Product Sales Forecasting using Quantitative Methods

Abinmon Bousally
Department of Computer Application
Amal Jyothi College of Engineering Kanjirappally, India
abinmonbousally@mca.ajce.in

Miss Navyamol K
Department of Computer Application.
Amal Jyothi College of Engineering Kanjirappally, India
navyamol@amaljyothi.ac.in

Abstract- As the impact of the Internet on people's lives grows, e-commerce platforms are developing more quickly, with both their user bases and revenue growth. Strong government policy support in recent years has also helped to create a favourable climate for the growth of the e-commerce sector. The epidemic of this year has highlighted the e-commerce industry's rise in the national economy. In these conditions, e-commerce platforms and firms are becoming more numerous and competitive. A platform must be able to better satisfy user demands and perform admirably in all areas of coordination and administration if it hopes to preserve its competitive edge. Accurately predicting the sales volume of e-commerce platforms at this time is crucial. We are now looking on the prediction model that can be more useful in different situations despite the fact that there are already many studies on forecasting e-commerce sales. I choose to research and examine a number of quantitative forecasting methods for this paper's assessment of time series analysis and forecasting. This notebook discusses methods for forecasting product sales as well as in-depth justifications for the various metrics used to evaluate the projections.

Keywords: Machine Learning, Deep Learning, Regression, and Sales Prediction.

INTRODUCTION

Making a decision to acquire or sell any kind of asset is a highly challenging task today. Your choice may be affected by a variety of variables and complications, such as the ideal timing to acquire or sell certain things. In the financial market, buyers must be able to purchase goods at reasonable prices, and stockholders must know when to sell their shares. So, the problem of sales forecasting has emerged. Intuition was mostly used to predict sales for trading equities and commodities. As trading became more popular, people looked for strategies and instruments that could predict rates accurately, boosting profits and lowering risks. several techniques, including deep learning, machine learning, and regression.

The objective of this project was to anticipate the sales of the items included in the dataset received from Kaggle using a variety of quantitative techniques, such as Times Series Models and Causal Models.

The notebook features the following models:

- Seasonal Naive Model
- Holt-Winters Model (Triple Exponential Smoothing)
- ARIMA and Seasonal ARIMA Models;
- Linear Regression Model

REFERENCE

Compare the forecasts produced by linear, deep learning, and machine learning models. Using a set of sales data over the 1941-day period, we assessed two linear models, three machine learning models, and two deep learning models.

However, for predicting sales, complex machine learning and deep learning models do not outperform straightforward linear regression models for predicting sales. So, we try to find the best from linear models. So, we take another set of data and try to understand the variations in different models in linear regression. So, we take and specialized on trying 4 linear models and one new model: Seasonal Naive Model, Linear Regression Model, Holt-Winters Model (Triple Exponential Smoothing), ARIMA, and Seasonal ARIMA Models. We will select the ideal model from this set.

DATASET

The data set used in this paper is dataset contains historical sales records of 10 stores and 50 products, from the year 2013 through 2017. For the purpose of this project, we will only look at the sales of 'item' - 1 from 'store' - 1 The data set is divided into train and test sets, with the train set including sales records from January 2013 to September 2017 and the test set (validation set) including sales records from the final three months of 2017. For the sake of exploratory data analysis and causal modelling, certain new features derived from the date field have been developed.

LITERATURE REVIEW

Mohamed Ali Mohamed, Ibrahim Mahmoud El-Hanaway and Ahmad Salah, the proposed models are assessed using the following four metrics: MAE, RMSE, MAPE, and R2. The results demonstrate that the random forest model produces the best outcomes, followed by the ARIMA model with a narrow performance gap.

Anurag Bejju, Palani, Dubai campus, P.O Box 345055, Dubai, UAE, McCarthy's 4Ps are integrated into this model based on web mining to provide a comprehensive study of e-business strategies. In order to create more successful and productive judgments, managers can employ a structured and exact approach.

Karan deep Singh Booma P M and Umapathy Eaganathan By doing this project of using machine learning for forecasting the ecommerce sales, it was noticed that the in this project, there are many different methods of forecasting the sales of the ecommerce platform but the researcher was only able to focus on only four algorithms which are commonly being used when forecasting the sales of the future.

LIJUAN HUANG, ZIXIN DOU, YONGJUN HU, AND RAOYI HUANG This study not only adds to the body of knowledge by demonstrating the influence of sentiment topic distribution on sales forecast, but it also has practical ramifications for e-commerce practitioners by helping them better manage inventory and advertise using this prediction approach. TERMS Sentiment analysis, distribution of SCOR topics, and sales forecasting.

