GREAT LEARNING

FINANCE AND RISK ANALYTICS

ASSIGNMENT-2

2021

GUNJAR FULEY - 9938126651

Credit Risk

Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company which in turn reduces its chances of getting credit in the future and may have to pay higher interests on existing debts as well as any new obligations. From an investor's point of view, he would want to invest in a company if it is capable of handling its financial obligations, can grow quickly, and is able to manage the growth scale.

A balance sheet is a financial statement of a company that provides a snapshot of what a company owns, owes, and the amount invested by the shareholders. Thus, it is an important tool that helps evaluate the performance of a business.

Data that is available includes information from the financial statement of the companies for the previous year (2015). Also, information about the Networth of the company in the following year (2016) is provided which can be used to drive the labeled field.

Explanation of data fields available in Data Dictionary, 'Credit Default Data Dictionary.xlsx'

1.8 Build a Random Forest Model on Train Dataset. Also showcase your model building approach

Solution-

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

In the previously done exercise, we had successfully completed the necessary steps such as exploratory data analysis, null value and outlier treatment. Then, the data we did split the data into train and test set. Then, we built Logistic Regression Model (using statsmodel library) on most important variables. Most important features were identified manually using backward elimination approach.

Then, we applied the Random Forest Model.

While using the Random Forest Model, we have used the hyper parameters and the best parameters were identified using fit in the grid. It basically gives the combination of best hyper parameters.

Then we will use these best hyper parameters as our final model.

Best parameters-

```
{'max_depth': 15,
 'min_samples_leaf': 5,
 'min_samples_split': 30,
 'n_estimators': 10}
```

On the Train data-

	precision	recall	f1-score	support
0.0	0.98	0.99	0.99	2142
1.0	0.93	0.85	0.89	260
accuracy			0.98	2402
macro avg	0.95	0.92	0.94	2402
weighted avg	0.98	0.98	0.98	2402

Recall- 85%

Precision- 93%

1.9 Validate the Random Forest Model on test Dataset and state the performance matrices. Also state interpretation from the model

Solution-

The model was applied to the test data and the following was observed-

	precision	recall	f1-score	support
0.0	0.98	1.00	0.99	1056
1.0	0.95	0.80	0.87	128
accuracy			0.97	1184
macro avg	0.97	0.90	0.93	1184
weighted avg	0.97	0.97	0.97	1184

Recall- 80%

Precision- 95%

Here we can say that the model is overfitting in nature.

1.10 Build a LDA Model on Train Dataset. Also showcase your model building approach

Solution-

For Linear Discriminant Analysis we have imported the necessary libraries.

After bring the libraries, we have fitted the model and predicted on train as well as test set.

Based on which we have the following confusion matrix-

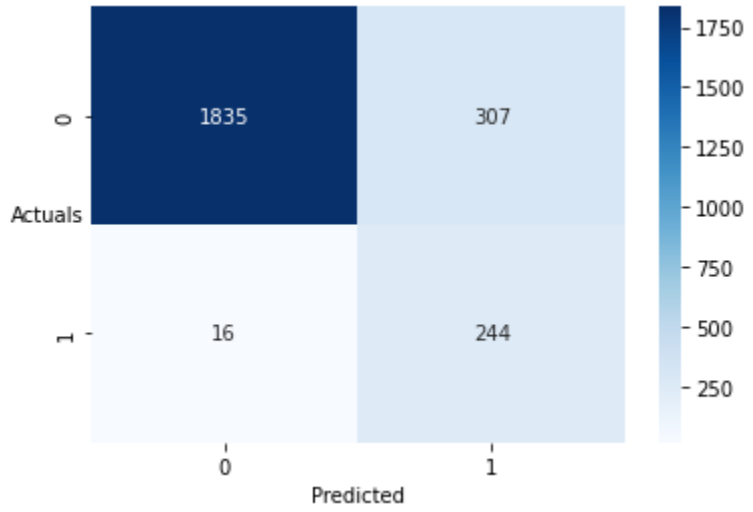
	precision	recall	f1-score	support
0.0	0.93	0.99	0.96	2142
1.0	0.83	0.42	0.56	260
accuracy			0.93	2402
macro avg	0.88	0.71	0.76	2402
weighted avg	0.92	0.93	0.92	2402

The confusion matrix isn't encouraging as Recall is only 42% and precision is 83%. Recall is our priority here.

Therefore, we can try adjusting the thresholds to get better results.

LDA comes from the scikit learn family, so we will have to separately predict the probability and only take the probability of 1s. We have used the function predict_proba. It gives probabilities of 0s as well as 1s, here we are more concerned for 1s. By doing this we have up come with threshold to tune the model further. The threshold got is 0.07817435035548496.

After using the threshold we have got the following confusion matrix-



The number of TP and TNs are high i.e. 2079(1835+244) out of 2402.

	precision	recall	f1-score	support
0.0	0.991	0.857	0.919	2142
1.0	0.443	0.938	0.602	260
accuracy			0.866	2402
macro avg	0.717	0.898	0.760	2402
weighted avg	0.932	0.866	0.885	2402

The model achieved here has Recall 93% and Precision 44%.

