Forecast of Consumer Behavior Using Time Series and Ensemble Techniques

MSc Research Project
Data Analytics

Tellaeche Macias Andrés Enrique
Student ID: X21226814

School of Computing
National College of Ireland

Supervisor: Prashanth Nayak

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Tellaeche Macias Andrés Enrique
Student ID:	X21226814
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Prashanth Nayak
Submission Due Date:	14/08/2023
Project Title:	Forecast of Consumer Behavior Using Time Series and Ensemble Techniques
Word Count:	4,862
Page Count:	15

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:	12th August 2023

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):	

Forecast of Consumer Behavior Using Time Series and Ensemble Techniques

Tellaeche Macias Andrés Enrique
X21226814

Abstract

Time series analysis and forecasting is a very important part of machine learning since it has many applications. Time series consists of observations taken in a chronological order. Some of the applications related to time series are the forecasting of the stock price in the stock market, forecasting the demand for certain products, natural language processing, etc. In this research project, we explore the possibility of improving the forecast of the consumer behavior by implementing ensemble techniques. For this reason, we retrieved publicly available data about the sales of alcohol in the United States, particularly in the state of Alabama, and combined them with weather data from the same state. We implemented Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX), Gated Recurrent Unit (GRU), and Long-Short Term Memory (LSTM) separately and then applied ensemble models using the results obtained by the individual models to improve the final forecast.

1 Introduction

The importance of a good relationship with the customers has become a big focus of attention in recent years. Acquiring new customers and retaining the ones that a business already has are two factors that are as important as having products to offer. Having a tool capable of forecasting what the consumers will want is the dream of every business and organization, not only because they could predict the trends and promote themselves using a more effective marketing campaigns, but also because they would always be ready to supply the demand, they would never have an excessive inventory taking valuable space in the warehouse, and at the same time they would not run out of the products that the customers are looking for.

Considering there is no tool like that yet, the closer that the organizations can get to it is by analyzing the consumer behavior. There are a lot of things that the organizations can get from analyzing them, like if a customer is close to change providers of certain services, if they will try a new brand of products, or what the customers meal would be depending on the weather.

Customers decisions are affected by many factors, and their emotions are a very important one. The way a person feels depends on many variables, and no one can control some of those variables, a clear example of this would be the weather. In the work presented by Kato (2021), the authors described how the weather affects the consumers mood and their perspective towards certain products, which at the end affects their purchase behavior.

The weather could be a very strong ally for the organizations, but no one can control it, therefore it can become a struggle for a variety of reasons. However, these struggles are not limited to the consumers change in mood but also including logistic limitations. Nonetheless, the technological advances of the last years have seen an improvement in the capabilities of collecting, processing, and analyzing large amounts of data to find patterns in the customers. One recent successful example of an actual application of these capabilities was shown in the article published by Hershey (2016), in which the author describes the relationship between weather and consumer behavior. The author also exposes the case of the dynamic menus implemented by McDonald's in Canada, which recommended certain products depending on the weather. After the testing period of the menus ended it was reported that the customers in the stores where the dynamic menus were implemented tended to expend more money than the clients from stores that did not have these dynamic menus.

The focus in forecasting consumer behavior is currently focused on the online to offline market, since it is easier to gather data from the specific individuals and in more detail. Some examples of the kind of data they could gather are the products they looked at, how much time they spent looking at a certain product before buying it, or before not buying it, etc. However, a market that has not been analyzed as deeply is the offline to offline market, like physical retail stores, because it is harder to gather data as detailed as in online to offline. The retail stores have started to gather relevant data from their returning customers by using systems of loyalty points, card systems, memberships, etc. They have also started to present special offers to the customers participating in these systems based on their preferences, nonetheless, it is still hard to approach the customers that are not participating in these systems.

In this project, we analyzed and forecasted the consumer behavior in the state of Alabama, United States, focusing on the sales of alcohol in general. We focused on alcohol because it was the easiest type of product to group, and it is a type of product consumed in most of the world. In this forecasting we ignored the physical retailer of preference and took into consideration the weather conditions presented in the state. The techniques applied were Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX), Gated Recurrent Unit (GRU), Long-Short Term Memory (LSTM) and ensemble techniques combining the forecasted results of the three individual models.

The main aim of this research was to understand: How well can a model using time series and ensemble techniques forecast the consumer behavior?

2 Related Work

2.1 Consumer Behavior Prediction Using Machine Learning

Anshu et al. (2021) implemented multiple techniques to analyze and predict the effects of effective online marketing in the customers and establish which technique performed the best. Among the techniques tested were Decision Tree (DT), Logistic Regression (LR), Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). The results obtained by the authors established that the RF outperformed all the other models tested in terms of accuracy, classification error, F-measure, precision, and recall. In a similar fashion, Zhang and Wang (2021) proposed an improved deep forest (DF) model to predict the consumers behavior related to re-purchase of various products from the Chinese e-commerce market. The improvements proposed by the authors consisted of adding two layers of extreme randomized trees and two additional layers of Extreme Gradient Boosting (XGBoost) in order to improve the cascade layer of the model. The results obtained in this research indicated that the improved version of DF proposed by the authors could outperform traditional methods, which were also tested in this research, and even a deep learning model. Both works treated the prediction of consumer behavior as a classification problem, dividing the results in the customers that took action and those who did not.

