Market Risk Analysis

Balasubramaniyam.R

Problem statement: 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.
You are expected to do the Market Risk Analysis using Python.

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

The Random Forest model is built on the entire dataset. The dataset is highly imbalanced and therefore SMOTE is used to balance the dataset. A model is built using SMOTE data and found that it is generating an overfitted model. Therefore, we try to build the model without SMOTE and let's observe the output.

The dataset is then split into train and test dataset and a random forest model is built on train data.

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

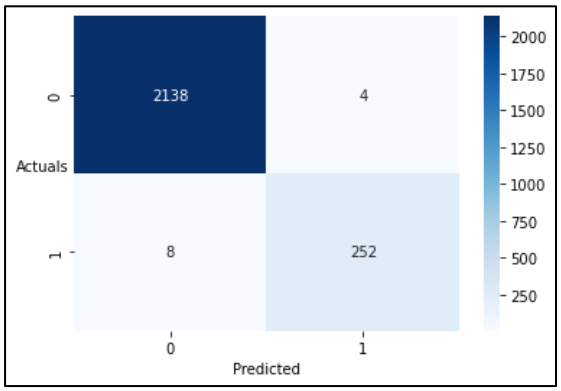
	precision	recall	f1-score	support
0	1.00	1.00	1.00	2142
1	0.98	0.97	0.98	260
accuracy			1.00	2402
macro avg	0.99	0.98	0.99	2402
weighted avg	0.99	1.00	0.99	2402

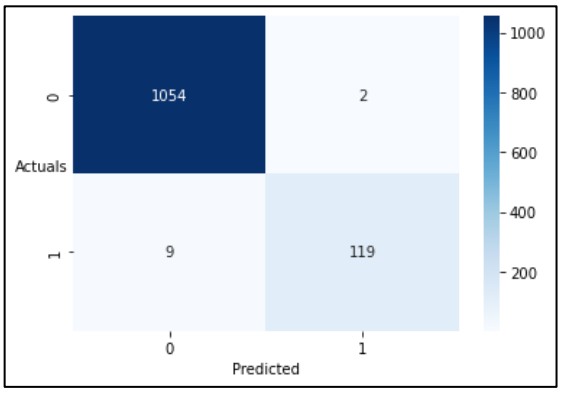
	precision	recall	f1-score	support
0	0.99	1.00	0.99	1056
1	0.98	0.93	0.96	128
accuracy			0.99	1184
macro avg	0.99	0.96	0.98	1184
weighted avg	0.99	0.99	0.99	1184

1.9.1. Classification_report(Train data), Classification_report(Test data).

From the Classification report we can observe the RF model is performing well on the train dataset and test dataset.

The model has a recall value of 0.97 on train data and 0.93 on the test data. This means the model has performed well on both train and test datasets.



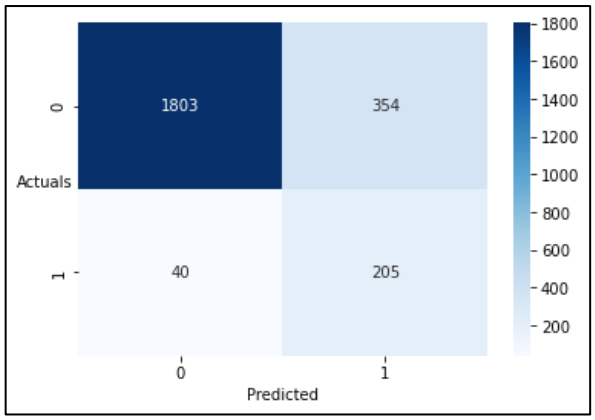


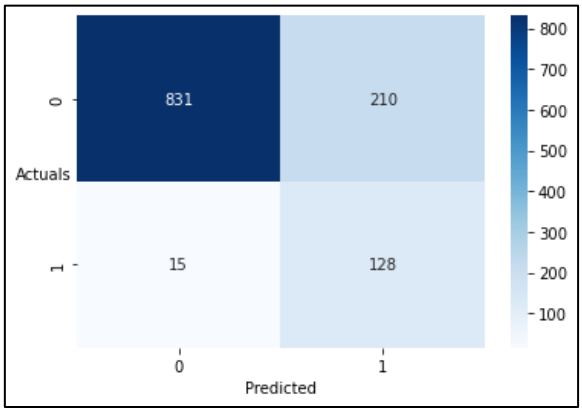
1.9.2. Confusion matrix (Train data), Confusion matrix (Test data).

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

A LDA model is built on the Train and Validated using the test dataset.

