Supply Chain Analytics

Group 8

CONTENT

- Problem Statement
- Dataset
- Moving Average
- Exponential Smoothing
- Linear Regression
- Arima
- Machine learning
- LSTM
- Results
- Conclusion

PROBLEM STATEMENT

Daily data for female births of a California was selected.

The reason for selection of this dataset was that the prediction of number of births is extremely important for any government and most of its constituting bodies for medium- and long-term planning for infrastructural requirements like schools, teachers, utilities, commodities, etc.

In addition to the government, many organizations depend heavily on birth estimates especially companies like Babies-R-Us, Toys-R-Us, Diaper Pampers, Huggies, etc., Strollers and Crib manufacturing companies etc.

- Female births’ data was picked up for analysis and forecasting.
- We checked the predictability of data for female births and found that it has some relevant components and is predictable.
- The goal of this analysis was to forecast the number of births in a California for next year in future using one of the most optimum forecasting models applying the learnings of the Time Series Analysis and Forecasting Course and utilizing Jupyter notebook as the software tool.

DATASET

Daily Female Births Dataset

This dataset describes the number of daily female births in California in 1959.

The units are a count and there are 365 observations. The source of the dataset is credited to Newton (1988).

Below is a sample of the first few rows of

Date	Births
01-01-1959	35
01-02-1959	32
01-03-1959	30
01-04-1959	31
01-05-1959	44
01-06-1959	29
01-07-1959	45
01-08-1959	43
01-09-1959	38
01-10-1959	27
01-11-1959	38
01-12-1959	33

Moving Average

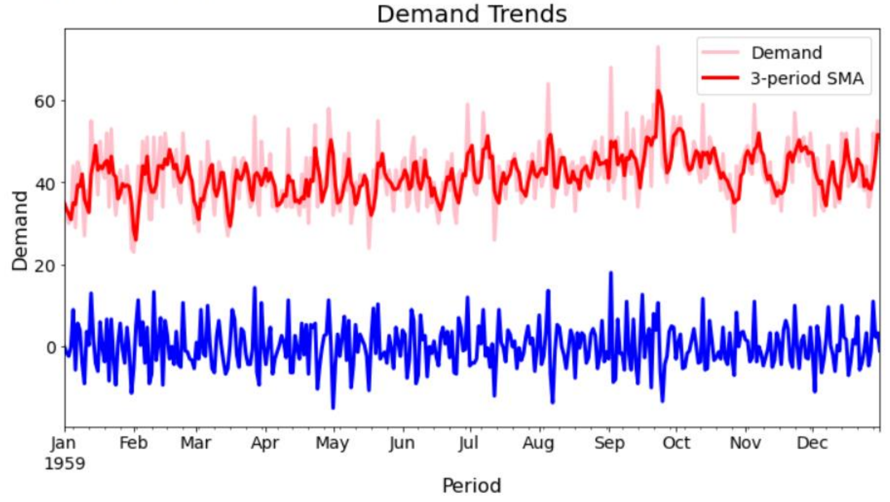
- MA is also known as rolling mean as it is calculated by averaging data of the time series within a certain time period.
- MA is also considered as simple forecasting model.
- Forecast is the average of the demand during the last n periods.
- We calculate MA by using below formula:

$$f_t = \frac{1}{n}\sum_{i=1}^{n} d_{t-i}$$

  - n is the number of periods we take the average of
  - d_t is the demand we observed during the periods t.
  - f_t is the forecast we made for period t

- As per our Dataset we have taken period 3.
- It has best values prediction for our dataset.
- Accuracy measures were found using accuracy() function for all these models as shown above.