2.2 Consumer Behavior Prediction Using Deep Learning

The task of getting ahead of the future consumers demands has proven to be a complex and everchanging one, for this reason many approaches have attempted to use the latest technology available to accomplish the task. Lang and Rettenmeier (2017) made use of a Recurrent Neural Network (RNN) instead of LR, which had been applied to predict the probability of a customer to purchase a product from an e-commerce store and categorize those who would purchase it and those who would not. The implemented RNN had a simple architecture, with the addition of a Long-Short Term Memory (LSTM) layer, and ten-dimensional cell states with the hidden state at the last time step which was combined with a binary non history feature to perform the final prediction in a logistic layer. The obtained results showed that the RNN gave equally accurate or even more accurate predictions than the LR, adding the advantages of less preprocessing needed and less training time. Following with the neural networks implementation, Tian et al. (2022) presented a Deep Neural Network (DNN) that was improved by using a Rectified Linear Unit (ReLU) activation function to solve the gradient disappearance problem and by selecting the appropriate proportion of positive and negative samples in the training data. The obtained model slightly outperformed the traditional DNN and only took one third of the time to train. These research works explored the possibility to implement deep learning models to predict consumer behavior and showed that these techniques could outperform more traditional machine learning techniques.

2.3 Forecast of Consumer Behavior Using Time Series

Another tested approach for the prediction of consumer behavior has been using time series data. Salehinejad and Rahnamayan (2016) applied three variations of RNN to predict how recently a customer made a purchase, how often the customer made purchases, and how much the customer spent each time by tracking each of them with an ID. The results obtained in this research showed that the variation of RNN using the ReLU activation function outperformed the other variations tested, which included a LSTM and simple RNN. With a similar objective in mind, Abbasimehr and Shabani (2020) aimed to predict the same behaviors as the previous authors but applying Autoregressive Integrated Moving Average (ARIMA) instead of a RNN or a variation of it. The authors did the forecasts by two approaches, the first one consisted in an aggregated forecast in which they took all the customers into consideration for the forecasts. The second approach consisted of a segmented wise forecasting, by performing clustering in the data. The obtained results showed that by segmenting the data using a Segment-Wise-Customer-Wise approach the forecasts had better results. Even though these papers focused on the same problem, it is hard to approximate which one performs the best forecasts, since they were done and tested with different datasets. Like the last reviewed paper, Abbasimehr and Baghery (2022) clustered three different datasets containing customers information and attempted to predict their behavior by applying time series analysis and a variety of machine learning techniques like ARIMA, K-Nearest Neighbor (KNN), support vector regression (SVR), and a combination of SVR and grasshopper algorithm (SVRGOA). The results obtained by the authors showed the importance of the selection of suitable features in the datasets and that there was not a single best model that could outperform the others in every scenario, since the models outperformed each other depending on the dataset.

Liu et al. (2022) analyzed the relationship between online food ordering and the weather conditions in order to attempt to forecast the orders of different food categories. The experiment was performed using three different viewpoints in the dataset which were all data, seasonal data, and weekly data. The datasets were clustered in food categories, there was an ARIMA model build for each of the food categories. The authors chose to implement ARIMA since it has been extensively used in studies where the influence of the factors is being analyzed. The study concluded that depending on the weather conditions some categories got more affected than others in different ways, some were requested more, and others were requested less. The authors also found that the preferences of the consumers also changed with the different seasons of the year. Meanwhile, Omar et al. (2023) proposed the employment of four machine learning techniques to create a model that could be able to improve the forecasting of the inventory needs and fulfillment performance of a business network. The base techniques were autoregressive integrated moving average with exogenous variables (ARIMAX), linear regression model (LRM) with one lag of sales, polynomial regression model (PRM) with one lag of sales, and a machine learning model based on light gradient boosting machine. The results obtained by the authors indicated that the joint forecasting of the sales in online and offline stores provided a more accurate forecast and a better inventory performance. These papers highlighted the continuous application of ARIMA and ARIMAX in analyzing and creating forecasting models to predict the consumer behavior in different sectors and be more prepared to supply the market demands.

