FRA Project: Milestone-2 Business Report

By

Ritusri Mohan

CONTENTS

PROBLEM 1

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.

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

Random forest model was built, using GridSearchCV for hyperparameter tuning.The following grid was used:

```
{
    'max_depth': [20,30,40],
    'max_features': [2,3,4,5,6,7,8],
    'min_samples_leaf': [25,50,75,100],
    'min_samples_split': [25,50,75,100],
    'n_estimators': [50,100,150]
}
```

```
RandomForestClassifier(max_depth=20, max_features=7, min_samples_leaf=25,
                       min_samples_split=50, n_estimators=150, n_jobs=-1,
                       oob_score=True)
```
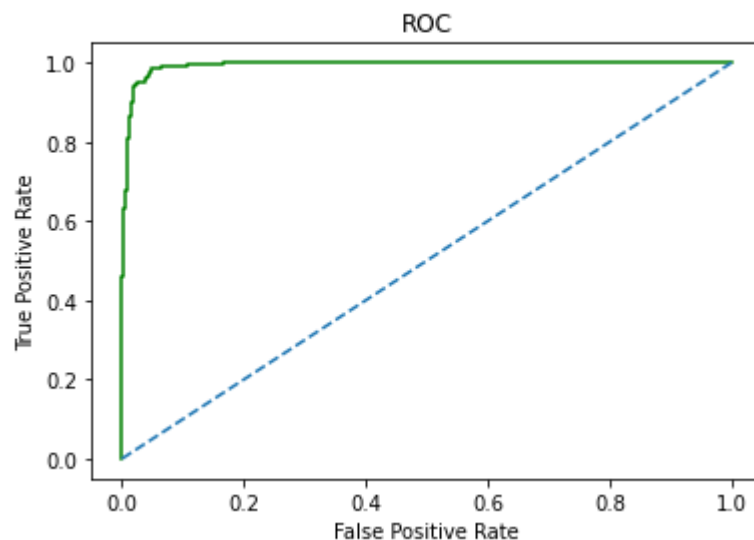
Model On Train Data

- Confusion matrix

```
array([[2141,   31],
       [  31,  199]], dtype=int64)
```

- Classification report

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2172
1	0.87	0.87	0.87	230
accuracy			0.97	2402
macro avg	0.93	0.93	0.93	2402
weighted avg	0.97	0.97	0.97	2402

- ROC Curve

```
Area under Curve is 0.9932300424373448
```


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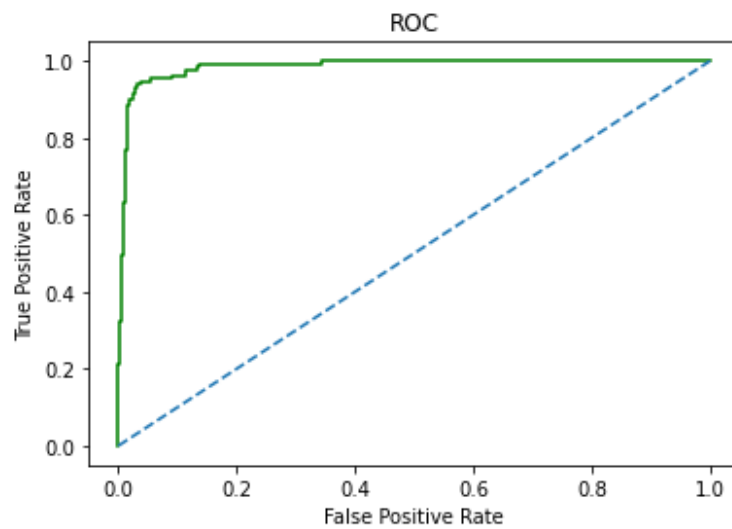
Model On Test Data

- Confusion matrix

```
array([[1032,   19],
       [  13,  120]], dtype=int64)
```

- Classification report

	precision	recall	f1-score	support
0	0.99	0.98	0.98	1051
1	0.86	0.90	0.88	133
accuracy			0.97	1184
macro avg	0.93	0.94	0.93	1184
weighted avg	0.97	0.97	0.97	1184

- ROC Curve

Area under Curve is 0.9855132598384639

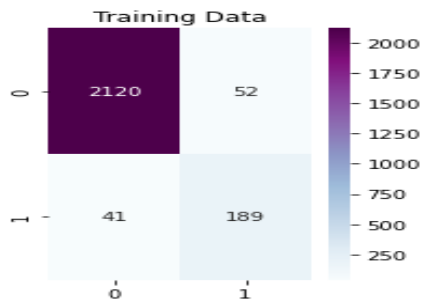## 1.10 Build a LDA Model on Train Dataset. Also showcase your model building approach.

