Bondai

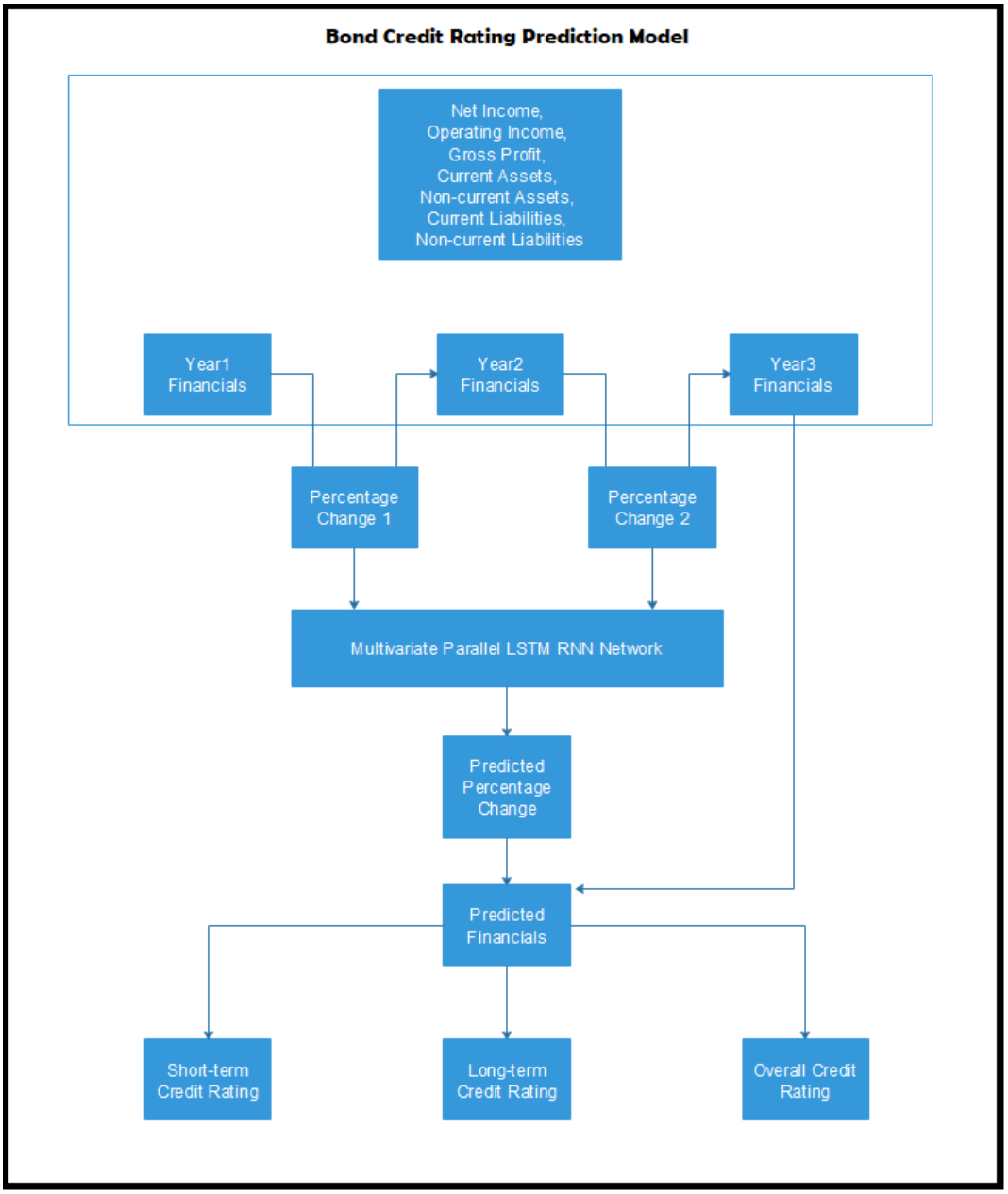
1. Solution Design –

1.1. Prototype – In our protype we have developed the two critical components of the proposed credit rating system.

i. Bond Credit Rating Neural Network and
ii. Financial news-based sentiment analysis

- **Bond Credit Neural Network –**

We have developed a *Multivariate Parallel LSTM RNN Network* which predicts the financials of the company, more specifically the current and non-current assets, the current and non-current liabilities by looking at the financials of the trailing three years.