2.4 Time series forecasts in other fields

Zhai et al. (2020) presented a model that merged the capabilities of extracting hidden relationships of multiple control variables from the XGBoost and the gating unit from the gated recurrent unit (GRU) to extract the timing information, the authors named this model XGB-GRU. The main objective of this research was to improve the forecast of the temperature in the heating furnace used in industrial processes based on sixty-one characteristic variables. The results obtained showed that the proposed model outperformed the base models, giving more accurate forecasts. In the work presented by Zheng and Chen (2021), the authors implemented a deep learning model based in a GRU to improve the forecast of demand and production of energy power. The novelty of the model was that they implemented it with a selective update strategy and with a proposed stochastic optimization that the authors called adaptive mixed gradient, giving as a result a model that the author called GRU-SSU-AMG. The results obtained by this model outperformed the traditional GRU model in the five tested datasets.
Yamak et al. (2019) performed a comparison of three techniques to forecast the bitcoin price. The data set obtained by the authors consisted in the daily exchange rate of bitcoin going from 28 November of 2014 to 5 June 2019. The compared techniques were ARIMA,

GRU, and LSTM, and the metrics to compare the performance of the models were the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The results obtained showed that the best performance was obtained by the ARIMA model, followed by the GRU but the authors recognized that to improve the forecast of the models it would be necessary to take into account external factors that also affects the price of the cryptocurrency, like county policies and social media. Sharmin et al. (2022) implemented an ARIMA model and a stacked LSTM to accurately forecast the crime rate in London using the available data about crimes committed in that city. The implemented ARIMA consisted of a classic auto-ARIMA and the stacked LSTM contained multiple LSTM layers stacked to improve the forecasts. The results obtained by the authors showed that both models could accurately forecast the crime rate from 2014 to 2019 but if the analyzed time got extended to 2020 instead of 2019 the ARIMA models was the only one that could capture the downfall in the crime rate due to the pandemic.

2.5 Ensemble methodologies

Maaliw et al. (2021) implemented an ensemble machine learning model to forecast the number of Covid-19 cases. This ensemble model consisted of a dynamically weighted average between the forecasts of the two base models, an ARIMA model and a stacked LSTM. The obtained results confirmed that by taking into consideration both models forecast it was possible to achieve more accurate forecasts. Even though this research was focused on a different topic than consumer behavior it presents an interesting methodology to apply using time series and ensemble techniques for this project.

3 Dataset Description

For this project we retrieved the sales of alcohol in the state of Alabama from public data made available by the US Department of Agriculture (USDA). Then we obtained the historic weather data of Alabama from the National Oceanic and Atmospheric Administration (NOAA). Finally we merged both datasets, making sure that the dates matched from both datasets.

3.1 Sales Data From the USDA

This dataset contains information of retail sales in 43 states of the US including Alabama, Arizona, Colorado, Florida, Massachusetts, New Mexico, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, etc. It includes a weekly summary of retail sales, by category of products, volume and revenue [1]. The dataset records start on the week of October 6, 2019 and has been updated until January 26, 2023.
The columns of the original dataset are listed and described in Table 1.

3.2 Weather Data From the NOAA

Since the weather data has to be requested online it is possible to only gather the desired columns instead of the whole information available in each station [2]. The columns that

[1] https://www.ers.usda.gov/data-products/weekly-retail-food-sales/
[2] https://www.weather.gov/help-past-weather

Table 1: Sales dataset description

Column name	Description
Date	Day that the week ended
State	State of the sales
Category	product category
Dollars	Total value of sales in US dollars
Unit sales	Total units sold of the product in any presentation / size
Volume sales	Total volume sales in standardized units
Dollars last year	Total value of sales in US dollars in the same week of last year
Unit sales last year	Total units sold of the product in any presentation / size in the same week of last year
Volume sales last year	Total volume sales in standardized units in the same week of last year
Dollars 3 years ago	Total value of sales in US dollars in the same week 3 years ago starting in October 3, 2021
Unit sales 3 years ago	Total units sold of the product in any presentation / size in the same week of 3 years ago starting October 3, 2021
Volume sales 3 years ago	Total volume sales in standardized units in the same week 3 years ago starting October 3, 2021
Percent change dollars 1 year	Percentual change in the value of the sales compared to last year
Percent change units 1 year	Percentual change in the units sold compared to last year
Percent change volume 1 year	Percentual change in the volume sold compared to last year
Percent change dollars 3 years	Percentual change in the value of the sales compared to 3 years ago starting October 3, 2021
Percent change units 3 years	Percentual change in the units sold compared to 3 years ago starting October 3, 2021
Percent change volume 3 years	Percentual change in the volume sold compared to 3 years ago starting October 3, 2021

Table 2: Weather dataset description

Column name	Description
PRCP	How much rain fell in the day
SNWD	Snow depth of the day in centimeters
SNOW	Snow fall of the day
TAVG	Average temperature registered of the day in Celsius
TMAX	Maximum registered temperature of the day in Celsius
TMIN	Minimum registered temperature of the day in Celsius

we requested and their descriptions can be found in Table 2.