Linear Discriminant Analysis model was built.

Model On Train Data

```
LinearDiscriminantAnalysis()
```
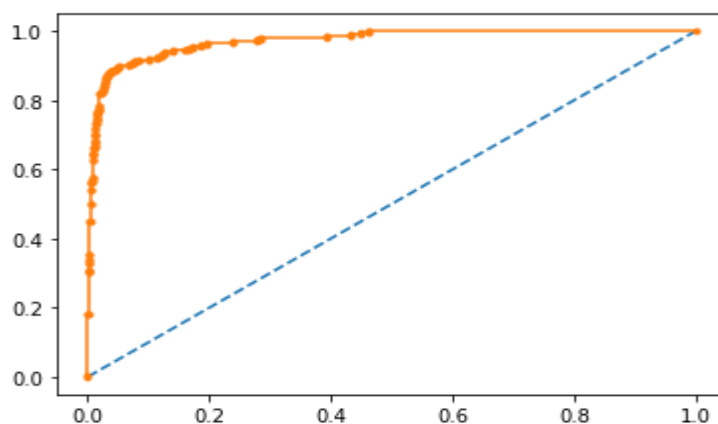
- Confusion matrix

- Classification report

	precision	recall	f1-score	support
0	0.98	0.98	0.98	2172
1	0.78	0.82	0.80	230
accuracy			0.96	2402
macro avg	0.88	0.90	0.89	2402
weighted avg	0.96	0.96	0.96	2402

- ROC Curve

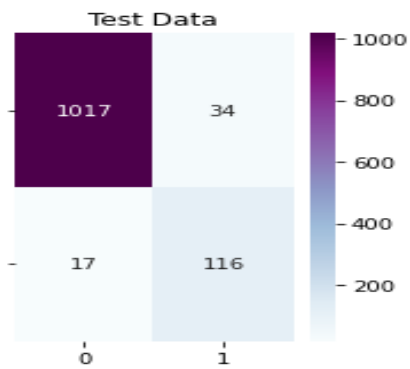AUC for the Training Data: 0.973

[<matplotlib.lines.Line2D at 0x18d8620ed88>]

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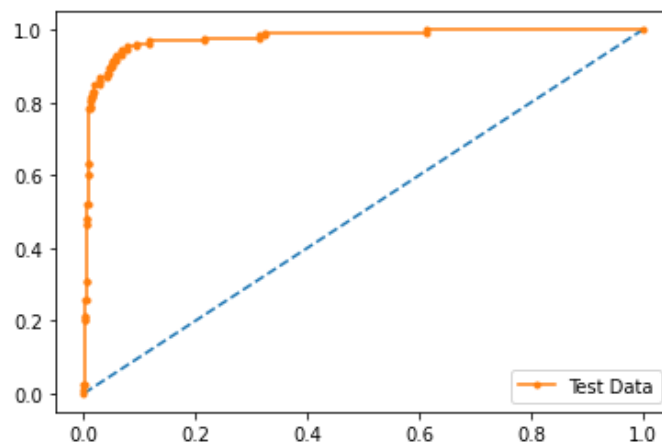
Model On Test Data

- Confusion matrix

- Classification report

	precision	recall	f1-score	support
0	0.98	0.97	0.98	1051
1	0.77	0.87	0.82	133
accuracy			0.96	1184
macro avg	0.88	0.92	0.90	1184
weighted avg	0.96	0.96	0.96	1184

- ROC Curve

AUC for the Test Data: 0.977

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

	Logistic reg Train	Logistic reg Test	LDA Train	LDA Test	RFCL Train	RFCL Test
Accuracy	0.91	0.89	0.96	0.96	0.97	0.97
Recall	0.99	0.99	0.98	0.97	0.99	0.98
Precision	0.90	0.90	0.98	0.98	0.99	0.99
F1 Score	0.94	0.94	0.98	0.98	0.99	0.98

From the above table it can be seen that the Random Forest Classifier model is the best with an accuracy of 97%

1.13 State Recommendations from the above models

LOGIT MODEL:

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	0.0876	0.250	0.351	0.726	-0.402	0.577
ROG_Gross_Block_perc	-0.0008	0.000	-5.190	0.000	-0.001	-0.001
Current_Ratio_Latest_	-0.0087	0.001	-11.432	0.000	-0.010	-0.007
Debtors_Ratio_Latest_	-0.0006	0.000	-4.293	0.000	-0.001	-0.000
Cash_Flow_From_Investing_Activities	0.0004	0.000	3.087	0.002	0.000	0.001
Total_Debt	0.0005	9e-05	5.850	0.000	0.000	0.001
Cash_Flow_From_Operating_Activities	-0.0006	0.000	-4.950	0.000	-0.001	-0.000
Cash_Flow_From_Financing_Activities	-0.0002	0.000	-2.195	0.028	-0.000	-2.62e-05
Value_Of_Output	-0.0003	8.2e-05	-3.823	0.000	-0.000	-0.000

- Lower the Gross_Block_perc, higher is the chance of a default, which would mean the Gross block percent is expected to be negative.
- Lower the Current_Ratio_Latest, higher is the chance of a default, which would mean the Current ratio is expected to be negative.
- Lower the Debtors_Ratio_Latest, higher is the chance of a default, which would mean the Debtor ratio is expected to be negative.
- Higher the Cash_Flow_From_Investing_Activities, higher is the chance of a default, which would mean the cash flow is expected to be positive.
- Lower the Cash_Flow_From_Operating_Activities, higher is the chance of a default, which would mean the cash flow is expected to be negative.

Curr_Ratio_Latest is most important criteria amongst the above parameters, while the Gross_Block_perc is the least important.

PROBLEM 2

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

Stock price graphs of Mahindra & Mahindra vs Time:
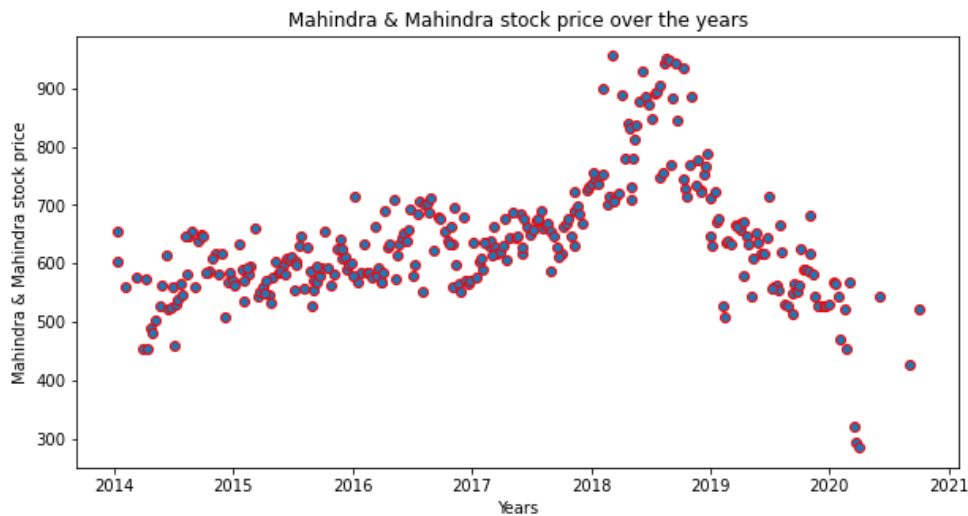
Stock price graphs of Infosys vs Time:
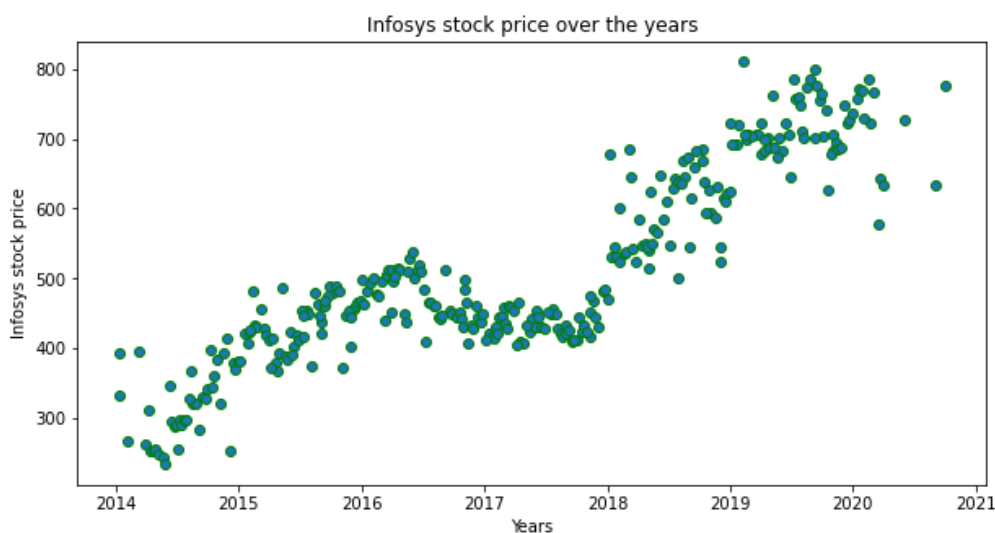
2.2 Calculate Returns for all stocks with inference

