

FULL STACK DATA ENGINEERING PROJECT FOR BUSINESS PROCESS

Jeevankarthick B

Kowshik R

Bannari Amman Institute of Technology, Sathyamangalam

ABSTRACT

Designing an integrated sales forecasting application and predict the future sales of a business or a company. A web application will be created using Angular framework. A login page will be created for entering the user's credentials and then navigating into the dashboard where the user will be provided with the inputs. The inputs contain the upload button to upload the dataset and then the input box for selecting the target feature and entering the frequency of years / months to which the sales need to be predicted further from the current one. Application contains the features like Data extraction, data cleaning, pre-processing, and Training the model, prediction and visualization. These features have been carried out by using machine learning and the algorithm used here is SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with exogenous factors). The Angular acts as front-end, Machine learning (Python) as back-end and Flask- API acts as middle-ware to integrate or connect front-end and back-end. Finally, the prediction results will be shown in the interactive dashboard using Power BI.

Keywords:

SARIMAX, PowerBI, Flask-API

INTRODUCTION

The Sales forecasting is used to estimate a company's or an organization's revenue for a specific time frame in the future. It primarily establishes how many products have been sold by the business and how long it has taken the business to market those products. Businesses that have forecasting results that are 10% more accurate than the preceding year's help them generate sales revenue that is twice as high. These studies demonstrate the value of sales promotion and how it contributes to business success. Since it aids in business decision-making, sales forecasting is an important indicator.

Additionally, it is useful for planning the resources based on utilisation. Company or business owners will take decisions for the procurement of the raw materials , and hiring the persons or the staffs and dealing with the marketing on the desired goods. The procedure of sales validation is essential to the development and growth of a company.

OBJECTIVES

The main objective of the study is to identify the challenges in the implementation of the advanced time series algorithms for real time data processing and to improve the accuracy and error less metrics of the real time data and to provide knowledgeable interactive dashboards to the user

METHODOLOGY

These features have been carried out by using machine learning and the algorithm used here is sarimax (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors). Flask-api is used to integrate or link the front-end and back-end, with Angular acting as the front-end and machine learning as the back-end. Finally, using powerbi and the super-market dataset, the prediction findings will be displayed in an interactive dashboard.

Angular provides the modules and the components. A module is a group of interconnected components. Furthermore, it provides the full HTML page's application functionality. Directives, controllers, and services, among other components, are contained within modules. We use angular modules to handle the login and dashboard components for our application. The login screen will have a form for users to fill out with their username and password. The username and password will be checked. The page's structure and the validation portion are provided in HTML files. The structure of the ts file is made using FormControls, validators, and regular expressions.

Once both inputs have been submitted, a function that is triggered by the button's 'click' event will permit the submit button and direct the user to the dashboard page. On the dashboard screen is a form with input fields for the dataset file and time frame input, such as the amount of months or years, as well as an upload button. Once both entries have been made,

After pressing the submit button, apis will be called to post the dataset, collect the input data, and launch the Python prediction application. The view model button will also display the ML program's success metrics. API stands for application programming interface. We'll create a Python REST API using the Flask framework. It functions as a compact framework for building web apps similar to Angular. Since no specific frameworks or tools are required, it is known as a microframework. Because they are already accessible, third-party libraries aid in providing standardized functionality. You can use endpoints to contact specific methods. The '/upload file' and '/get input' commands were created using the post method to obtain the dataset and convert it to a dataframe. The '/post input' and '/get input' commands were created using the get method to post and get input. Various HTTP tricks and vulgar behavior include, GET POST is a method for obtaining a property.

Model: The ARIMA family of models includes a subset known as SARIMAX. ARIMA models are composed of the moving-average term (MA) and the autoregressive term (AR). (MA). The former views a value as the weighted sum of earlier values at any particular time. The value of the second model is the same as the weighted sum of the past residuals.

There is also an integrated word to distinguish the time series. (I). This vanilla version's inability to manage seasonality is a serious flaw.

$$\text{SARIMA}(p,d,q) \times (P,D,Q,S)$$

P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = duration of seasonal pattern. p = non-seasonal AR order, d = non-seasonal differencing, and q = non-seasonal MA order.