Since the financials can vary from company to company depending on its market capitalization, it was very important the data used to train the network be accordingly normalized to obtain good and reliable results. So, instead of directly feeding the financial numbers we have trained our network based on the percentage change of the financial numbers in between two consequent annual reports. And based on these percentage change for the last three annual reports, our network predicts the expected percentage change for the next year as compared to the current financials.

The predicted percentage change is then used to calculate the actual financials of the company followed by the short-term, long-term and the overall credit ratings for the respective companies.

The formulas used for credit ratings –

$$Short\ term\ credit\ rating = \frac{Total\ current\ assets}{Total\ current\ liabilities}$$

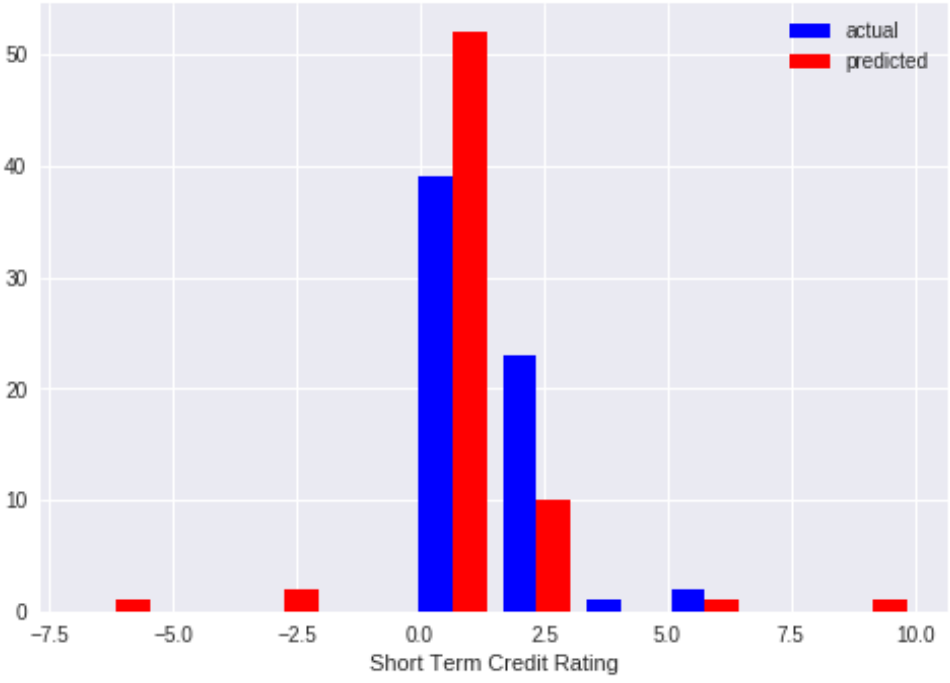
$$Long\ term\ credit\ rating = \frac{Total\ non\ current\ assets}{Total\ non\ current\ liabilities}$$