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





1.11.1. Confusion matrix (Train data), Confusion matrix (Test data).

	precision	recall	f1-score	support
0	0.978	0.836	0.901	2157
1	0.367	0.837	0.510	245
accuracy			0.836	2402
macro avg	0.673	0.836	0.706	2402
weighted avg	0.916	0.836	0.862	2402

	precision	recall	f1-score	support
0	0.982	0.798	0.881	1041
1	0.379	0.895	0.532	143
accuracy			0.810	1184
macro avg	0.680	0.847	0.706	1184
weighted avg	0.909	0.810	0.839	1184

1.11.2. Classification_report(Train data), Classification_report(Test data).

From the Classification report we can observe the LDA model is performing well on the train dataset and test dataset.

The model has a recall value of 0.83 on train data and 0.89 on the test data. This means the model has performed well on both train and test dataset.

In fact, the model has performed really well on test dataset than the train data.

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

	precision	recall	f1-score	support
0	0.981	0.857	0.915	2157
1	0.404	0.853	0.549	245
accuracy			0.857	2402
macro avg	0.693	0.855	0.732	2402
weighted avg	0.922	0.857	0.878	2402

	precision	recall	f1-score	support
0	0.983	0.821	0.895	1041
1	0.408	0.895	0.560	143
accuracy			0.830	1184
macro avg	0.695	0.858	0.727	1184
weighted avg	0.913	0.830	0.854	1184

1.12.1. Logistic Regression Classification_report(Train data), Classification_report(Test data).

The logistic regression and LDA models are having identical performance on the test dataset.

Based on this we can conclude both these models are suitable for our dataset.

The RF model is having higher recall than LDA and regression model however there may be a case of overfitting of the RF model.

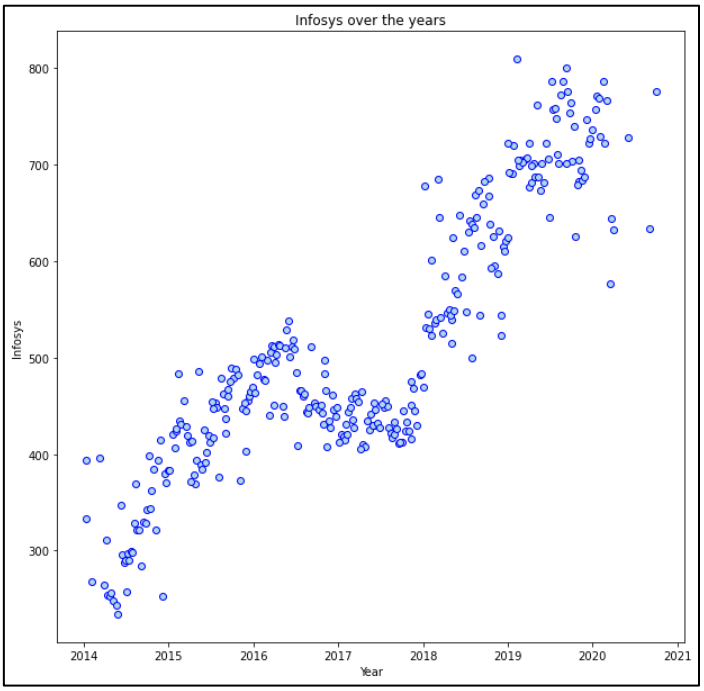
1.13 State Recommendations from the above models

It is recommended to use the Logistic regression model for finding the default variable in our dataset.

The Logistic regression model is recommended because it has performed well on the train data set as well as the train data set.

The RF model is having really good results by observing its confusion matrix, The RF model can be used for prediction by collecting a more balanced dataset and tuning the model to avoid overfitting of the model.

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



2.1.1. Infosys Stock prices

The above plot of Infosys indicates that there is an overall upward trend in the stock price from the year 2014 to 2020. There was a decrease in the stock price from mid of 2016 to the mid of 2017 after this period the stock prices have recovered well to a upward trend.

2.1.2. Idea_Vodafone Stock prices

The above plot of Idea_Vodafone indicates that there is an overall downward trend in the stock price from the year 2014 to 2020. There was a decrease in the stock price from mid of 2015 after this year the stock prices have faced a downward trend.