Therefore, we can say that the model has improved after using the threshold for tuning.

1.11 Validate the LDA Model on test Dataset and state the performance matrices. Also state interpretation from the model

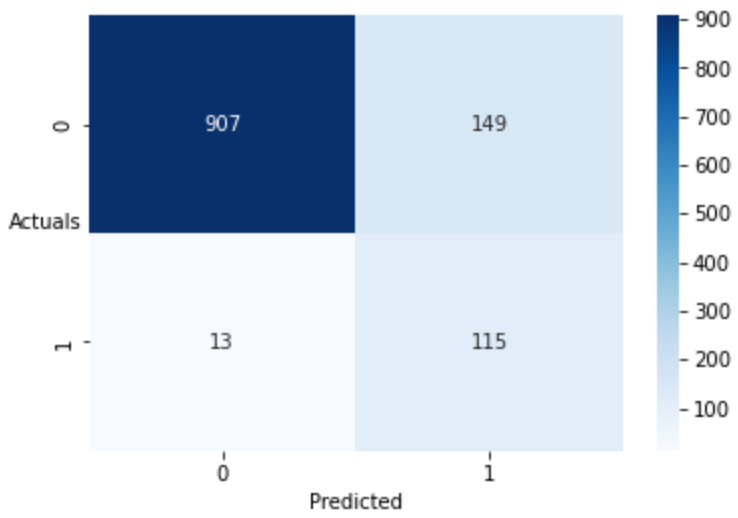
Solution-

Continued from the above sequence, we have checked the confusion matrix for test data and following are the result-

	precision	recall	f1-score	support
0.0	0.93	0.99	0.96	1056
1.0	0.78	0.40	0.53	128
accuracy			0.92	1184
macro avg	0.86	0.69	0.74	1184
weighted avg	0.92	0.92	0.91	1184

It is not a good model as the Recall is only 40% and Precision is 78%. We are more concerned for Recall value here.

On both train as well as test data set the threshold was applied for model tuning and received the following result for the test data.



Here we can see that the total number of TP and TN are 1022 which is fairly good as compared to its performance before applying the threshold.

	precision	recall	f1-score	support
0.0	0.986	0.859	0.918	1056
1.0	0.436	0.898	0.587	128
accuracy			0.863	1184
macro avg	0.711	0.879	0.752	1184
weighted avg	0.926	0.863	0.882	1184

In case of test data as well the model has improved performance significantly.

The Recall value is 89% and the Precision is 43%.