Text(0, 0.5, 'Demand')




```
[45] # For KPI Calculation for period 3
     MAE = df["Error_3"].abs().mean()
     print("MAE:",round(MAE,2))
     RMSE = np.sqrt((df["Error_3"]**2).mean())
     print("RMSE:",round(RMSE,2))
     MSE= (df["Error_3"]**2).mean()
     print("MSE:",round(MSE,2))
     MAPE= np.mean(np.abs((df["Demand"] - df["SMA_3"])/df["Demand"]))*100
     print("MAPE:",round(MAPE,2))
     PA = 100-MAPE
     print("PA:",round(PA,2))
```

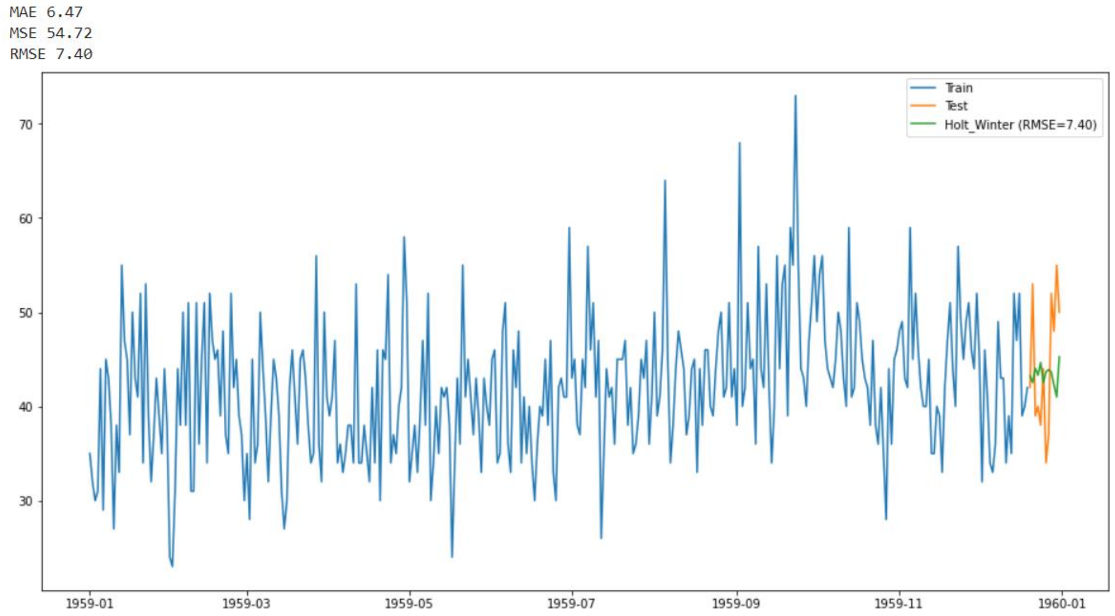
```
MAE: 4.35
RMSE: 5.42
MSE: 29.38
MAPE: 10.58
PA: 89.42
```

Exponential Smoothing

- Exponential Smoothing's methods are appropriate for non-stationary data.
- It’s forecasting methods are similar in that a prediction is a weighted sum of past observations, but the model explicitly uses an exponentially decreasing weight for past observations.
- There are 3 exponential smoothing methods
  - Simple Exponential Smoothing (SES)
  - Double Exponential Smoothing (DES) also called Holt method
  - Triple Exponential Smoothing (TES) also called Holt-Winters models
- Exponential Smoothing is also an extension of MA that explicitly handles trends, and the most advanced approach adds support for seasonality.

HOLT-WINTERS MODELS(TES)

- We have used Triple Exponential Smoothing method as our dataset has strong seasonality.
- Below are the values obtain from this model.
- RMSE is 7.40