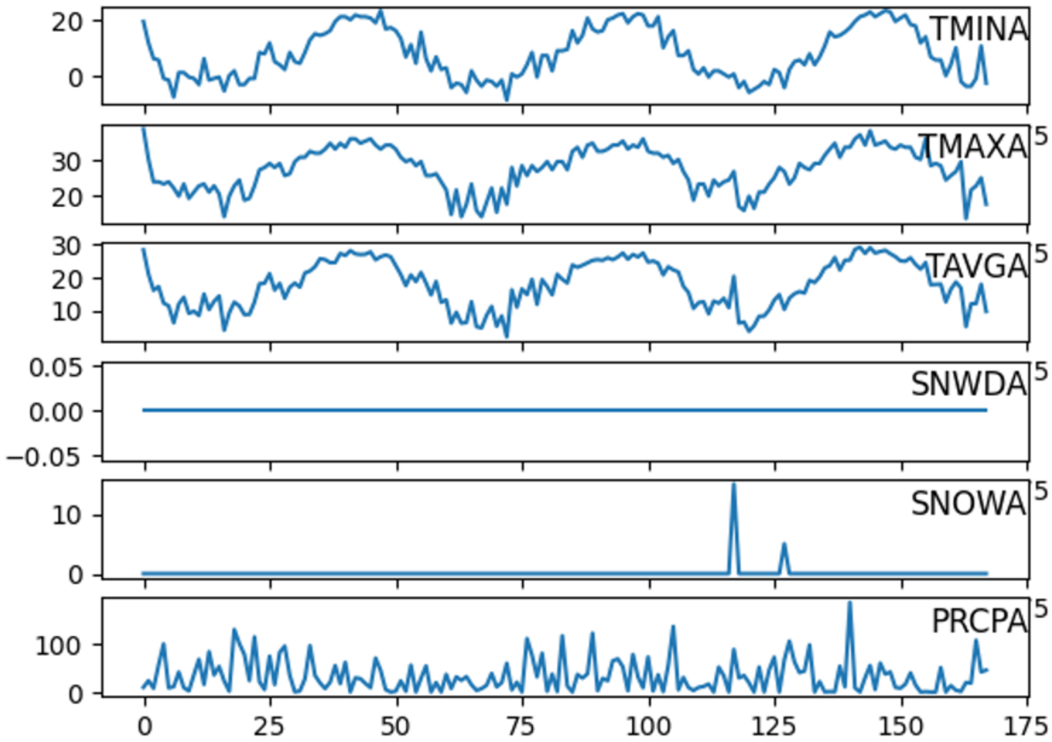

Figure 1: Weather conditions in Alabama

As seen in figure 1, the snow depth and snow fall variables are mostly 0 and would not help in the creation of the models, therefore we dropped them.

3.3 Merged Dataset

Before we started the preprocessing of the final dataset, we had to merge both datasets. The first challenge to achieve this was to convert the weather data into a weekly format instead of daily format to match the sales data. For this purpose, we took the lowest temperature shown on the week and selected is as the TMIN of the weekly dataset. Then we took the maximum temperature registered in the week and placed it as the TMAX for the weekly dataset. For the average temperature we calculated the overall average of the average temperatures in each day and took it as the new TAVG. Finally, for the

Table 3: Merged dataset description

Column name	Description
Date	Day that the week ended
Unit sales	Total units sold of the product in any presentation / size
PRCP	How much rain fell in the day
TAVG	Average temperature registered of the day in Celsius
TMAX	Maximum registered temperature of the day in Celsius
TMIN	Minimum registered temperature of the day in Celsius

remaining variables, we took the weekly totals of each of them and used it for the weekly dataset.
The resulting dataset consists of the columns described in Table 3.

4 Preprocessing of the Data

We performed two separated preprocesses in the dataset, the first one was done to fit the data into the two deep learning models implemented, LSTM and GRU. The second preprocess was exclusively done for the ARIMAX, since the dataset shape requirements were different for the deep learning models than for the ARIMAX.
The first steps were the same for both preprocesses, they consisted of replacing the index of the dataset with the dates and deleting the unnecessary columns. After this we just gave the appropriate format to the remaining columns, by converting them into numeric values.
The first step that we performed differently for both preprocesses was changing the shapes of the dataset used for the deep learning models. The reshaping consisted in adding the batch size, making the shape of the dataset three-dimensional. We choose to use a batch size of 6. This change of shape was unnecessary for the ARIMAX.
Then we defined the distributions of training, validation, and testing data. The deep learning models benefit from having validation data during to improve the training of the model, meanwhile the ARIMAX does not. For this reason the ARIMAX only had training and testing data.
The testing data was the same for both separations, the last ten weeks presented in the dataset going from 16 of October 2022 to 18 of December 2022. The validation data for the deep learning models consisted of the ten weeks prior to the testing data, going from 7 of August 2022 to 9 of September 2022. Finally, the training data differs for both separations, the training data used in the deep learning models goes from the first week recorded in the data set, which starts in the 6 of October 2019 until the 31 of July 2022, meanwhile the training data for the ARIMAX model goes from the 6 of October 2019 until the 9 of October 2022.
Due to the magnitude of the variables in the dataset differed greatly, it was necessary to standardize the magnitudes of all the variables in the training, validation, and testing data. The standardization consisted in obtaining the mean and standard deviation of each variable and then subtract the mean from each value and have it divided by the standard deviation. Additionally, we performed the standardization separately for each variable and we only used the training data to obtain the mean and standard deviation

from each variable. For this reason, the standardized values for the deep learning models and the ARIMAX are different.