1.12 Compare the performances of Logistics, Radom Forest and LDA models (include ROC Curve)

Solution-

Businesses

1.13 State Recommendations from the above models

Solution-

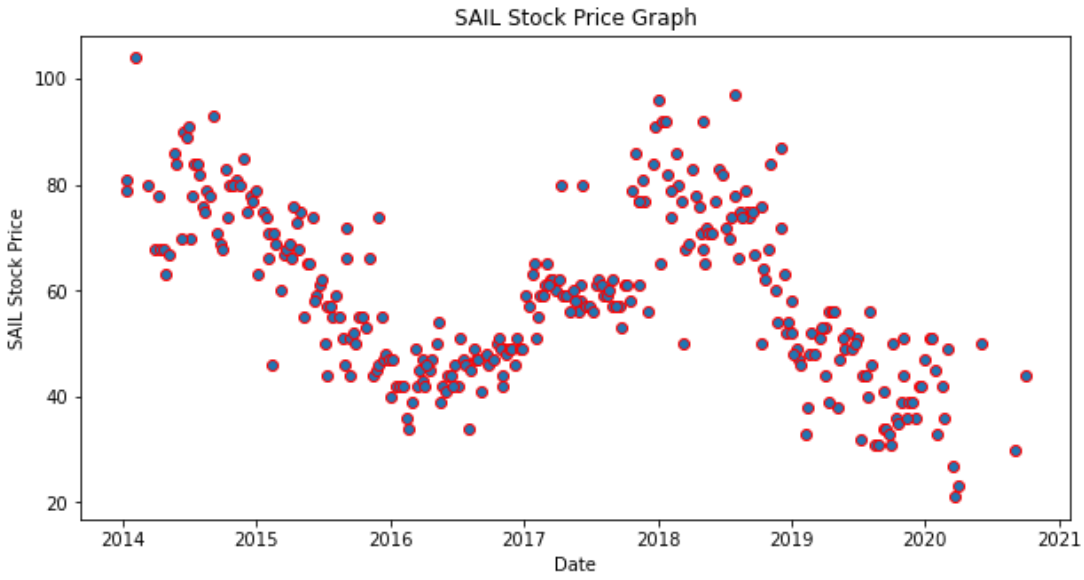
Businesses

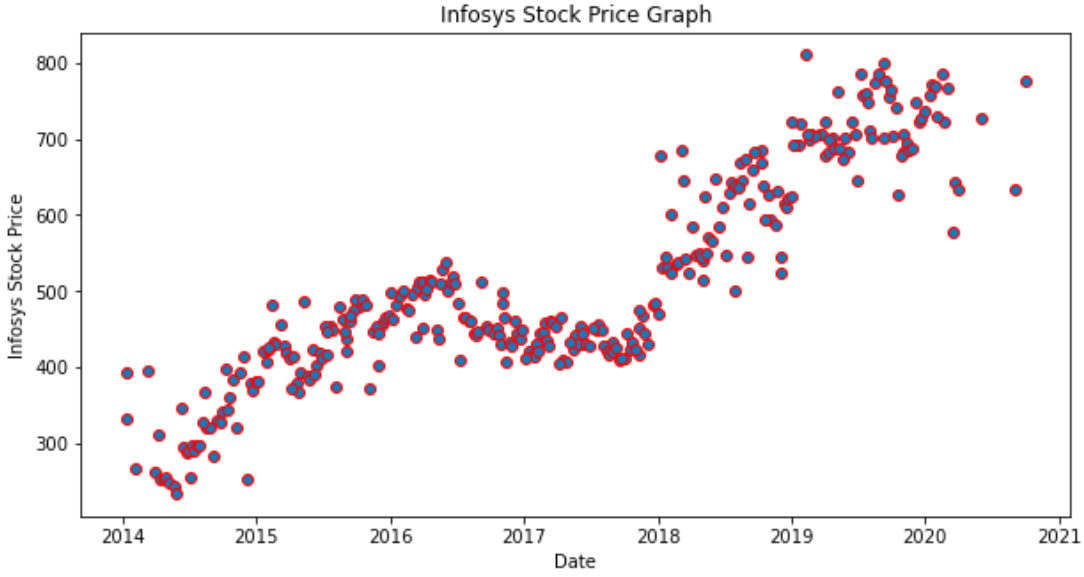
Market Risk

The dataset contains 6 years of information (weekly stock information) on the stock prices of 10 different Indian Stocks. Calculate the mean and standard deviation on the stock returns and share insights.

2.1 Draw Stock Price Graph(Stock Price vs Time) for any 2 given stocks with inference

Solution-





SAIL Stock Price Graph is oscillating between 100 and 20. The SAIL price reached highest during 2014 and has reached to lower value around 40 now. It is not a good choice for short term investors as the prices are going down. However, in long term they might go up therefore people looking for long term investment can go for it.

In case of Infosys, the trends are different. There is a clear visible linear growth in the stock prices. It was in between 300 and 400 till 2015 but now has reached between 700 and 800 in FY 2020-21. So Infosys can be good choice for an investor.

2.2 Calculate Returns for all stocks with inference

Solution-

Below is the return of all the stocks with returns.

Here the model is suggesting that in the portfolio to get 0.60% return we should go for Infosys and Shree Cement. Here the volatility is 0.182%. The study encourages us not to include Indian Hotel, Mahindra & Mahindra, Axis Bank, SAIL, Sun Pharma, Jindal Steel, Idea Vodafone, Jet Airways in the portfolio.

	Weights	Average Return
Infosys	0.747212	0.34%
Indian Hotel	0	0.14%
Mahindra & Mahindra	0	-0.07%
Axis Bank	0	0.22%
SAIL	0	-0.15%
Shree Cement	0.769544	0.45%
Sun Pharma	0	-0.05%
Jindal Steel	0	-0.13%
Idea Vodafone	0	-0.51%
Jet Airways	0	-0.48%

2.3 Calculate Stock Means and Standard Deviation for all stocks with inference

Solution-

Standard Deviation-

Stock	Standard Deviation
Infosys	0.004717203

Indian Hotel	0.007246377
Mahindra & Mahindra	0.018655065
Axis Bank	0.00167406
SAIL	0.032913165
Shree Cement	0.020084069
Sun Pharma	0.02883949
Jindal Steel	0.007446648
Idea Vodafone	0.006024096
Jet Airways	0.044964029

Standard Mean-

	Average Return
Infosys	0.34%
Indian Hotel	0.14%
Mahindra & Mahindra	-0.07%
Axis Bank	0.22%
SAIL	-0.15%
Shree Cement	0.45%
Sun Pharma	-0.05%
Jindal Steel	-0.13%
Idea Vodafone	-0.51%
Jet Airways	-0.48%

The

2.4 Draw a plot of Stock Means vs Standard Deviation and state your inference

Solution-

2.5 Conclusion and Recommendations

Solution-

Here the model is suggesting that in the portfolio to get 0.60% return we should go for Infosys and Shree Cement. Here the volatility is 0.182%. The study encourages us not to include Indian Hotel, Mahindra & Mahindra, Axis Bank, SAIL, Sun Pharma, Jindal Steel, Idea Vodafone, Jet Airways in the portfolio.