Madhuvanthi.K, Nallakaruppan. M.K, Senthilkumar N C, Siva Rama Krishnan S- Sales forecast is a current, widespread trend in which all business enterprises prosper. It also helps an organisation or concern identify its future objectives and the strategy it will use to reach those objectives. The information on car sales is gathered from a number of sources.

METHODOLOGY

In this study, we evaluated the effectiveness of four models.

- Seasonal Naive Model
- Holt-Winters Model (Triple Exponential Smoothing)
- ARIMA and Seasonal ARIMA Models;
- Linear Regression Model

Quantitative Methods to Forecast Product Sales

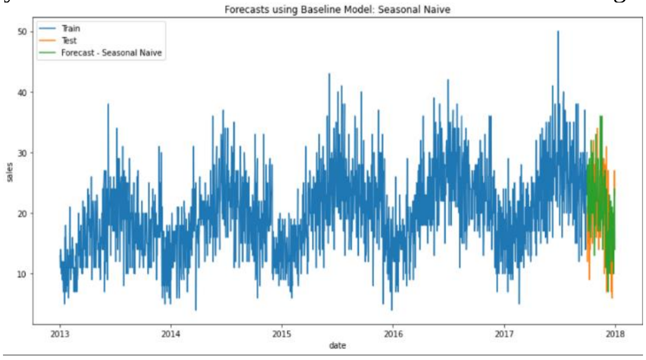
From the data exploration plots above, it can be reliably inferred that there is seasonality present in the product sales data as well as an overall rise in the volume of sales throughout the years. The linear trend and seasonality evident in the product sales will thus be taken into consideration when we anticipate the volume of sales for the final three months of 2017.

There are several approaches to tackle the forecasting problem. Either traditional time series models or causal models, such linear regression, can be used. We will investigate each of these methods and try to evaluate our projections using the validation set.

- **Seasonal Naive Model**

By comparing what occurred at the same time last year, it is a crude technique that considers seasonal tendencies. For instance, the seasonal naïve technique will assume the same number of sales for December 2017 as it did for December 2016 if we wish to anticipate the sales during that month. Fortunately, we have sales data going back at least a year; otherwise, this approach may not make sense. In the following code, the dates in the test data are subtracted by one year, and the resultant difference is increased by one day. Then, the resulting difference is searched up in the training data to return the sales for each of the resulting dates.

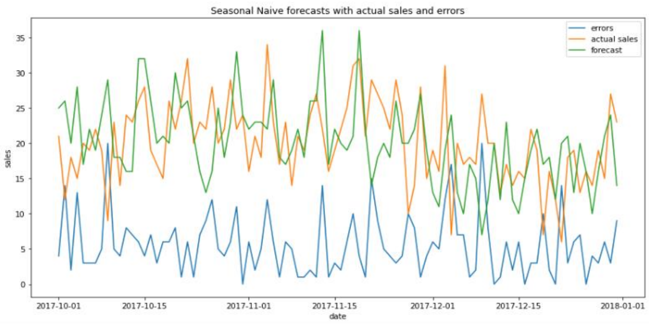
Taking into account seasonality, we now have our fundamental estimates based on the volume of sales from a year ago.



The predictions look good at first glance because our crude algorithm accurately depicts the declining trend. But we shall explicitly measure the performance using measures for forecast accuracy.

Evaluating the Forecasts

One of the widely recognised forecasting metrics is forecast error, which is the most frequent forecasting statistic. You can quickly calculate the error by identifying the difference between the actual and expected sales figures. For instance, you would have a -4 mistake if 14 things were projected to be sold but only 10 were. To evaluate the forecast as a whole, metrics like mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error will be looked at (MAPE).



