely manner. The novelty is the transmission of information through email system which required making a connection with the server and generating the port number. The smtplib module is used for connecting to the client server to verifying login credentials and sending emails. When an email is sent via a secure connection (SSL), port 465 is used. I have used SSL for a secure connection, the starttls() function is used for connection to an SSL server, this provides a more secured connection than standard SMTP connection.

6 Evaluation

The objective of the project was to estimate the amount of pizza sales and then compute the amount of cheddar cheese required which depends on those sales. The error parameters RMSE, MAE, MAPE, MASE, ME and MPE were used to assess the models. This evaluation parameters takes the real values input into the model and the model's fitted values, and calculates the difference between the two values. There are different ways how these parameters works and calculate the results. For both the models I have forecasted the quantity of pizza sales for 55 days and compared to the historic sales. The values predicted by the ARIMA was utilized to estimate sales for the next day since it's accuracy was best. The table shown in figure 9 shows the comparison of the values for models. There is a significant difference in MSE and RMSE values where the ARIMA has a better accuracy. The MSE and RMSE values dictate the model performance. The ME value for both the models are negative and there is no significant difference in the values, the MAE values are also similar where TBATS model have an edge over ARIMA in terms of MAE. The MPE values were almost similar for both the models. The MAPE value is better in terms of accuracy for ARIMA compared to TBATS.

Based on the results predicted by these evaluation parameters. It was clear that ARIMA outperformed TBATS. The quantity of pizza sales were made on the basis of ARIMA.

The prediction of the quantity of sales of pizza and cheese required for the next day was done based on the hyper parameter tuning and it was found that the values of (p, d, q) for ARIMA is (6, 1, 0) gave best accuracy. The predictions were made based on these parameters. For hyper-parameter tuning, I have used 66 % training data and 34 % as test. To implement ARIMA model, I have specified AR(p) Auto-regression, I(d) Differencing, MA(q) Moving Averages. In the code snippet refer to figure 10 and figure 11. I have passed different combinations of these parameter (p,d,q) values. As the performance of the forecasting model is evaluated using Root Mean Squared Error

Model Comparison

MODELS	MSE	RMSE	ME	MAE	MPE	MAPE
TBATS	437.57	20.91	-7.78	16.86	-0.02	6.67
ARIMA	318.73	17.85	-0.04	18.95	0.006	6.01

Figure 9: Comparison of Models

(RMSE). Furthermore, to conclude why I have used ARIMA (6,1,0) I can see the best model have got from hyper-parameter tuning is (6,1,0) with least (RMSE) of 26.962. The results are shown in figure 12 which shows that the model has predicted 254 pizzas and the recommended quantity of cheese needed as predicted by model is 12.73 kgs.

```
[73]: # evaluate an ARIMA model for a given order (p,d,q)
def evaluate_arima_model(X, arima_order):
    # prepare training dataset
    train_size = int(len(X) * 0.66)
    train, test = X[0:train_size], X[train_size:]
    history = [x for x in train]
    # make predictions
    predictions = list()
    for t in range(len(test)):
        model = ARIMA(history, order=arima_order)
        model_fit = model.fit()
        yhat = model_fit.forecast()[0]
        predictions.append(yhat)
        history.append(test[t])
    # calculate out of sample error
    rmse = sqrt(mean_squared_error(test, predictions))
    return rmse

# evaluate combinations of p, d and q values for an ARIMA model
def evaluate_models(dataset, p_values, d_values, q_values):
    dataset = dataset.astype('float32')
    best_score, best_cfg = float("inf"), None
    for p in p_values:
        for d in d_values:
            for q in q_values:
                order = (p,d,q)
                try:
                    rmse = evaluate_arima_model(dataset, order)
                    if rmse < best_score:
                        best_score, best_cfg = rmse, order
                    print('ARIMA%s RMSE=%.3f' % (order,rmse))
                except:
                    continue
    print('Best ARIMA%s RMSE=%.3f' % (best_cfg, best_score))

# load dataset
def parser(x):
    return datetime.strptime(x, '%Y-%m-%d')
series = read_csv(historical_pizza_sales_data, header=0, index_col=0, parse_dates=True, squeeze=True, date_parser=parser)
# evaluate parameters
p_values = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
d_values = range(0, 6)
q_values = range(0, 6)
warnings.filterwarnings("ignore")
evaluate_models(series.values, p_values, d_values, q_values)

ARIMA(0, 0, 0) RMSE=30.555
ARIMA(0, 0, 1) RMSE=28.351
```

Figure 10: Hyperparameter Optimization of ARIMA

6.1 Discussion

The research was done to solve the problems faced by the restaurant owners including intense competition in this space with low profit margins and high cost associated with the business. Due to recent success of ARIMA and TBATS in forecasting, these models were chosen for the prediction, where ARIMA outperformed TBATS in terms of accuracy, when the results of both models were compared for this dataset. The predictions were based on ARIMA model which predicted the next day sales for the restaurant and cheese required for next day operations. After hyperparameter tuning, it was confirmed that the best model was with the values of (6,1,0) with least (RMSE) of 26.962. The results indicated that the model has predicted 254 pizzas and the recommended quantity of cheese needed as predicted by model is 12.73 kgs. This information was sent to the owner of the restaurant using auto email systems.