5 Experiment setup

In this section we explain each component that took part in this research, starting with a short description of the techniques applied and following with the structures of the implemented models.

5.1 Long Short-Term Memory

The LSTM consist of a variation of RNN which are capable of learning long term dependencies, especially in problems related to sequence predictions, making it a favorite deep learning method to analyze and forecast time series problems [3]. The LSTM architecture consists of three gates, which are the forget gate that determines if the information from the previous timestep is relevant and should be kept or not, the input gate that learns new information from the current timestep, and the output gate that gives the updated information to the next timestep.

5.2 Gated Recurrent Unit

The GRU could be consider a variation of the LSTM, therefore, also a variation of the RNN, and has also been implemented in problems involving long sequences of data, giving equally good or better results than the LSTM using less computational power as mentioned by Dey and Salem (2017). It was designed to solve the vanishing gradient problem by employing an update gate and a reset gate that determines the information that passed to the output and are capable of keeping information from long ago or remove information that is irrelevant to the predictions.

5.3 Autoregressive Integrated Moving Average with Explanatory Variable

The ARIMAX consists of a statistical model used for analyzing and forecasting future values in time series data [4]. In the base ARIMA model the time series is made stationary by differentiating it, meanwhile the autoregressive and moving average components are used to model the time series in an equation but is not capable of taking other variables into consideration. The ARIMAX model adds external or exogenous variables that are able to affect the original time series. The model is formed by three components which are p that refers to the order of the autoregressive, d that represent the degree of differencing to make the time series stationary and the q that refers to the moving averages order.

5.4 Ensemble techniques

Ensemble techniques combines multiple models to improve the prediction performance, generalization and robustness, by combining the results of multiple models the overall

[3] Extracted from: `https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/`

[4] Extracted from: `https://365datascience.com/tutorials/python-tutorials/arimax/`

performance could be enhanced compared to the individual models. Ensemble techniques can be effective because they can improve the model generalization by reducing overfitting, capturing different patterns in the data and compensating the individual models weaknesses. Even though there are many forms of performing ensemble techniques, we choose to apply the voting method, in which the models are train separately and the results of the individual models gets averaged to get the final result.

5.5 Model Building

To improve the reproducibility of the project, we settled the TensorFlow random seed to be 30, which was the value that gave the best result among the tested seeds.
The LSTM model that we built consists of an input layer configured to receive a shape of 6 by 5, two LSTM layers, the first one with 64 nodes and the second one with 32, this combination was the best among the configurations that we tested, and it was necessary to include two layers since having only one was not enough to get good results. The number of nodes were also kept low to avoid overfitting, since the amount of data available was not particularly big to train a complex model with a high number of nodes. Finally, we added two dense layers, the first one has a Rectified Linear Unit (ReLU) activation function to help the neural network to learn and approximate complex relationships in the data, the second dense layer have a Linear activation function since the final objective of the model is to predict a linear value, in this case the standardized value of the units sold. To avoid overfitting the model ran 50 epochs and the best run based in the Root Mean Squared Error (RMSE) was saved as the final model.
The GRU model configuration is very similar to the LSTM, having an input layer to receive the same 6 by 5 shape, and the final two dense layers at the end. This model has one GRU layer with 140 nodes and a ReLU activation function, followed by a flatten layer to do the transition from the GRU layer to the dense layers. To avoid overfitting the model ran 50 epochs and the best run based in the RMSE was saved as the final model.