total_sales	total_sn_pred_sales	overall_error	MAE	RMSE	MAPE

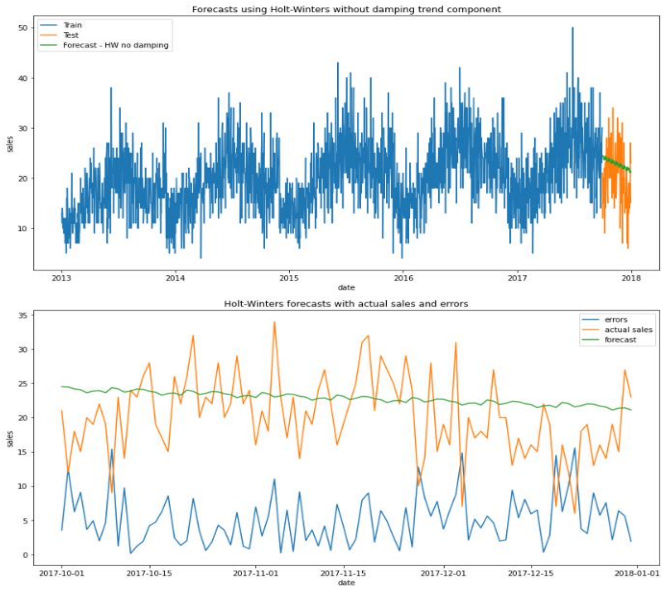
The overall error is not as bad, and we were able to achieve a MAPE of 27.8%. We will use this as a benchmark to judge the forecast performance of the other models.

- **Holt-Winters Model (Triple Exponential Smoothing)**

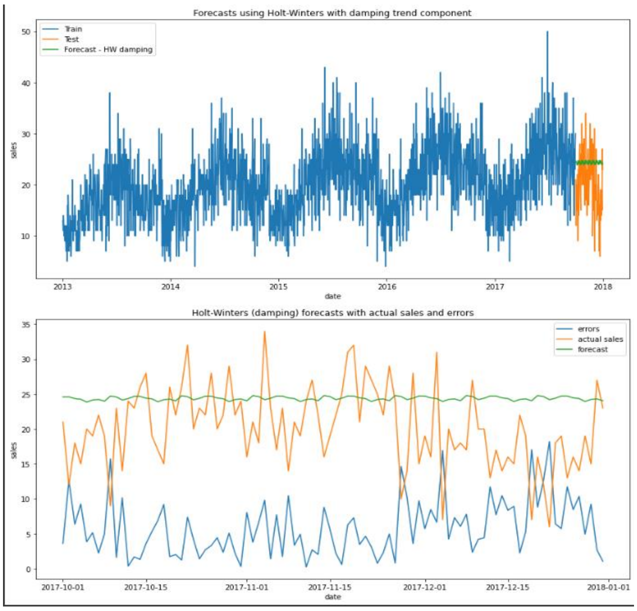
We may see a time series' seasonality, trend, and error/remainder terms using a time series decomposition diagram. The Exponential Smoothing models depend heavily on these three elements. As a result, choosing the kind of exponential smoothing model to utilise for our projections is assisted by the decomposition graphic.

In contrast to Single Moving Averages, which give earlier data equal weights, exponential smoothing gives older observations weights that go off exponentially. So, while creating forecasts, recent data are given a relative advantage

over historical data. But with exponential smoothing, there are one or more smoothing parameters that need to be chosen (or approximated), and these decisions impact the weights given to observations.
Three different smoothing methods are available. Smoothing with a single, double, or triple exponential.
When a trend and seasonality are present in the data, triple exponential smoothing is applied. A third equation is used to account for seasonality (sometimes called periodicity). The Holt-Winters (HW) technique is the name given to the resulting system of equations in honour of its creators.



Inference: The decreasing trend is clearly captured by the Holt-Winters method, and the MAPE 22.8% is better in comparison to our baseline model. Let's try Holt-Winters method with a damped parameter, and see if we can improve the results



Inference: To the naked eye, forecasts seem alright, however the MAPE 29.7% is worse than our baseline model.