The results will provide an additional medium to support the owners if the model can be employed in the real world. There are a couple of limitations which can halt it's

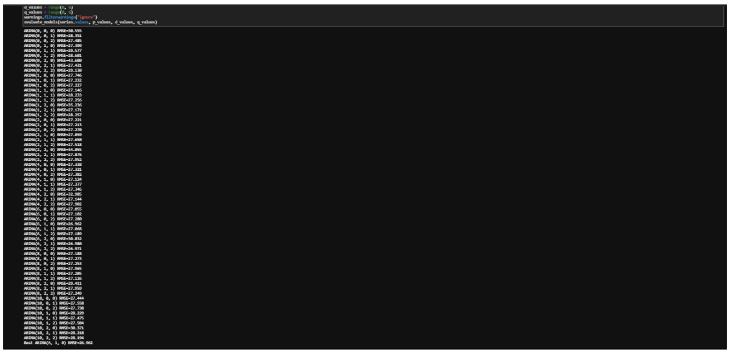

Figure 11: Output of Hyperparameter optimization of ARIMA

Figure 12: Prediction made by ARIMA

application in the real world.

- **High volatility** The restaurant business is very volatile where sales can fluctuate based on factors like seasons, festivals and events like Covid-19. These models are not best in those scenarios.

- **Lack of historical sales data** One of the other limitation is that small restaurant owners don't have the records of historical data needed to forecast the future sales which can go against the deployment in the real world.

7 Conclusion and Future Work

The research achieved the objective of the optimizing the workflow in the restaurants using time series model. The research was based on improving the efficiency of restaurants. The pizza sales for the next day were forecasted using time series models namely ARIMA and TBATS, The models were compared and it was found that ARIMA had better accuracy. The pizza sales and quantity of cheese needed for next day were then predicted using ARIMA, this information was then passed on to the owner of the restaurant using an email system. The research will definitely contribute in improving the efficiency of the restaurant once deployed in the real world.

The research has a wide scope for future work since the restaurant business depends on variety of factors including seasons, festivals and factors like COVID-19, the prediction can be made taking into factors like seasons and festivals using deep learning methods and different time series models. The scope of research can be extended to more number of restaurant chains in order to form a better strategy.

8 Acknowledgement

I would like to thank Dr. Majid Latifi for his constant support, guidance and valuable inputs throughout the research work.

References

Abbasi, B., Babaei, T., Hosseinifard, Z., Smith-Miles, K. and Dehghani, M. (2020). Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management, *Computers & Operations Research* **119**: 104941.

Devi, M., Kumar, J., Malik, D. and Mishra, P. (2021). Forecasting of wheat production in haryana using hybrid time series model, *Journal of Agriculture and Food Research* p. 100175.

Duc, D. N. and Nananukul, N. (2020). A hybrid methodology based on machine learning for a supply chain optimization problem, *Journal of Physics: Conference Series*, Vol. 1624, IOP Publishing, p. 052022.

Europe, S. V. (2021). Crisp-dm methodology.
URL: *https://www.sv-europe.com/crisp-dm-methodology/*

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H. and Lachhab, A. (2018). Forecasting of demand using arima model, *International Journal of Engineering Business Management* **10**: 1847979018808673.

Gogolev, S. and Ozhegov, E. (2020). Asymmetric accuracy metrics in food retail sales forecasting: An empirical comparison, *Higher School of Economics Research Paper No. WP BRP* **61**.

Karabiber, O. A. and Xydis, G. (2019). Electricity price forecasting in the danish day-ahead market using the tbats, ann and arima methods, *Energies* **12**(5): 928.

Kazarian, K. (2021). Artificial intelligence is finding its mark in food manufacturing.
URL: *https://www.foodengineeringmag.com/articles/99365-artificial-intelligence-its-all-about-the-data*

Luo, Y. and Xu, X. (2019). Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp, *Sustainability* **11**(19): 5254.

Meller, J., Taigel, F. and Pibernik, R. (2018). Prescriptive analytics for inventory management: A comparison of new approaches, *Available at SSRN 3229105* .

Mihirsen, D. D., Joseph, J. T. and Renisha, B. (2020). Time series analysis for supply chain planning in restaurants, *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, IEEE, pp. 1–5.

Purkayastha, S. (2021). Pizza sales.
URL: *https://radiostud.io/*

Şahinli, M. A. (2020). Potato price forecasting with holt-winters and arima methods: A case study, *American Journal of Potato Research* **97**(4): 336–346.

Tanizaki, T., Hoshino, T., Shimmura, T. and Takenaka, T. (2019). Demand forecasting in restaurants using machine learning and statistical analysis, *Procedia CIRP* **79**: 679–683.

Tanong, K., Jatuporn, C., Suvanvihok, V. and Seerasarn, N. (2021). Forecasting import demand for soybean meal in thailand using box-jenkins method, *Journal of Hunan University Natural Sciences* **48**(5).

Taş, B. (2018). *Electricity price forecasting using hybrid time series models*, Master's thesis, Middle East Technical University.

Versluis, H. (2016). *Daily sales forecasting in foodservice: developing a model for application in an expert system*, PhD thesis, University of Pretoria.

Wang, X., Wang, L., Wang, S., Chen, J.-f. and Wu, C. (2021). An xgboost-enhanced fast constructive algorithm for food delivery route planning problem, *Computers & Industrial Engineering* **152**: 107029.

Xing, E. and Cai, B. (2020). Delivery route optimization based on deep reinforcement learning, *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, IEEE, pp. 334–338.

Zháo, M. A. and Jayadi, R. (2021). Forecasting daily visitors and menu demands in an indonesian chain restaurant using support vector regression machine, *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, IEEE, pp. 1–6.