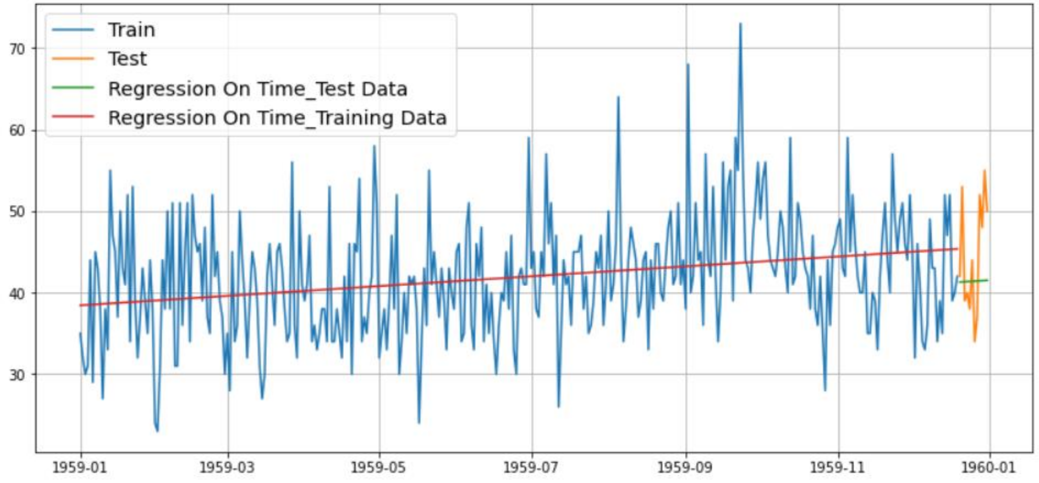


LINEAR REGRESSION

Linear programming is a simple technique to show complex relationships through linear functions to find the optimum solutions

LP is one of the simplest ways to perform optimization

Helps solving some very complex optimization problems by making a few simplifying assumptions



	Model	Test RMSE	Test MAPE	Test MAE	Test MSE
Model 1	RegressionOnTime	7.34	13.72	6.08	53.81

- For our dataset the value of RMSE is 7.34 and MAE is 6.08.

ARIMA

- ARIMA stands for Auto regressive integrated moving average model.
- AR model uses the dependent relationship between an observation and some number of lagged observations.
- Integrated is the use of differencing of raw observations in order to make the time series stationary.
- MA model uses the dependency between an observation and a residual error(white noise) from a moving average model applied to lagged observations.
- Each of these components are explicitly specified in the model as a parameter.
- AR and MA are two widely used linear models that work on stationary time series, and I is a preprocessing procedure to “make the time series stationary if needed.

Using Dickey fuller method stationarity was checked

It was made stationary.

The value obtain for RMSE and Mae model is 8.2 and 6.7 respectively for test models

Machine Learning

- Machine learning techniques allow predicting the amount of products/services to be purchased during a defined future period.
- ML learns patterns from the entire dataset while the statistical models treat each data point separately.
- Here We have converted a time series prediction problem (i.e. predict demand over time) into a supervised machine learning problem.
- We created Benchmark using linear Regression

BENCHMARK

We train our data and tested and split it to 75-25 ratio.

Created benchmark using Linear Regression.

The value obtain was:

- Tree on train set MAE%: 13.0
- Tree on test set MAE%: 11.8

Decision Tree

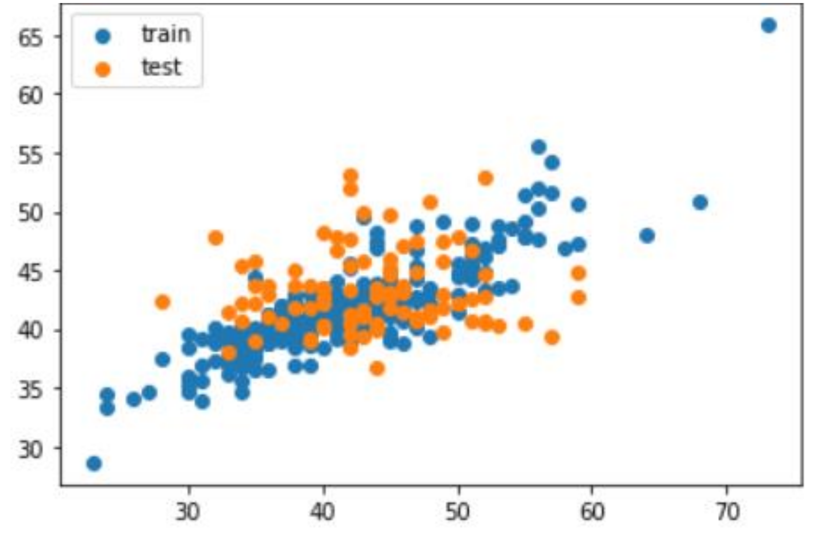
- Decision trees are a class of machine learning algorithms that will create a map (a tree actually) of questions to make a prediction.
- We call them Regression trees as we are predicting a number here.
- Prediction with this algorithm is not possible for our dataset as the value we obtain for our dataset are negative.
- The prediction values can be worst with negative values.

