

TIME  
SERIES  
ANALYSIS

**Final Project Technical Report**

**Submitted by (GROUP #9)**

* VENKATA THEJASWI MULLAPUDI
* NANDITH MALYADHEESH KURRA
* VIKAS KUMAR VEJENDLA
* SUSENDRANATH REDDY MUSANI

**Submitted for DSCI-6001**

**(FALL 2023)**

**DR. ARDIANA SULA**

**CONTENTS**

**1. Introduction…………………………………………………....3**

**2. Executive Summary…………………………………………...3**

**3. Background Theory…………………………………………...4**

**4. Methodology………………………………………………...…4**

**5. Exploratory Data Analysis………………….…………….…..8**

**6. Results Section…………………………………………….…...10**

**7. Application Deployment………………..……………………..11**

1. **Conclusion ……………………………………………….……12**
2. **Future Improvements………………………………………....13**

**10. References…………………………………….……………….13**

### LIST OF ABBREVIATIONS

1. ARIMA – Auto regressive integrated moving average
2. LSTM – Long short-term memory
3. SARIMA – Seasonal auto regressive integrated moving average
4. ACF/PCAF- Auto correlation function/ partial auto correlation function

### ABSTRACT

Sales is a crucial part of any business. It can make or break a business. Without a powerful forecasting tool and right management, it is impossible be competent in the industry. One can plan for stock management, inventory and advertisement of products at the right time. With advancement in technology and data being so readily available in large amounts, it is very expensive or a complex job to get the useful information out of it by previous methods.

Machine learning is needed to get the forecasts with more accuracy and with less work. Patterns to be detected is left upto algorithms. This is useful to stand your ground in industry now with less cost. Time consumption is minimised also as training is done by the machine and no manual effort is needed.

Again if one wants to start a new business, sales forecasting might be the most crucial aspect of success ahead. Moreover loans are a must in most business or capital is needed to buy all the machines, inventory they need to get off the ground. One will get nothing with a bright thought and a lot of enthusiasm when you have to go to the bank. Trust is vital to get huge amounts of capital and this is visible if a lot of investors are there. A business plan is all one needs. Thus, our project uses ARIMA model for store item forecasting and trains the model on it for give store item sales.

**1. INTRODUCTION**

This technical report presents the findings of a research project into the use of machine learning algorithms for time series analysis. The algorithms used in the process are Z-Score, Isolation models, (EMA), Mainly ARIMA and SARIMA. The dataset used for the research was the holiday\_events from the store sales. The results of the study showed that SVC achieved the highest accuracy of 77% followed by Logistic Regression (76%), Random Forest (75%) and KNN (70%). The implications of these results are discussed, and possible future research directions are suggested.

**2. EXECUTIVE SUMMARY**

This project focuses on predicting the time series analysis and provides a comprehensive understanding of store item demands, enabling informed decision-making for inventory management. By leveraging these insights, the store can enhance operational efficiency, reduce costs, and ultimately improve customer satisfaction through optimized product availability. Regular updates and continuous monitoring are essential to adapt to changing market dynamics and maintain a competitive edge.

**3. BACKGROUND THEORY**

Time series analysis is a growing globally for the predictions. It is one of the most leading fundamental concepts and methodologies used to analyze and interpret temporal data patterns, and its prevalence is increasing every day. Machine learning algorithms and the concepts we use in the background are components of time series, stationary, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) , Forecasting Models like ARIMA, SARIMA., Outlier Detection and Correlation Analysis Understanding these theoretical foundations is crucial for effectively applying time series analysis techniques to store item demands, enabling accurate forecasting and informed decision-making in inventory management.

**4. METHODOLOGY**

**4.1.Data collection**

The dataset used for the research was the store Items demanding Forcasting time series from train.csv Machine Learning Repository. This dataset contains 38 to 40 versions, there are more than 314 datasets .

This dataset contains 314 sets, and here in this we use train.csv which contains more than 50 variable , for the analysis mainly we are using 4 datasets from the store data which are:

1. DATA
2. ITEM
3. STORE
4. SALES

**4.2. Decomposing Time Series**

**The Seasonal component:** A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known frequency. A time series can contain multiple superimposed seasonal periods.

