INTERNSHIP REPORT

SUMMER - 2019

NAME – PARITOSH KALLA

COMPANY NAME – SYNECHRON INC.

EXECUTIVE SUMMARY –

Synechron is one of the leading digital transformation companies focused exclusively on financial services industry. It is helping global financial services and insurance companies embrace the most cutting-edge innovations to evolve their businesses. Synechron uniquely delivers these firms an end-to-end Digital, Consulting and Technology capabilities with expertise in capital markets, wealth management and insurance as well as emerging technologies like Blockchain and Artificial Intelligence. Synechron's well-regarded data practice has strong experience in leading major engagements in Enterprise Data Architecture and Governance, Data Analytics and AI, Data Visualization, and Data Quality Improvement. The company works with clients to improve data architecture for business and IT, reduce data management and sourcing costs by helping them to improve data quality and reporting, monetize the value of data, improve compliance, and ultimately, ensure the effectiveness of business decisions.

Th use case for my internship was to use Statistical models and Deep Learning to predict the volume and amount for different traded currency pairs with respect to the Forex products.

Challenges Faced –

1. **Learning Python and Machine Learning using Python** – The internship involved using R and Python for creating a predictive model for currency pairs. Since I had not studied Python for Data Science, so I started the internship by learning Machine Learning using Python (Udemy), along with Machine Learning I also practiced Python coding (Hackerrank).
   Learning - This challenge helped me to learn Python libraries like Scikit Learn, NumPy, Pandas, math, Keras.

2. **Getting the data in proper format to apply the statistical models –** The data given to me was in pdf format and was only very few rows, the challenge here was to convert the data into readable format and simulate it so as apply the Statistical and Deep Learning models. I used online tool to convert the data into excel and used random library in Python and arima.sim function in R to simulate data.
   Learning - The challenge helped me to use Python Data structures, for loops to iterate through the dataframe, using nonlinear series, simple exponential smoothing, complex exponential smoothing, heterogenous variance.

3. **Using Univariate or Multivariate Analysis for the model –** The simulated dataframe had 10 currency pairs for Forex forward product. The currency pairs had one currency as USD so I started with multivariate analysis assuming that the currency pairs would be correlated to each other, but the results of predicted values when compared to test set did not match, after discussing the results with my manager, he suggested univariate analysis as there is no causation between the currency pairs, For example - BRL(Brazilian Real) depreciation against USD may not have any causal effect relating to MXN(Mexican peso). Depreciation of BRL may be accounted for changes in economic policies related to Brazil which, if not trading with Mexico, may not show up in MXN at all. He also suggested to use R for time series analysis as R gives better results than Python for time series analysis.
   Learning – Learned Vector Auto Regression model for Multivariate time series analysis and ARIMA model for univariate time series.

4. **Using Deep Learning Model –** The results from VAR model and ARIMA model were not satisfactory as it showed a lot of variation in predicted values and test set. To get better results, I had to apply deep learning algorithm. Recurrent Neural Network are required when there are time dependencies in the data, but RNN is suitable only for short-term dependencies because of Vanishing Gradient problem. To tackle this problem, I used Long short-term memory (LSTM) network which is an improvement over RNN and works fine with long term dependencies.
   Learning – Talking of Deep learning models, there are five types of LSTM models which can be used to conduct predictive analytics. The best predictive results are obtained by

CNN LSTM. A CNN model can be used in a hybrid model with an LSTM backend where the CNN is used to interpret subsequences of input that together are provided as a sequence to an LSTM model to interpret.

5. **Study the external factors which effect the currency pair trading** – All the above steps involved creating a predictive model using only historical data but in real world there are also a lot of external factors that affect the trading volume and amount. Facebook has a Python library 'Prophet' which deals with such external factors (Ex – political, economic etc.), it takes into account external factors in form of dates and add it to the historical data available.

   Learning – 'Prophet' library for Time Series Analysis

Reference –

www.synechron.com
www.machinelearningmastery.com
www.analyticsvidya.com