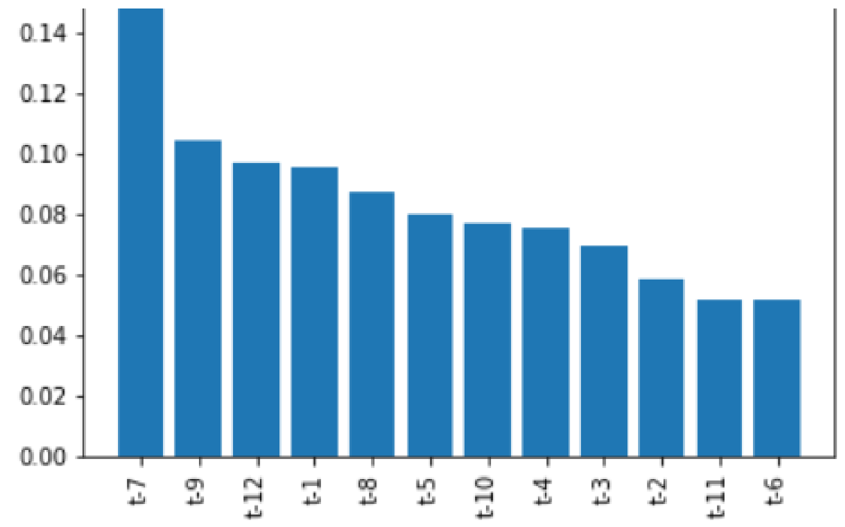
Random Forest Tree

- Random forests are a large number of trees, combined using averages at the end of the process.
- A single decision tree is a weak predictor, but is relatively fast to build.
- More trees give a more robust model and prevent overfitting.
- For our dataset the random forest tree value is negative so it’s not the best fit.

GRADIENT BOOSTED TREE MODEL

- Gradient boosting is a set of decision trees where tress are built successfully .
- It builds one tree at a time which is also called additive model.
- It combines its result on the way unlike random forest tree.
- For our dataset value predicted successfully and the values are MAE 8.7% for train and 12.5% for test.





LSTM(LONG SHORT-TERM MEMORY NETWORK)

- They are Types of RNN’s.
- the LSTM model reads one time step of the sequence at a time and builds up an internal state representation that can be used as a learned context for making a prediction.
- It solves many time series tasks unsolvable by feedforward networks using fixed size time windows.

	Model Name	RMSE
1	LSTM-model-2	6.69
2	LSTM-model-3	6.73
0	LSTM-model-1	6.73

- We have used 3 models with different epoch.
- Epoch used are
  - Model-1 epoch 50
  - Model-2 epoch 100
  - Model-3 epoch 80
- For model-2 with epoch 100 has best RMSE values 6.69

RESULTS

- The table shows the value of RMSE.
- The lower the value of the RMSE the best model.
- For our dataset Moving Average is the best model followed BY LSTM with 100 epoch

Model	RMSE
Moving Average	5.42
Exponential Smoothing(TES)	7.40
Linear Regression	7.34
ARIMA	8.72
LSTM	6.69

MAE VALUE FOR MODELS

- MAE value for Moving Average is less followed by linear regression
- For our dataset Moving Average is the best model

Model	MAE
Moving Average	4.35
TES	6.47
Linear Regression	6.08
ARIMA	6.77
Benchmark model	12.5
Gradient Boosted Tree	12.4

CONCLUSION

After utilizing all the models for forecasting it was concluded that moving average and LSTM models were the best model for forecasting with respect to smoothing, linear regression and Arima model.

So, it was felt that perhaps moving average should be used for forecasting the number of female births in the region.

The forecasting results were quite promising with MAE (Mean Absolute Error) being 4.35 and RMSE (Root Mean Square Error) as 5.42 respectively.

This project provided a great learning opportunity for the group because it presented many challenges in achieving the goals, and overcoming those challenges improved the group's understanding of the topic.

Initially, a dataset on house prices, airline passengers, and crude oil production was used, but after preliminary analysis, it was found that the data was not predictable and thus we used this dataset.

THANK YOU

Khushboo Patel

Ruchita Tamboli

Shubham Bhatnagar

Aayush Seth

Arshie Fatima