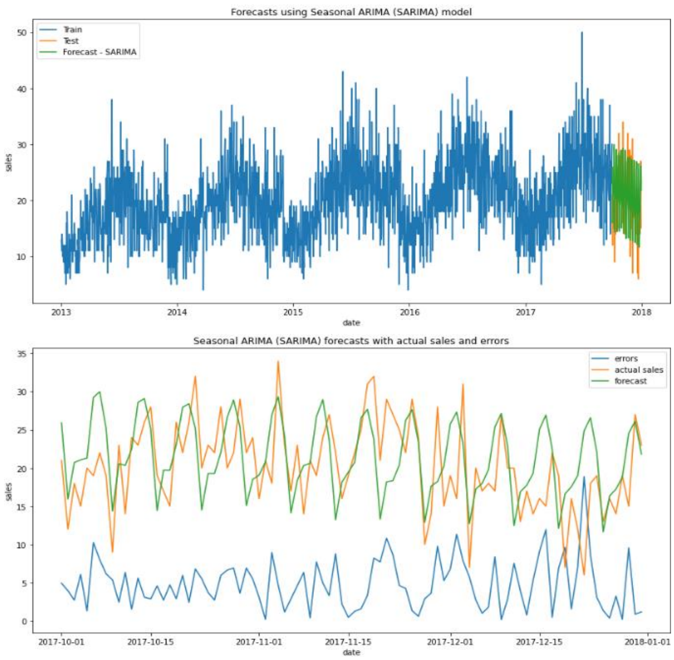
- **ARIMA and Seasonal ARIMA Models**

Step 1: Check stationarity
Our series has to be made stationary before we can continue with our analysis.
The quality of stationarity is the manifestation of steady statistical features (mean, variance, autocorrelation, etc.). A time-series is not stationary if its mean grows with time.
Methods to Check Stationarity:

Plotting rolling statistics: A first useful method for visually inspecting our series is to plot rolling means and variances. If the rolling statistics reveal a distinct trend (either upwards or downwards) and fluctuating variance (increasing or decreasing amplitude), you can draw the conclusion that the series is extremely unlikely to be stable.

Augmented Dickey-Fuller Test: If a time-series is stationary, it will pass this test. It produces a result known as a "test-statistic," on the basis of which you can determine the time-series' stationary or non-stationary status with varying degrees (or percentages) of confidence. The test statistic must be more negative (less negative) than the crucial value since it is expected to be negative in order for the hypothesis to be rejected and it to be determined that the series is stationary.

ACF and PACF plots: An autocorrelation (ACF) graphic demonstrates how a series' autocorrelation with respect to its own lag evolves over time. On a partial autocorrelation (PACF) plot, the correlation between a series and a lag of itself that cannot be explained by correlations at any lower-order lags is displayed. In a perfect world, the series should not be linked with any of its own lags. In terms of the illustration, we want each spike to land in the region of blue.



Inference: The SARIMA model with MAPE of 23.7% performed better than our baseline model, but couldn't better Holt-Winter's Triple Exponential Smoothing method.

- **<u>Linear Regression Model</u>**

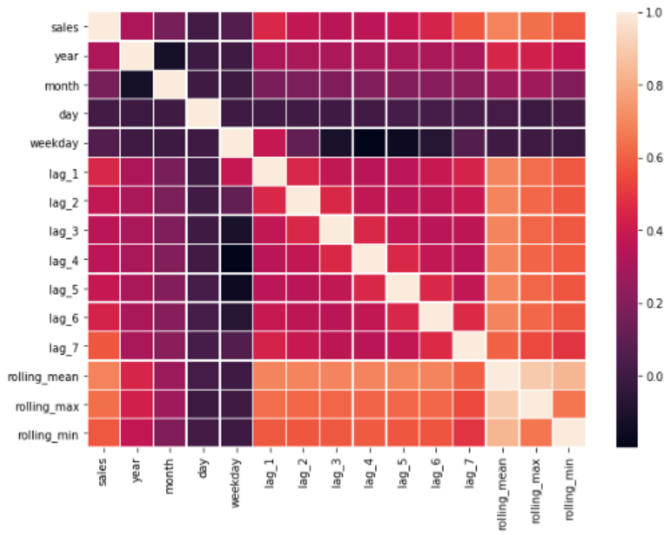
A linear regression model describes the relationship between a dependent variable (y) and one more independent variables (X). The term "response variable" also applies to the dependent variable. Explanatory or predictive variables and independent variables are both used interchangeably. Both categorical and continuous predictor variables also go by the name's covariates and factors. The term "design matrix" often refers to matrix X containing observations on predictor variables.

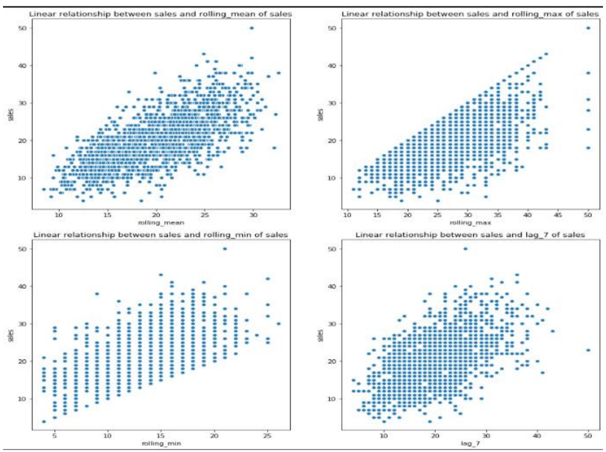
<u>Interpret Linear Regression Results</u>

The data into a table in step one.
Making a fitted model is step two.
Finding and eliminating outliers is step three.
Refine the model in step four.
Predict how fresh data will be received in step five.
Disseminate the model step six.

