Analysis of National Stock Exchange of India

Ruchita Shinde, Nikita Shendkar, Shahnawaz Khan, Sohong Chakraborty

Department of Information Systems, California State University

Los Angeles

**Abstract:** When faced with the task of analyzing nearly one million rows, Microsoft Azure Machine Learning Studio and Databricks proved to be of real help. The NSE data set was the perfect data set to study the co-relation between columns and predict the number of stocks exchange (volume).

1. **Introduction**

Nifty50.csv The NIFTY 50 index is National Stock Exchange of India's benchmark stock market index for Indian equity market. It is a well diversified 50 stock index accounting for 22 sectors of the economy. It is used for a variety of purposes such as bench-marking fund portfolios, index based derivatives and index funds.

The National Stock Exchange of India Limited (NSE) is the leading stock exchange of India, located in Mumbai. The NSE was established in 1992 as the first demutualized electronic exchange in the country. NSE was the first exchange in the country to provide a modern, fully automated screen-based electronic trading system which offered easy trading facility to the investors spread across the length and breadth of the country. Vikram Limaye is Managing Director & Chief Executive Officer (MD & CEO) of NSE.

National Stock Exchange has a total market capitalization of more than US$1.41 trillion, making it the world’s 12th-largest stock exchange as of March 2017. NSE's flagship index, the NIFTY 50, the 50 stock index is used extensively by investors in India and around the world as a barometer of the Indian capital markets. Nifty 50 index was launched in 1996 by the NSE. However, Vaidyanathan (2016) estimates that only about 4% of the Indian economy / GDP is actually derived from the stock exchanges in India.

Unlike countries like the United States where nearly 70% of the GDP is derived from larger companies and the corporate sector, the corporate sector in India accounts for only 12-14% of the national GDP (as of October 2016). Of these only 7,800 companies are listed of which only 4000 trade on the stock exchanges at BSE and NSE. Hence the stocks trading at the BSE and NSE account for only around 4% of the Indian economy, which derives most of its income related activity from the so-called unorganized sector and households.

Economic Times estimated that as of April 2018, 60 million (6 crore) retail investors had invested their savings in stocks in India, either through direct purchases of equities or through mutual funds. Earlier, the Bimal Jalan Committee report estimated that barely 1.3% of India's population invested in the stock market, as compared to 27% in USA and 10% in China.

The reason for choosing this data set was that it had enough rows and columns to carry out a successful analysis and to predict the volumes of stocks exchanged.

1. **Related Work**

We chose two models:

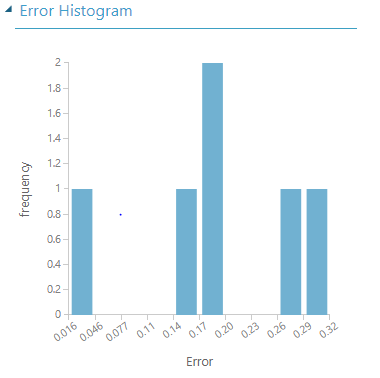
* Linear Regression
* Neural Networks

Linear Regression: In Linear regression, a single independent variable is used to predict the value of a dependent value.

Neural Networks Regression: They are one of the learning algorithms used within machine learning. They consist of different layers for analyzing and learning data.

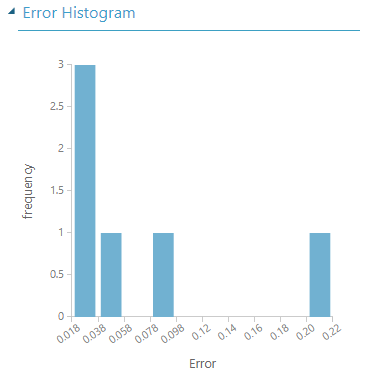
While running the dataset on the Azure ML platform, linear regression, the root mean square came out to be 0.207358 and the Coefficient of determination was 0.967752. For neural network regression it was 0.98181

for root mean square error and 0.99277 for the coefficient of determination.



We tried using the Hyper tune model as well. But it was not fetching a good coefficient of determination value. So, we concluded that the neural network works best in Azure ML studio.

In Databricks PySpark, the RMSE was 24751.8266 for linear regression and 23882.667 for neural network regression.



1. **Conclusion**

After concluding the data analysis using both Azure ML Studio and Databricks, we can conclude that Neural Network regression was the best for our data set. The RMSE though came out to be very different for both the platforms as RMSE values are different for different domains.

**References**

**1.**<https://docs.databricks.com/spark/latest/mllib/decisin-trees.html#gbt-regression-using-mllib-pipelines>

**2.**<https://docs.databricks.com/spark/latest/mllib/mllib-pipelines-and-stuctured-streaming.html>

**3.**<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/machine-learning-modules>

**4.**<https://docs.microsoft.com/en-us/azure/machine-learning/desktop-workbench/data-prep-python-extensibility-overview>