$$Overall\ credit\ rating = \frac{Total\ assets}{Total\ liabilities}$$

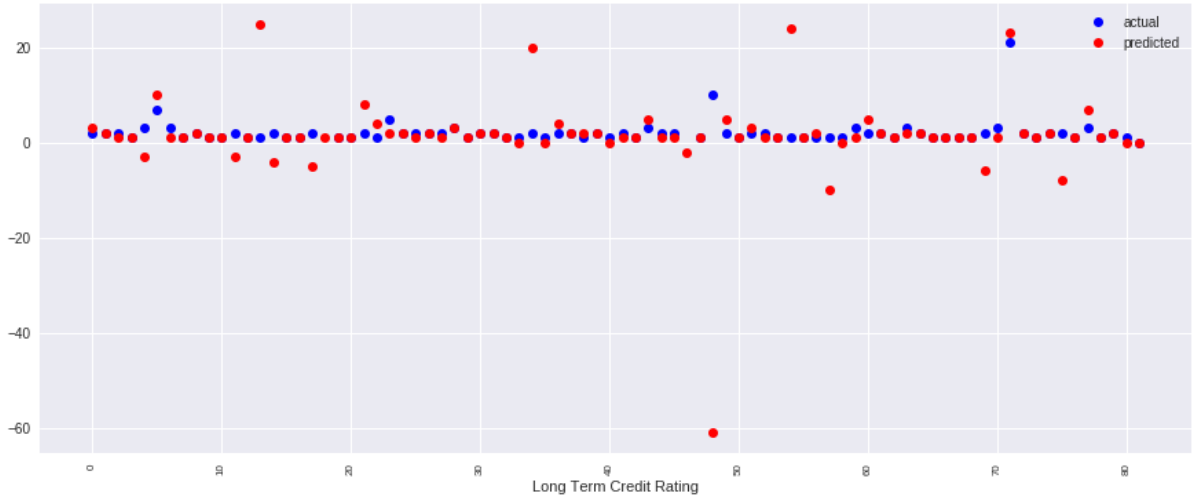
We do understand that the formula used for credit ratings is naïve and un-reliable, that is why we have trained our model to predict the financials of the company rather than credit ratings. To understand better the performance of our model we have plotted our result on different graphs for visualization from multiple perspectives rather than numerical accuracy.