Returns for all the stocks i.e. difference of log of price at t and the log of price at t-1 are shown below.

	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	-0.026873	-0.014599	0.006572	0.048247	0.028988	0.032831	0.094491	-0.065882	0.011976	0.086112
2	-0.011742	0.000000	-0.008772	-0.021979	-0.028988	-0.013888	-0.004930	0.000000	-0.011976	-0.078943
3	-0.003945	0.000000	0.072218	0.047025	0.000000	0.007583	-0.004955	-0.018084	0.000000	0.007117
4	0.011788	-0.045120	-0.012371	-0.003540	-0.076373	-0.019515	0.011523	-0.140857	-0.049393	-0.148846

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

	Stock Means	Stock Standard Deviation
Infosys	0.002794	0.035070
Indian Hotel	0.000266	0.047131
Mahindra & Mahindra	-0.001506	0.040169
Axis Bank	0.001167	0.045828
SAIL	-0.003463	0.062188
Shree Cement	0.003681	0.039917
Sun Pharma	-0.001455	0.045033
Jindal Steel	-0.004123	0.075108
Idea Vodafone	-0.010608	0.104315
Jet Airways	-0.009548	0.097972

Shree Cement has the highest value of stock means and Idea Vodafone has the highest value of stock standard deviation.

2.4 Draw a plot of Stock Means vs Standard Deviation and state your inference
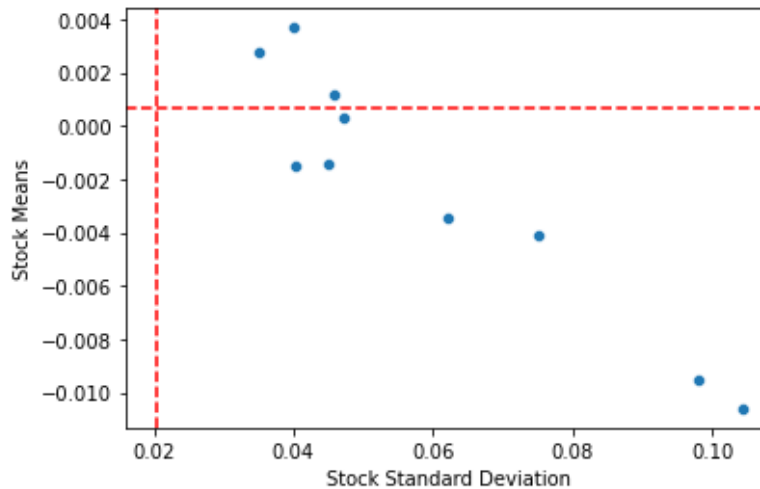
While the stocks on the bottom right imply low volatility and good returns, the stocks higher up and to the far left indicate high volatility and low returns. This graph can be used to balance risk and benefit when considering investments in various businesses.

2.5 Conclusion and Recommendations

When there are competing stocks with higher returns and lower risk in a portfolio, stocks with a lower mean and higher standard deviation do not have a place. Thus, there are just a few stocks left for the data we have here:

two options:
- one with the best return and the lowest risk
- one with the best risk and the lowest return

Therefore, in this dataset, Shree Cement appears to be the best performer from a pure Returns standpoint, followed by Infosys and Axis Bank. In this dataset, Infosys appears to be the least risky company, followed by Mahindra & Mahindra, Shree Cement, and the standard deviation.

To evaluate the risk to reward ratio, we suggest utilising the stock means vs. standard deviation graphic. Even though more volatile stocks may produce short-term gains, they may not be wise long-term investments. In contrast, a stock with minimal volatility may not make for a smart investment in the short term but may do so in the long run. The above-mentioned plot should therefore be inferred based on the type of investment that one is seeking.

THE END