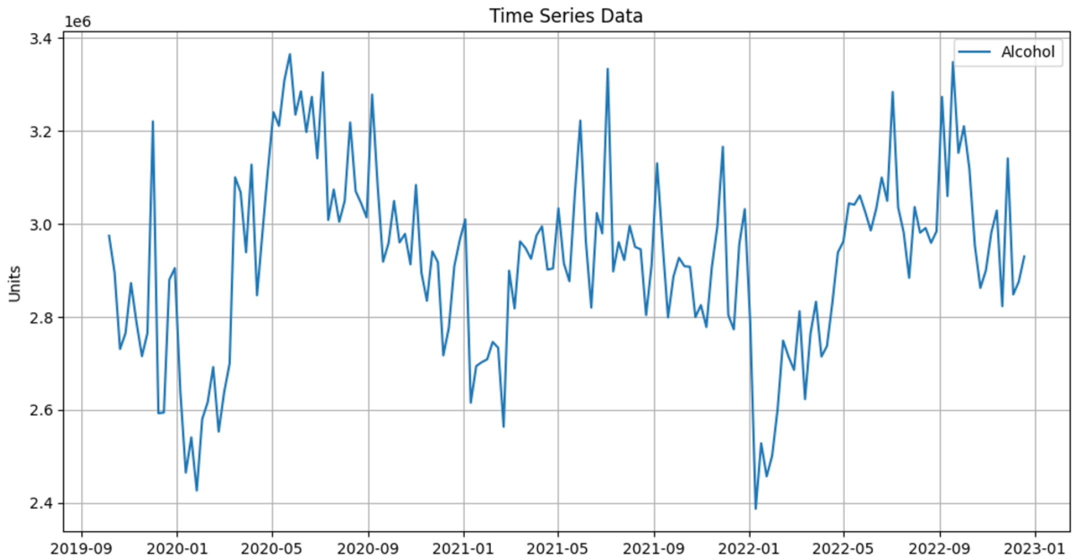

Figure 2: Units of alcohol sold in Alabama

We set the p, d, and q values for the ARIMAX model to be 1, 0, and 1 respectively. The d is 0 since, as seen in Figure 2, the data is stationary, therefore there is no need to derivate. The p and q were determined by checking the plots presented in Figure 3. Looking at the Partial Autocorrelation plot, the first lag is far above the limit, therefore the value of p was equal to 1. Even though the next three lags are also above the limit, they are very close to it, that is why we could ignore them. By looking at the Autocorrelation plot, the first lag is particularly above the limit and then no other lag reaches it, therefore we set the q value to 1.

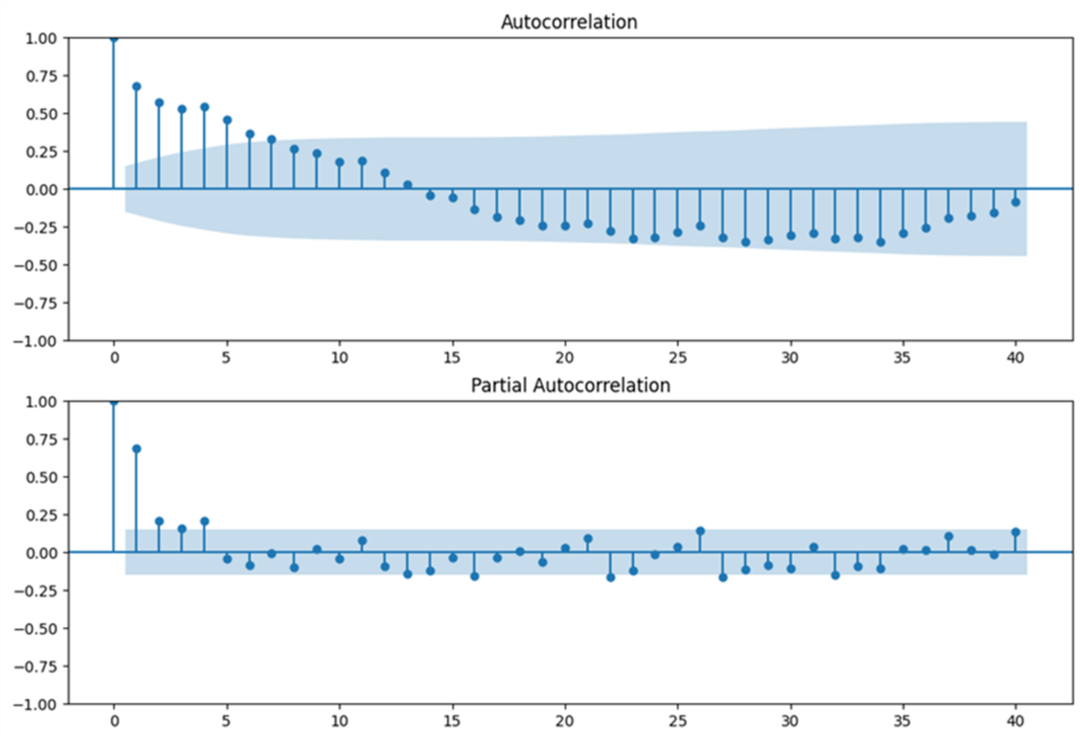

Figure 3: Autocorrelation and Partial Autocorrelation of the data

The forecasted values of the models previously described are in a standardized format. Therefore, it is necessary to bring them back to the original format in order to be able to do a meaningful interpretation of the results.
Then we started with the ensemble combinations. We tested three combinations using the ARIMAX as the common model, since among the three individual models, ARIMAX was the only one to have a negative bias that could compensate the positive biases obtained by the deep learning models. The first combination that we tested combined the results of the ARIMAX and the LSTM. The second combination used the results of the ARIMAX and the GRU. The final combination makes use of the forecasted values of the three individual models. Since the ensemble combinations used the postprocessed values there was no need to perform any postprocessing to the results of the ensemble models.