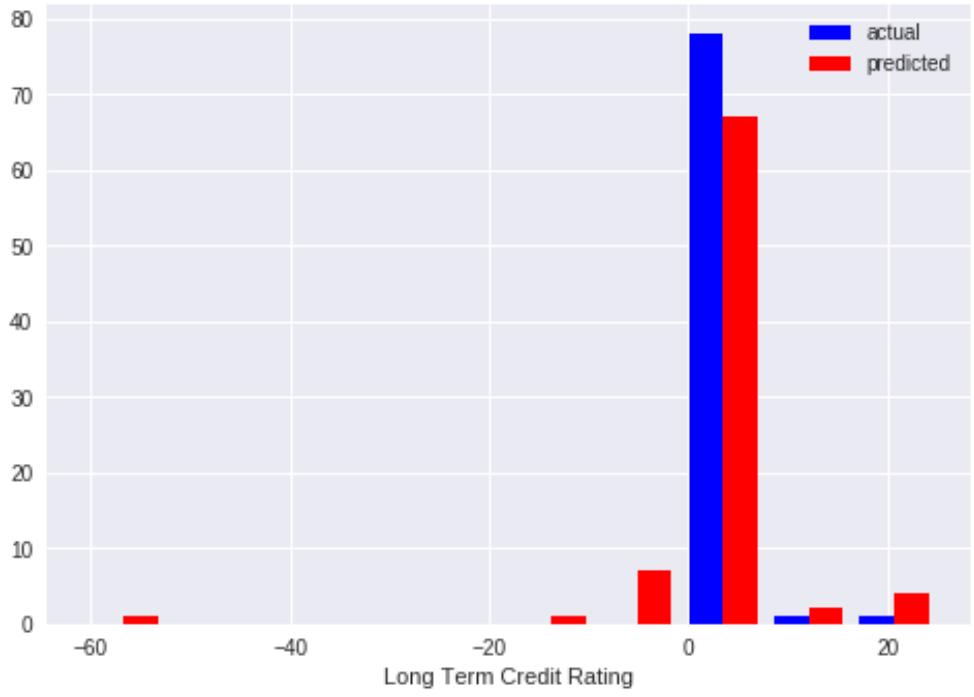
Obtained Results – *(Next Page)*

Short Term Credit Ratings -

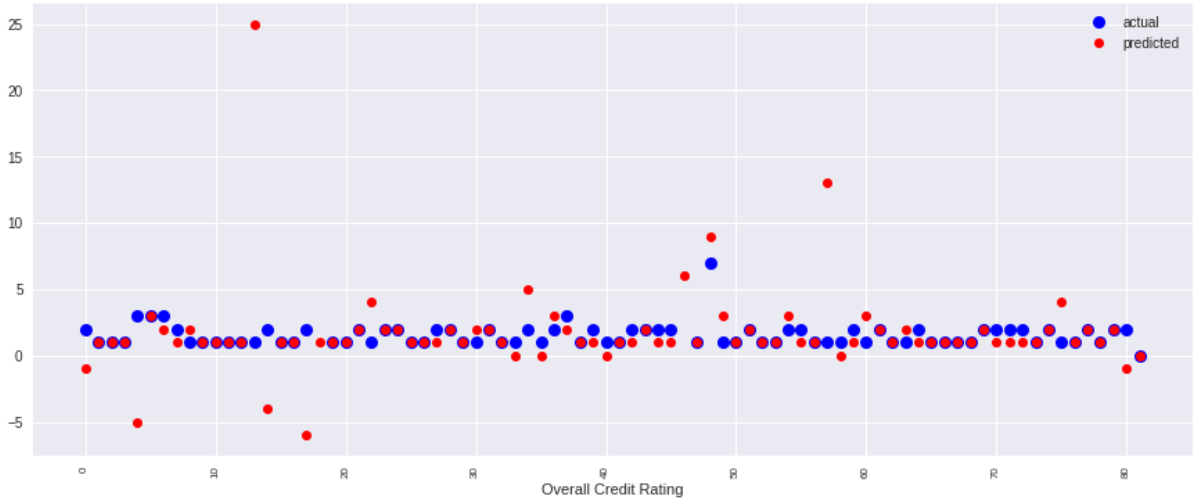


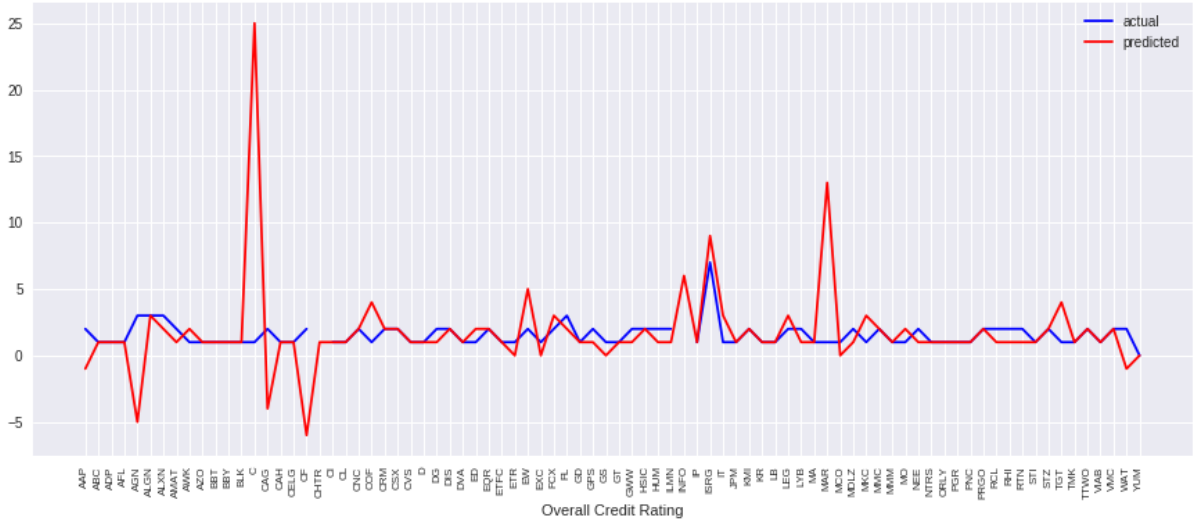
Long Term Credit Ratings –

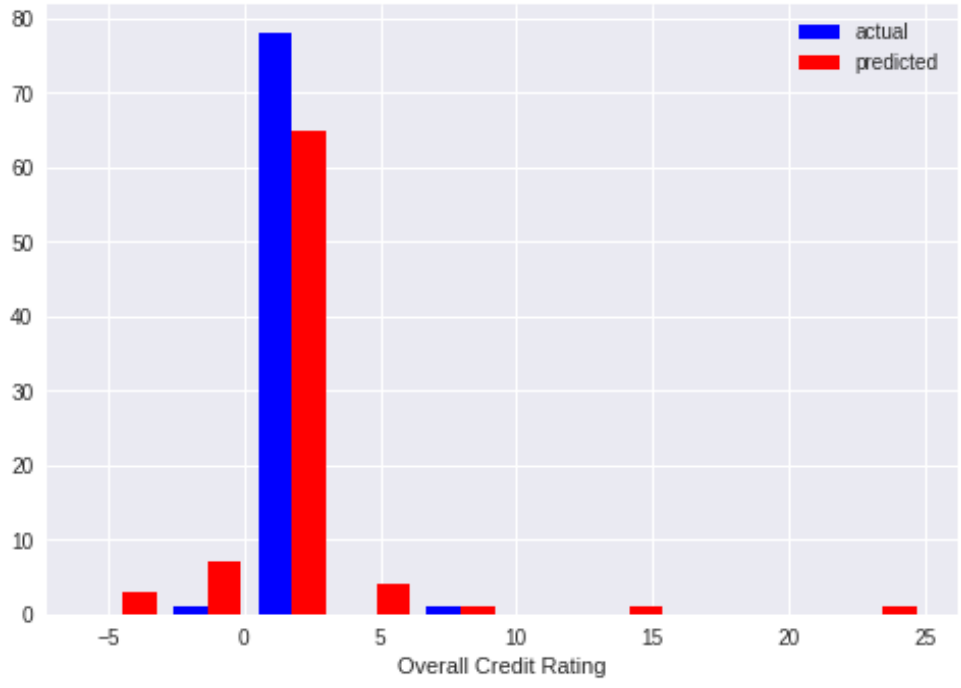




Overall Credit Ratings –







Our Analysis and Viewpoints – Though the model performed according to our expectations in case of long-term and overall credit ratings, the results for short-term credit ratings were very far from expected. This is probably due to our model design which make its predictions based on past three years data, and thus fails to model out the short-term trends.