**The Trend component:** A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes a trend is referred to as “changing direction” when it might go from an increasing trend to a decreasing trend.

**The Cyclical component:** The cyclical component represents phenomena that happen across seasonal periods. Cyclical patterns do not have a fixed period like seasonal patterns do. The cyclical component is hard to isolate and it's often ‘left alone’ by combining it with the trend component.

**The Noise component:** The noise or the random component is what remains behind when you separate out seasonality and trend from the time series. Noise is the effect of factors that you do not know, or which you cannot measure. It is the effect of the known unknowns, or the unknown unknowns.

**4.2.1 Analysing Seasonality**

To identify additive or multiplicative model for decomposition -  
There are basically two methods to analyze the seasonality of a Time Series: additive and multiplicative

The **additive model** is Y[t] = T[t] + S[t] + e[t]

The **multiplicative model** is Y[t] = T[t] \* S[t] \* e[t]

**4.3 Checking Stationarity**

Before applying any statistical model on a Time Series, the series has to be stationary or time invariant, which means that, over different time periods, it should have constant means, constant variance and constant covariance. It means that the data should have constant mean throughout, scattered consistently and should have same frequency throughout.

* The mean of the series should not be a function of time.
* The variance of the series should not be a function of time. This property is known as homoscedasticity.
* Finally, the covariance of the i th term and the (i + m) th term should not be a function of time.

Here we are going to check the stationarity using 2 methods:

1. **Moving Mean:** Plot the moving average or moving standard deviation to see if it varies with time.
2. **ADCF Test — Augmented Dickey–Fuller test:** This is used to gives us various values that can help in identifying stationarity. The Null hypothesis says that a Time-series is non-stationary.

**4.3 Prediction using SARIMA Model**

Take the last 30 days in training set as validation data

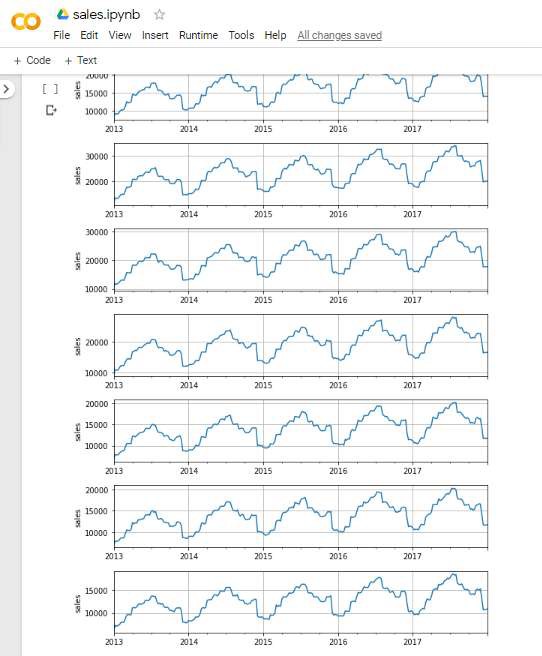
We can see that this model is better than simple ARIMA model.

**def** smape\_kun(y\_true, y\_pred):

mape **=** np.mean(abs((y\_true**-**y\_pred)**/**y\_true))**\***100

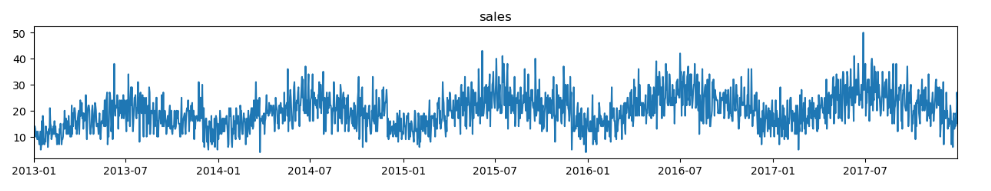
smape **=** np.mean((np.abs(y\_pred **-** y\_true) **\*** 200**/** (np.abs(y\_pred) **+** np.abs(y\_true))).fillna(0))

print('MAPE: %.2f %% \nSMAPE: %.2f'**%** (mape,smape), "%")