2.2 Calculate Returns for all stocks with inference.

There are ten companies in our dataset and the returns for these stocks are calculated, returns for the first 10 records are as follows,

	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
5	-0.031749	-0.015504	0.040656	0.061875	0.061558	0.011400	-0.008217	0.024898	0.012579	-0.016598
6	0.019961	0.060625	0.011881	0.076961	0.112795	0.067622	-0.016639	0.097543	0.048790	0.020705
7	-0.036221	0.199333	0.038615	0.059898	0.136859	0.056790	-0.049881	0.105732	-0.024098	0.169258
8	-0.041847	-0.012121	0.064183	-0.014642	-0.023530	0.048090	0.044835	-0.010084	-0.012270	-0.181630
9	0.135666	0.081917	-0.003559	0.071154	0.213574	0.105167	-0.018724	0.132686	0.024391	0.072031

2.2.1. Returns for all stocks.

The maximum and minimum returns from each company from 2014 to 2020 are as follows,

Infosys	0.135666
Indian_Hotel	0.199333
Mahindra_Mahindra	0.089407
Axis_Bank	0.127461
SAIL	0.309005
Shree_Cement	0.152329
Sun_Pharma	0.166604
Jindal_Steel	0.243978
Idea_Vodafone	0.693147
Jet_Airways	0.300249

Infosys	-0.167300
Indian_Hotel	-0.236389
Mahindra_Mahindra	-0.285343
Axis_Bank	-0.284757
SAIL	-0.251314
Shree_Cement	-0.129215
Sun_Pharma	-0.179855
Jindal_Steel	-0.283768
Idea_Vodafone	-0.693147
Jet_Airways	-0.458575

2.2.2. Maximum, Minimum Returns from each stock.

The Idea Vodafone company has a maximum positive return of 0.69 and Mahindra & Mahindra has a least positive return of 0.089.

The Idea Vodafone company has a maximum negative return of -0.69 and Shree cement has a least negative return of -0.129. The Idea Vodafone company is having a significant reduction in its stock price.

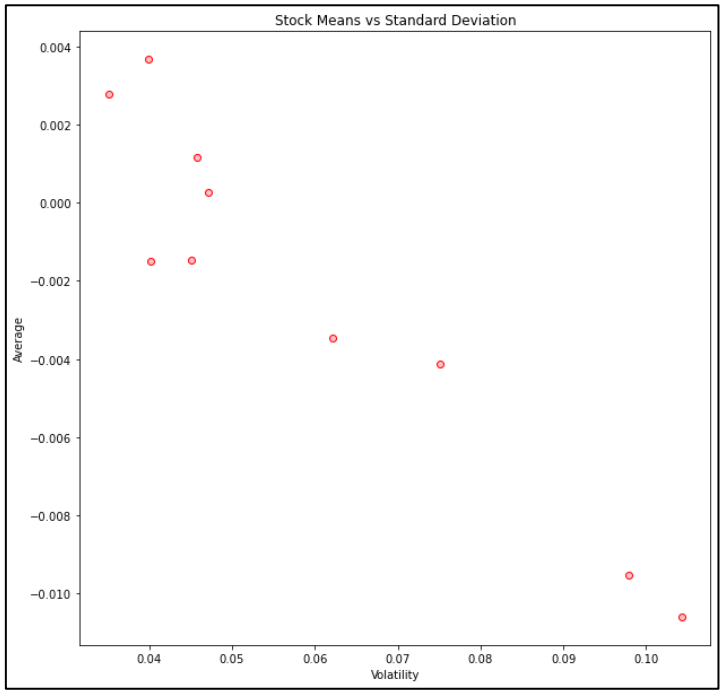
2.3 Calculate Stock Means and Standard Deviation for all stocks with inference.

	Average	Volatility
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

2.3.1. Average & Volatility of each Company.

The average and volatility of the stock represents the average returns and the risk associated with them. The Idea Vodafone company has maximum volatility and therefore it is the stock with high risk in these ten companies. Infosys is having the least risk in terms of investment. Shree cement is having good avg. returns with moderate volatility.

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



2.4.1. Average Vs Volatility.

The average vs volatility plot indicates that the Idea Vodafone and Jet airways are having higher risk associated with them with high negative returns.

Shree cement and Infosys companies are having the higher returns with the least risk involved among these ten companies.

Axis bank and Indian hotel stocks are having the 3rd and 4th highest returns with moderate risk associated with them.

Mahindra and Mahindra, SAIL, Sun Pharma, Jindal Steel, Idea Vodafone, Jet airways are having negative returns and high risk associated with them.

2.5 Conclusion and Recommendations

The dataset contains ten companies stock prices. From the analysis of this data, we can conclude that only Infosys, Indian hotel, Axis bank, Shree cement are suitable for investment. Other six companies are not suitable for investment at the moment as they haven't generated any positive returns and also highly volatile.

Shree cement, Infosys are most recommended companies to make an investment. The Shree cement is higher return and comparatively more risk than the Infosys company.

Axis bank and Indian hotel are second most recommended companies to make an investment. Axis bank is having good returns and less risk compared to Indian hotel.