The following can be done to improve the predictions of the network –

1. Hyper-parameter tuning
2. Different models for different credit ratings
3. Modelling out the exact number of time-series steps for each type of model.

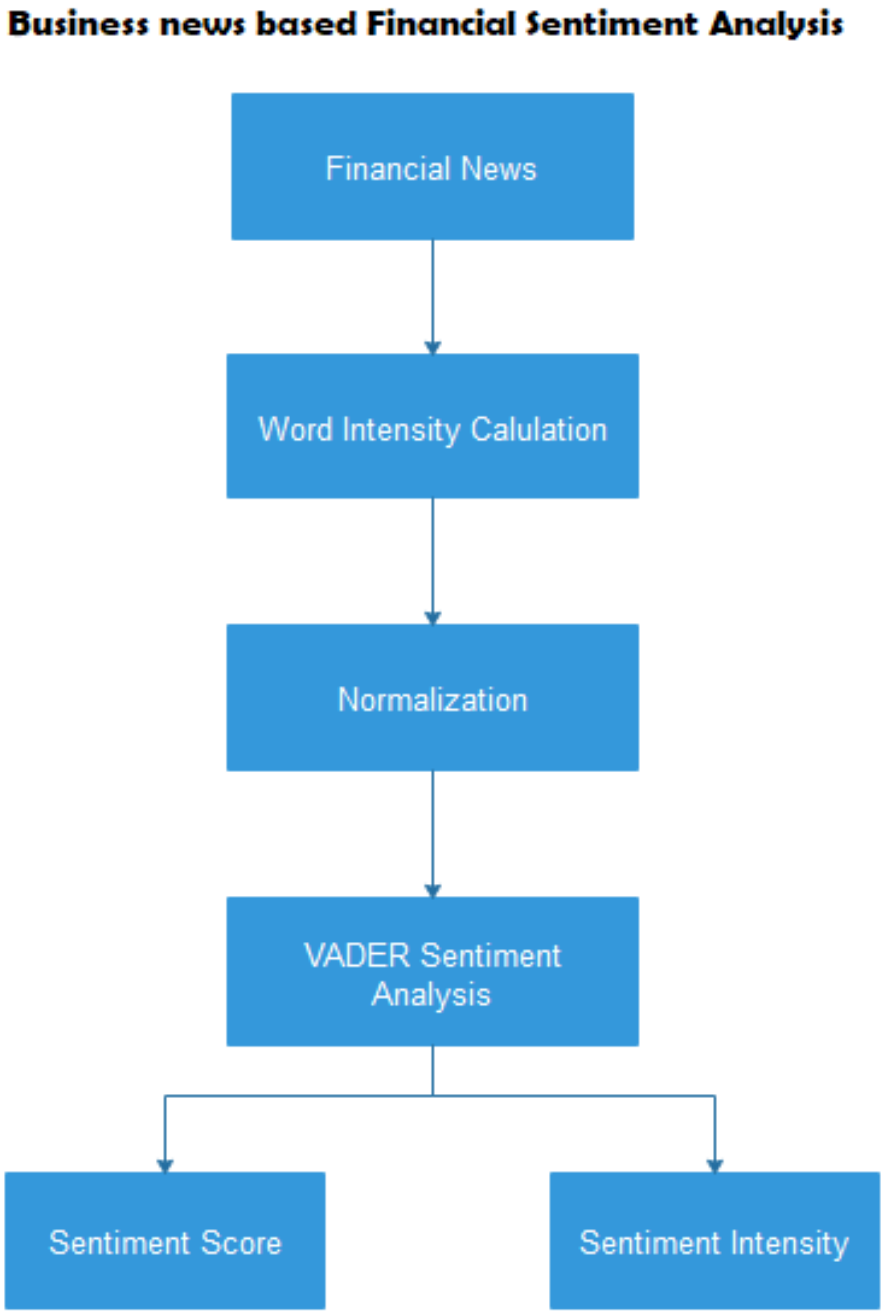
Our next approach – To address the proposed improvements using hit-and-trial manual approach will make the task to sorting out the best fit models for each type of predictions very tedious. Therefore, we propose Genetic Algorithm based approach to create the best models for each type of predictions.