**5. EXPLORATORY DATA ANALYSIS**

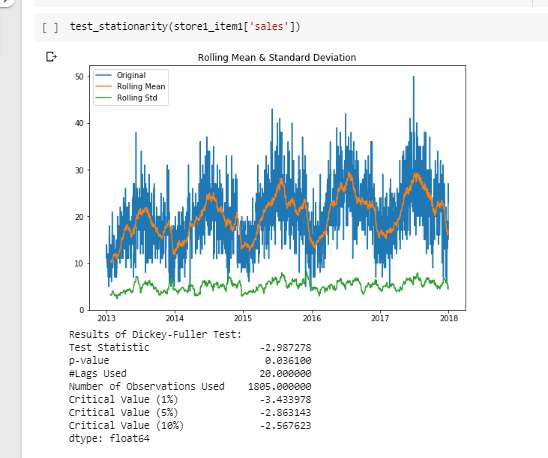
##### Plot sales for 10 stores wrt year

By looking at the graphs we can say that all stores show same seasonality and trend. So, we make our predictions taking 1 store and 1 item and then apply to all 10 stores.



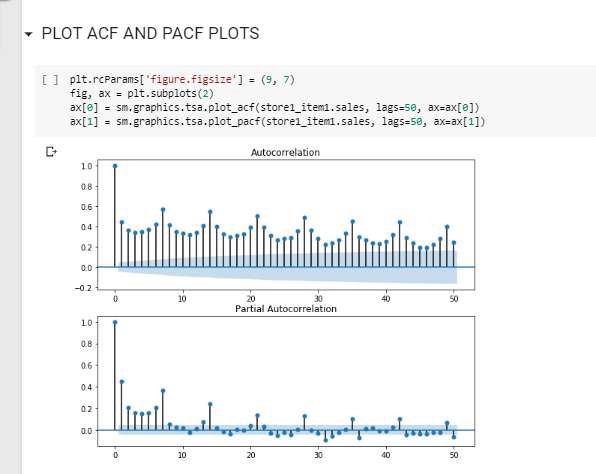
### Plot of monthly and yearly sales

##### **RESULT**



From the p value and critical values we observe that the sales data given to us is not stationary. It should have constant mean as well as variance to fit in the model.

which is brute force and checks on few models by random to decide which fits the best.



**5. RESULTS**

Sales is the crucial part of any business. A business financial plan needs a powerful tool for predicting sales rise or fall. This not only decides the how successful the business will be but it could help to take decisions as to which product to drop or promote. Past data of product sales could be put to use and decide future sales making the business more lively. Seasonality and trend are two important aspect of sales and with right tools in hand, one can not only up the sales but save money also. Dropping some products is an important decision and with machine learning techniques, the business can bloom to its fullest. Products with upcoming season can be managed and stored beforehand. This will save panic buying or sudden rise in price. Sales forecasting is therefore a must with your business. In this study, ARIMA(0,1,1) & SARIMA(0,1,0,12) were the best candidate model for predicting next 3 months of store items demand using last 5 years of store data taken from the website kaggle. This shows the power of ARIMA model to accurately predict the sales for small period of time. Although it has its limits but values predicted had very less error.

**BENEFITS:**

###### **Sales Planning:** Without any forecasting tools, a business is doomed. One will not be to check whether the products or schemes they are deploying are working in the favour of customers or not.

###### **Demand Forecasting:** With increasing competition in business, efficient sales forecast methods are a must. Moreover with machine learning methods, one is at advantage in relative to others. This the best tool to manage and check for loopholes in your business.

###### **Higher OTIF Delivery:** With correct sales foretelling, you’ll be able to gain an excellent OTIF or delivery. Once demands rise arrangement of inventory must be maintained maximizing the customer satisfaction. Moreover, customers will have lesser complaints and manager could focus on other plans as all prediction is left to machine learning tools.

