**Submission Number: 1**

**Group Number: 08**

**Group Members:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Full Legal Name** | **Location (Country)** | **E-Mail Address** | **Non-Contributing Member (X)** |
| Monika Singh | India | Ms19626@gmail.com |  |
| Ananta Narayana Satapathy | India | ansatapathy@yahoo.com |  |
| Harshil Sumra | India | Harshilsumra1997@gmail.com |  |

**Statement of integrity:** By typing the names of all group members in the text box below, you confirm that the assignment submitted is original work produced by the group (*excluding any non-contributing members identified with an “X” above*).

Monika Singh

Ananta Narayana Satapathy

Harshil Sumra

11. Which model - PCA regression or lasso -- provides a better fit to the data? Why?

When we compared R-squared and MSE metrics for our model evaluation, we found that Lasso performed well than PCA regression. Compared to the PCA regression Lasso provides a better fit to the data.

R-squared values for Lasso is 97% and for PCA it is 96%. MSE values are 1372.8 and 2349.6 for Lasso and PCA respectively.

12. Which model provides better interpretation of the results?

The Lasso model provides better interpretation of results compared than the PCA regression. When we compared R-squared and MSE result based on those we found that Lasso model is better.

13. How did your group divide the work?

We have divided the work into three parts since we are three members. First part is Data importing, dimensional data summaries preparation and data exploration. Second part consist of model building such as PCA regression and Lasso. Third part consist of generating technical and non-technical report to senior management.

14. Write a TECHNICAL 1-page report of your findings to your FE boss.

**Problem Overview**

Our objective is to predict the performance of the stock market. We have used LUX Index (Stock market index of the Luxembourg stock Exchange) as dependent variable and MSCI Indexes as independent variables from various areas of the world.

**Algorithm Explanation**

We have used supervised algorithm to evaluate the relationship between independent and dependent variable. We have performed the dimensionality reduction to the input data to improve our regression model. We have used PCA technique for that and Lasso for feature selection and regularization.

**Experimental Evaluation**

To evaluate our model we have used R-squares and MSE metrics. We have divided our data training data into 2/3rd and 1/3rd to our testing data. R-squared help us to check goodness of fit. Higher the value of fitness better the model has performed. MSE is a mean squared error which is the calculation of the mean of error which is difference of actual value and predicted value.

**Results and Conclusion**

R-squared shows that Lasso has reached 97% and PCA 96% and MSE is also towards Lasso giving value of 1372.8 and for PCA 2349.6. On the basis of those metrics we can say that Lasso has performed better than PCA.

Chart, scatter chart

Description automatically generatedChart

Description automatically generated with medium confidence

15. Write a non-technical 1 paragraph email of your findings for senior management

We have performed an analysis to check the relationship between the stock market index in various parts of the world. Our investigation shows that there is a relationship between various MSCI indexes. We have used LUX index as our dependent variable and the others left as independent variable. We have used regression supervised algorithm for the analysis. To evaluate the model we have used metrics which basically capture the difference between the actual value and predicted value. We have concluded that there is an accuracy of 97% from our best performing model.

**References**

https://www.thebalance.com/msci-index-what-is-it-and-what-does-it-measure-330594 8

https://towardsdatascience.com/explainable-artificial-intelligence-part-2-model-interpr etation-strategies-75d4afa6b739

https://www.cs.utexas.edu/~mooney/cs391L/paper-template.html