The proposed genetic algorithm-based approach will consist of –

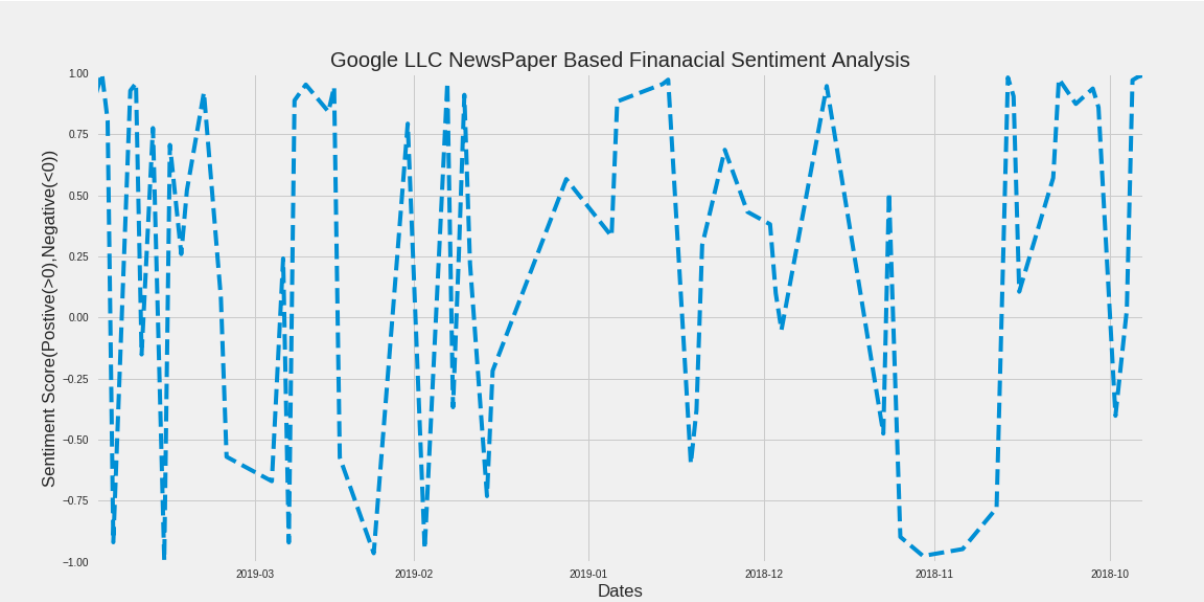
1. Population feature consisting of – no. of layers, no. of nodes in each layer, activation function and loss metrics function.
2. Multiple parallel selection strategy.
3. Crossover and mutation strategy with customized accuracy-based gradient-descent approach for faster result.