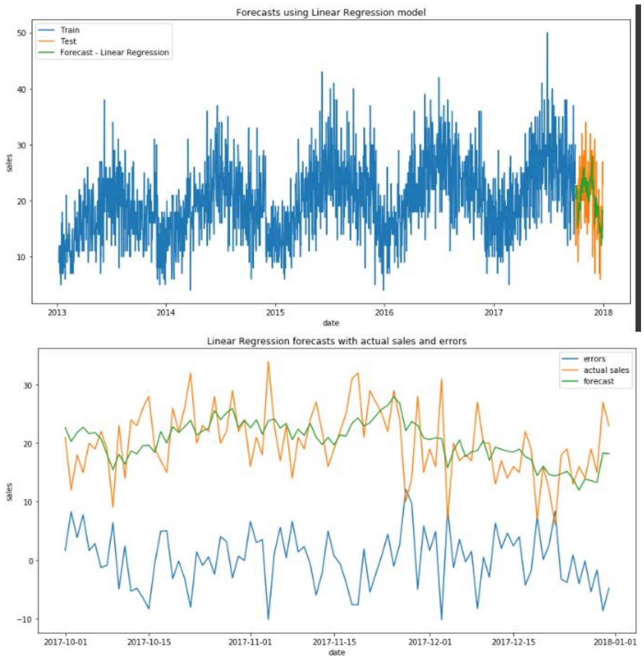
Let's apply Linear Regression to our time series data in order to forecasts sales

Feature Selection and Model Building





Model Evaluation and Predictions



The linear regression model accounts for both the upward and downward movement of the sales data. It outperformed the Holt-Winters model in terms of MAPE (19.07%), which had been performing the best up to this point.

RESULT

Naive Seasonal Model We were able to reach a MAPE of 27.8%, which shows that the overall error is not as terrible.

The Holt-Winters technique accurately depicts the declining trend, and the MAPE 22.8% performs better than our baseline model.

Although it outperformed our baseline model, the SARIMA model's MAPE of 23.7% was unable to surpass Holt-Triple Winter's Exponential Smoothing technique.

The upward and downward movement in the sales data is captured by the linear regression model. Its MAPE 19.07% performance beats that of the Holt-Winters model, which had previously been the best-performing model.

CONCLUSION

This paper offers a machine learning-based method for sales prediction.

We found that neither deep learning nor complicated machine learning models have any benefits after comparing the outcomes of several models. Overall, The best results have been obtained with simple linear regression models. In terms of a single time series, Triple Exponential Smoothing outperformed outperforming it on five different series. the LSTM model on two separate series.

Four linear models—the Seasonal Naive Model, the Holt-Winters Model (Triple Exponential Smoothing), the ARIMA and Seasonal ARIMA Models, and the Linear Regression Model—were the focus of our experimentation. For time series forecasting, we took into account various time-series models as well as a regression model. Our findings demonstrated that the linear regression model performed better than the other time-series models. Therefore, rather than using a time-series model to anticipate sales for this dataset, we may use a regression model. Regression models' core tenet is that previous patterns would repeat themselves in the future, and given our data was heavily seasonal and had a linear trend, it made obvious that the linear regression model would do better than the other models.

REFERENCES

(1) Sales Analysis of E-Commerce Websites using Data, Mining Techniques, Anurag Bejju, Department of Computer Science, Birla Institute of Technology & Science, Pilani, Dubai Campus, P.O.Box 345055, Dubai International Academic City, Dubai, UAE, volume 133 – No.5, January 2016

(2) E-Commerce System for Sale Prediction Using Machine Learning Technique, To cite this article: Karandeep Singh et al 2020 J. Phys.: Conf. Ser. 1712 012042

(3) Price Prediction of Seasonal Items Using Machine Learning and Statistical Methods
Faculty of Computers and Informatics, Zagazig University, Zagazig, Egypt *Corresponding Author: Ahmad Salah. Email: ahmad@zu.edu.eg, Received: 07 June 2021; Accepted: 08 July 2021

(4) Online Sales Prediction: An Analysis
Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong, Received May 5, 2019, accepted May 19, 2019, date of publication May 29, 2019, date of current version July 1, 2019.

(5) Sales Prediction based on Machine Learning
Zixuan Huo International School, Beijing University of Posts and Telecommunications, Beijing, China, 2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)