###### **Inventory Controls:** Stable inventory is vital for a successful business. Efficient forecasts of products demands over time can help avoid insufficient stock. Everything can be made according to plan which is desired. In case of rise in demands in season, one can arrange inventory beforehand and maximize the money earned with these stools.

###### **Supply Chain Management:** You get a more detailed demand chart for various products, which leads to higher management over your provide chain. This paves path to more opportunities to supply items in time and manage export import as well.

###### **Financial Planning:** It gives you more chances of better revenue and profit as you have already anticipated the hurdles as well as the aware of the predictions of various values leading to overall getter profit.

###### **Internal Controls:** Internal controls can be managed more efficiently as business have their future predictions. After predicting future sales one will be able to create selections concerning hiring – permanent or temporary – promoting and growth.

###### **Continuous Improvement:** With machine learning techniques can be used to predict sales with more accuracy as statistical model produce proper results only for small time span. Business performance will improve as accuracy is increased at a low labour with machine learning.

###### **Price Stability:** One can predict demands rise and downfall and arrange inventory according to it. This can help avoid the panic of price instability. Sales can be handled by time series forecasting more accurately.

###### **Marketing:** Sales forecasting helps in getting the most profit out of products. This could further provide the opportunity to promote products timely as popularity of some products could be increased. It can also give an insight to which products are not worth investing anymore.

**6.1. Classification report**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
|  |  |  |  |  |
| 0 | 0.41 | 0.39 | 0.40 | 38 |
|  |  |  |  |  |
| 1 | 0.78 | 0.99 | 0.87 | 137 |
|  |  |  |  |  |
| accuracy |  |  | 0.78 | 175 |
|  |  |  |  |  |
| Macro avg | 0.39 | 0.50 | 0.44 | 175 |
|  |  |  |  |  |

**Table 6.1.** Classification report for SVC

**7. APPLICATION DEPLOYMENT**

The demo application has been deployed using Gradio at the url: http://127.0.0.1:7860 or https://6f3139f20cb107cdd7.gradio.live

The code is hosted on Github:

<https://github.com/Tnvsproject/Time-Series-Analysis-on-Store-Item-Demand-Forecasting>

**8. CONCLUSION**

Sales is the crucial part of any business. A business financial plan needs a powerful tool for predicting sales rise or fall. This not only decides the how successful the business will be but it could help to take decisions as to which product to drop or promote. Past data of product sales could be put to use and decide future sales making the business more lively. Seasonality and trend are two important aspect of sales and with right tools in hand, one can not only up the sales but save money also. Dropping some products is an important decision and with machine learning techniques, the business can bloom to its fullest. Products with upcoming season can be managed and stored beforehand. This will save panic buying or sudden rise in price. Sales forecasting is therefore a must with your business. In this study, ARIMA(0,1,1) & SARIMA(0,1,0,12) were the best candidate model for predicting next 3 months of store items demand using last 5 years of store data taken from the website kaggle. This shows the power of ARIMA model to accurately predict the sales for small period of time. Although it has its limits but values predicted had very less error.

**9. REFERENCES**

[1] Adebiyi, Ayodele & Adewumi, Aderemi & Ayo, Charles. (2014). Stock price prediction using the ARIMA model. Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014. 10.1109/UKSim.2014.67.

[2] Rangsan Nochai , Titida Nochai. (2006). ARIMA MODEL FOR FORECASTING OIL PALM PRICE. Proceedings – 2 nd IMT\_GT Regioonal Conference on Mathematics, Statistics and Applications University Sains Malaysia, Penang, June 13-15,2006.

[3]Wei-Chiang Hong, “Rainfall forecasting by technological machine learning models”, Applied Mathematics and Computation ( Volume: 200 , Issue: 1 , June 2008 ).

[4] “Store Item Demand Forecasting Challenge”, kaggle.com, June 2018. [Online].Available:https://www.kaggle.com/c/demand-forecasting-kernelsonly/overview.

[5] Susan Li, “An End-to-End Project on Time Series Analysis and Forecasting with Python”, towardsdatascience.com, July 2018. [Online]. Available: https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-andforecasting-with-python-4835e6bf050b?gi=b708004bef61.