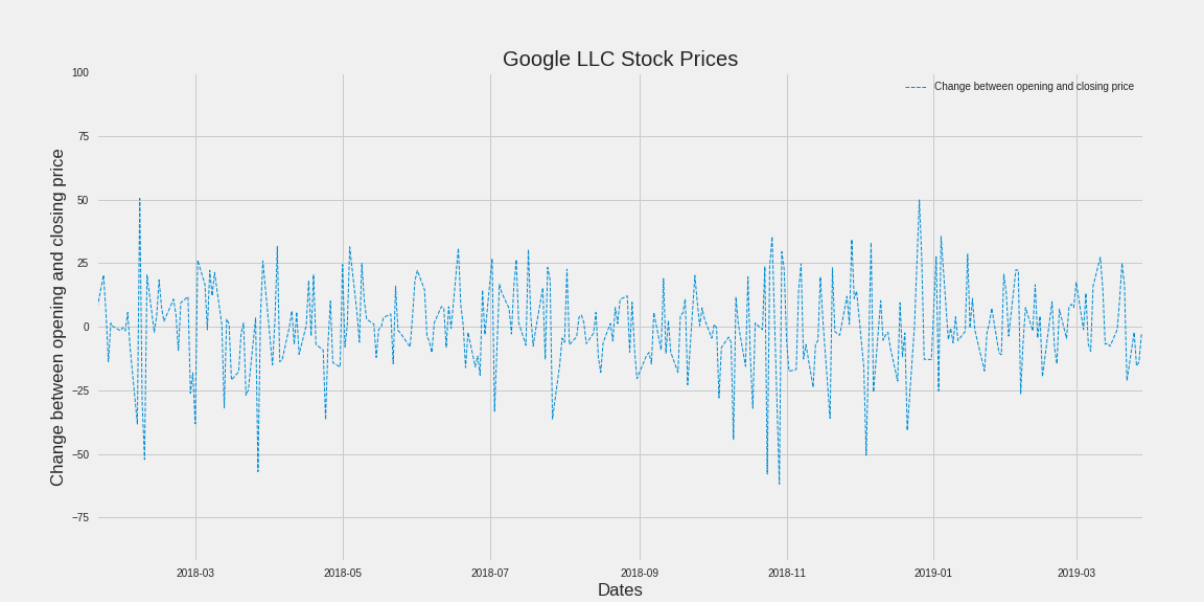
- **Business news based Financial Sentiment Analysis**

We have built a model to predict the financial sentiment from the business news and articles. The developed prototype crawls the financial news of the company of our choice from Business Times website using already developed python libraries (urllib and bs4). The intensity of each word in the obtained text is summed up followed by normalization and then we have used VADER sentiment analysis to obtain the final sentiment score. VADER (Valence Aware Dictionary for Sentiment Reasoning) is a pre-build sentiment analysis model included in the python NLTK package. It can give both positive/negative polarity of sentiments as well as the strength of the emotion (intensity) of a text.



Obtained Results –





Our Analysis and Viewpoints –

The model performed according to expectations. The plotted graph of sentiment scores and stock price variances when analysed carefully within the same range gave a good result of correlation between them.

However, to further improve our results we could use –

1. An improved lexicon of numerous financial words on top of VADER sentiment analysis lexicon.
2. A customized financial sentiment analysis model with rich training data.

2. Project Architecture –

- Primary programming language – Python

- Python libraries used –
    - keras: to create the bond credit rating RNN model
    - numpy and pandas: for data pre-processing, file handling and dataset management
    - matplotlib:
    - nltk: to create the financial sentiment analysis model
    - urllib and bs4: to crawl and filter news articles from financial news websites
    - yahoo-finance: to work with historical stock prices
    - matplotlib: to generate insight graphs of our results and analysis

- Development IDEs used –
    - Google Colab – as our primary development tool.
    - Jupyter Notebook
    - Spyder

- Machine learning models used –
    - Multivariate parallel LSTM RNN model

Project code, data results also available on –

1. https://git.ng.bluemix.net/vsharma4_be16/bondai
2. https://github.com/thevipulsharma/bondai