Table 4: Performance of the Models

Model	MAPE	MBE	RMSE
LSTM	2.78%	12,474.75	112,626.37
GRU	3.59%	3,881.2	129,691.1
ARIMAX	3.04%	-963.53	109,236.15
Ensemble 1	2.32%	5,755.61	96,823.07
Ensemble 2	2.74%	1,458.07	107,471.12
Ensemble 3	2.74%	5,130.8	104,837.28

6 Evaluation

The metrics that we used for the evaluation of the models were the Mean Absolute Percentage Error (MAPE), the Mean Bias Error (MBE), and the Root Mean Squared Error (RMSE), which are the most commonly used key performance indicators when we evaluate a forecast.
The MAPE compares the forecasted values against the actual values and gives a percentage of error of the forecast [5]. MAPE calculates the absolute percentage error for each of the data points and then calculate the average of the percentage error to give an overall percentage of the forecasts error. Since the percentage given by the MAPE represents the error in the forecast, the lower the error the better the model.
The MBE calculates the difference between the forecasted values and the actual values without taking into consideration the direction of the error, positive (overestimating) or negative (underestimating) [6]. The MBE gives an overall overview of the model tendency to overestimate or underestimate the actual values, if the result is positive it means that the model tends to overestimate and if it is negative it means that it is underestimating. The closer the MBE is to 0 it means that, in average, the predictions are unbiased and more accurate.
The RMSE gives the average magnitude of the errors between the forecasted values and the actual values [7]. The RMSE calculate the square root of the squared difference between the forecasted values and the actual values. The RMSE penalizes more larger errors than smaller ones, thanks to the squaring operation, therefore it tends to stress the impact of outstanding errors. The lower the RMSE the better are the obtained results, since it would indicate that the forecasted values are closer to the actual values.
The table 4 shows the resultant metrics of the models, even though the results in the MBE and RMSE seems very high, it is important to emphasize that the forecasted values are in the millions.

The Ensemble 1, which combines the results of the ARIMAX and the LSTM, obtained the best results among the Ensemble combinations and obtained the best overall MAPE and RMSE, which indicates that the average of the forecasted values was the closest to the actual values among the tested methods, and that the magnitude of the errors was the lower among the tested techniques.
Nonetheless, the MBE of the ARIMAX is still better than the MBE of the Ensemble 1.

[5]Extracted from: `https://www.statisticshowto.com/mean-absolute-percentage-error-mape/`
[6]Extracted from: `https://agrimetsoft.com/calculators/Mean%20Bias%20Error`
[7]Extracted from: `https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/`

This is due to the MBE of the LSTM being much bigger than the one of the ARIMAX, therefore the underestimation of the ARIMAX is not able to fully compensate the overestimation of the LSTM.

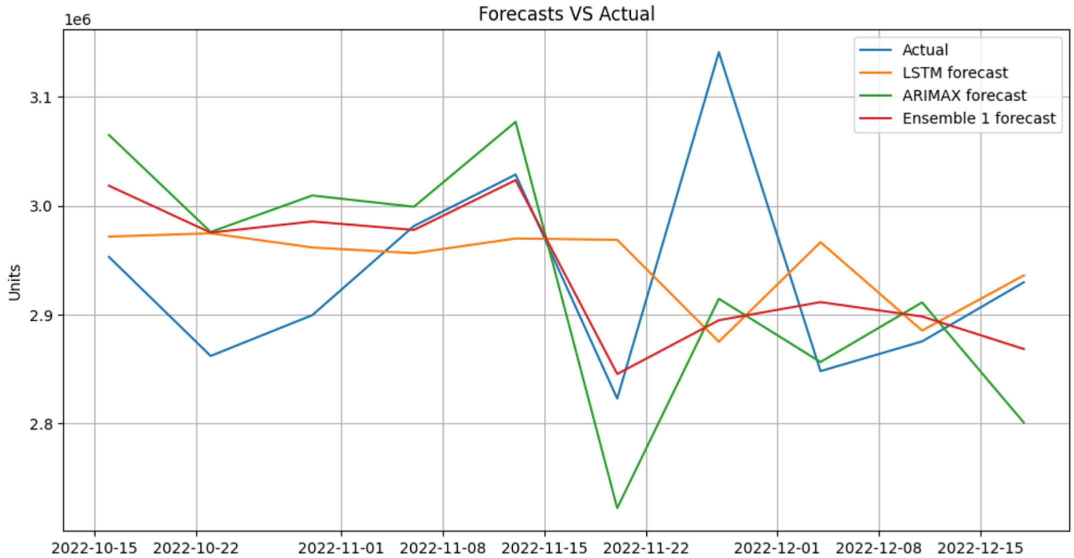

Figure 4: Forecasted values vs Actual values

As seen in Figure 4, the ensemble technique was able to improve the forecast of the two base models, ARIMAX and LSTM, and it shows that the ensemble model was capable of almost catching up to the actual values. However, the ensemble model was not capable of following all the sudden changes presented by the historical data, since it depends on the forecast of the base models, who were not able to follow up with the sudden changes either.