We first make sure the info we are using is reliable. The early plots make it clear that the data is not stationary because of the strong tendency and seasonality. But we can be more scientific in our assumption by using a statistical measure, the Augmented Dickey-Fuller test, which is also included in the statsmodels package. Strong evidence exists to disprove the null hypothesis, which is H_0 , if $p < 0.05$. There is weak proof against the null hypothesis if the time series is stationary because it has a unit root, which suggests that it is non-stationary. Data has no units root and is stationary.

The model loaded from the pickle file will be given the number of weeks, months, and years, and it will predict future sales for that period of time. For getting, visualising, and uploading using the Flask API, the data will be concatenated.

PowerBI turns sources into cohesive visualisations and insights. Data would be gathered using the api, and charts like bar charts or line charts would be used to plot the predicted data with drilling up/down giving it more information. To display the profit or sales in each area, etc., some pie charts are used. The info will also be updated with each refresh because an API is being used.

RESULTS AND DISCUSSION

User login in which user can enter into the prediction web app. If the user is new, he needs to create an account in signup page. When he enters his details, those data will be sent to backend and stored in Mongo dB database server. After all these processes he will be redirect to data entry page.



Output Chart

This figure shows the final output of prediction from our ML model used. It is shown in graphical representation for better understanding.



ACKNOWLEDGEMENT

We wish to express our sincere thanks to our faculty guide Dr. SUNDARAM S, Assistant Professor, Department of Physics for his constructive ideas, inspirations, encouragement, excellent guidance and much needed technical support extended to complete our project work. We would like to enunciate heartfelt thanks to our esteemed Bannari Amman Institute of Technology, Sathyamangalam Chairman Dr. BALASUBRAMANIAN S. V, and the respected trustee Dr. VIJAYAKUMAR M. P, for providing excellent facilities and support during the course of study in this institute.

CONCLUSION

SARIMAX algorithm is ease of use has made it a very good base model where your time series are easily split into simple time components. However, if your signal is noisy, fine-tuning the model's performance can be problematic.

Thanks to the native breakdown of the forecast into time components, the model's decisions are easy to interpret. However, it is not accurate enough to be used to measure the impact of an external event, because you never know if the effect of the event is already accounted for in the native time components. If your time series follow certain business cycles, you can achieve very decent performance quickly, without intensive feature engineering. However, the trend component is not always well estimated, which can cause significant performance drifts. Because of this, this model may require close monitoring and frequent human intervention.

Hence the population forecasting is done successfully and displayed to user. And our objectives are all completed. SARIMAX is definitely a good choice for making quick and accurate predictions. It has intuitive parameters that can be tuned by someone who has good domain knowledge but lacks technical skills in predictive models. Readers can also try and fit Sarimax directly to hourly data and discuss in the comments if they are able to get a better result.

REFERENCES

- [1] Fildes, Robert & Bretschneider, Stuart & Collopy, Fred & Lawrence, Michael & Stewart, Doug & Winklhofer, Heidi & Mentzer, John & Moon, Mark. (2003). Researching Sales Forecasting Practice. International Journal of Forecasting - int j forecasting. 19. 27- 42. 10.1016/S0169-2070(02)00033-X.
- [2] Hariyani, Mr. (2020). An Effective Approach for Sales Forecasting. International Journal for Research in Applied Science and Engineering Technology. 8. 2109- 2025.10.22214/ijrasat.2020.5329.
- [3] Gahirwal, Manisha. (2013). Inter Time Series Sales Forecasting.
- [4] Beheshti-Kashi, Samaneh & Karimi, Hamid & Thoben, Klaus-Dieter & Lütjen, Michael & Teucke, Michael. (2015). A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering: An Open Access Journal. 3. 154-161. 10.1080/21642583.2014.999389.
- [5] Hamel, Ayat & Ismael, Baydaa. (2022). Time series Forecasting Using ARIMA model.
- [6] Arunraj, Nari & Ahrens, Diane & Fernandes, Michael. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. International Journal of Operations Research and Information Systems. 7. 1-21. 10.4018/IJORIS.2016040101.
- [7] Armstrong, J. (1999).Sales Forecasting. SSRN Electronic Journal. 10.2139/ssrn.1164602.

IJETRM

International Journal of Engineering Technology Research & Management