7 Conclusion and Future Work

In this project, we explored the possibility to accurately forecast the sales of alcohol from the retailers in the state of Alabama, US, taking into consideration the weather conditions, and using ensemble techniques. Even though not all the ensemble combinations tested in this project provided better results than every base model, the possibility of implementing ensemble techniques to improve the overall forecasting capabilities and the reductions of errors was successfully proven with a positive result.

For future work, we would be interested in re-training all the models with more data as it is being released, in order to avoid the limitation presented by the relatively little amount of data available to train the deep learning models. Another possibility for future work could be trying different combinations of ensemble models with the upcoming deep learning techniques that are yet to come.

References

Abbasimehr, H. and Baghery, F. S. (2022). A novel time series clustering method with fine-tuned support vector regression for customer behavior analysis, *Expert Systems with Applications* **204**: 117584.

Abbasimehr, H. and Shabani, M. (2020). Forecasting of customer behavior using time series analysis, *Data Science: From Research to Application*, Springer, pp. 188–201.

Anshu, K., Singh, S. K. and Kumari, R. (2021). A machine learning model for effective consumer behaviour prediction, *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, IEEE, pp. 1–5.

Dey, R. and Salem, F. M. (2017). Gate-variants of gated recurrent unit (gru) neural networks, *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, IEEE, pp. 1597–1600.

Glen, S. (n.d.a). Mean absolute percentage error (mape), `https://www.statisticshowto.com/mean-absolute-percentage-error-mape/`.

Glen, S. (n.d.b). Rmse: Root mean square error, `https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/`.

Hershey, L. (2016). How businesses are influencing the relationship between the weather and consumer demand, *The Winnower* **3**: e146521.

Kato, T. (2021). Impact of weather conditions by time zone on soft drink purchases across offline and online channels considering the effect of tv advertising, *2021 5th International Conference on E-Business and Internet*, pp. 1–6.

Lang, T. and Rettenmeier, M. (2017). Understanding consumer behavior with recurrent neural networks, *Workshop on Machine Learning Methods for Recommender Systems*.

Liu, D., Wang, W. and Zhao, Y. (2022). Effect of weather on online food ordering, *Kybernetes* **51**(1): 165–209.

Maaliw, R. R., Ballera, M. A., Mabunga, Z. P., Mahusay, A. T., Dejelo, D. A. and Seño, M. P. (2021). An ensemble machine learning approach for time series forecasting of covid-19 cases, *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, pp. 0633–0640.

Omar, H., Klibi, W., Babai, M. Z. and Ducq, Y. (2023). Basket data-driven approach for omnichannel demand forecasting, *International Journal of Production Economics* **257**: 108748.

Padhma, M. (2022). End-to-end introduction to evaluating regression models, `https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/`.

Salehinejad, H. and Rahnamayan, S. (2016). Customer shopping pattern prediction: A recurrent neural network approach, *2016 IEEE symposium series on computational intelligence (SSCI)*, IEEE, pp. 1–6.

Saxena, S. (2023). Learn about long short-term memory (lstm) algorithms, `https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/`.

Sharmin, S., Alam, F. I., Das, A. and Uddin, R. (2022). An investigation into crime forecast using auto arima and stacked lstm, *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, IEEE, pp. 415–420.

Sidhharth, M. (2021). Basic understanding of time series modelling with auto arimax, `https://www.analyticsvidhya.com/blog/2021/11/basic-understanding-of-time-series-modelling-with-auto-arimax/`.

Smarten, M. (2018). What is arimax forecasting and how is it used for enterprise analysis?, `https://www.elegantjbi.com/blog/what-is-arimax-forecasting-and-how-is-it-used-for-enterprise-analysis.htm`.

Tian, Y., Lai, Y. and Yang, C. (2022). Research of consumption behavior prediction based on improved dnn, *Scientific Programming* **2022**.

Yamak, P. T., Yujian, L. and Gadosey, P. K. (2019). A comparison between arima, lstm, and gru for time series forecasting, *Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence*, pp. 49–55.

Zhai, N., Yao, P. and Zhou, X. (2020). Multivariate time series forecast in industrial process based on xgboost and gru, *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Vol. 9, IEEE, pp. 1397–1400.

Zhang, W. and Wang, M. (2021). An improved deep forest model for prediction of e-commerce consumers' repurchase behavior, *Plos one* **16**(9): e0255906.

Zheng, W. and Chen, G. (2021). An accurate gru-based power time-series prediction approach with selective state updating and stochastic optimization, *IEEE Transactions on Cybernetics* **52**(12): 13902–13914.