Bond Liquidity Prediction

-Final Project

A.Aravind Reddy(201301038)
G.Dilip Reddy(201330027)

Task Description :

A corporate bond is a debt instrument issued by companies to raise money for business operations. After a bond has been issued, it can be traded amongst investors, often via banks with dealers acting as an intermediary. A liquid bond is one which trades more frequently and in higher volumes. Knowing how liquid a particular entity would be on the market is very important.

We were given historical prices/volumes for a 3 month period, and our objective was to predict the total buy and sell volume for each bond over the 3 days immediately after the aforementioned 3 month period.

Input Data Format : There would be two input datasets for this challenge:

1) **Bond characteristics data :** This dataset contains static characteristics for the bonds being traded in the market. This would be in the following format:

```
isin,issuer,issue Date,market,amtIssued,amtOutstanding,collateralType,coupon,couponFrequency,couponType,industryGroup,industrySector,industrySubgroup,maturity,maturityType,securityType,paymentRank,144aFlag,ratingAgency1Rating,ratingAgency1Watch,ratingAgency1EffectiveDate,ratingAgency2Rating,ratingAgency2Watch,ratingAgency2EffectiveDate
```

2) **Historical price and volume data :** This dataset contains the price and volume at which bonds were traded at various points of time. This would be in the following format:

```
isin,time,price,side,volume,timeofday,date
```

Output Data Format : For each bond mentioned in the bond characteristics dataset, we need to predict the total buy and sell volume for the 3 days following immediately after end of the training dataset. The format is as follows:

```
Isin,buyvolume,sellvolume
```

Evaluation Criteria :

Solutions are evaluated based on the Numerical values of the volumes predicted. This criteria is selected because of the trivial reason as the output should be as nearer to the original values.

Initial Idea:

If we observe carefully the data here is a Time Series Data. We were given the data of previous 90 days and were expected to predict the next 3 days. Generally for such time series data prediction we use Recurrent Neural Networks. A **recurrent neural network** (**RNN**) is a class of Artificial Neural Network, where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. Unlike feedforward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs. This makes them applicable to tasks such as unsegmented connected handwriting recognition or speech recognition.

Numerous researchers now use a deep learning RNN called the long short-term memory (LSTM) network. It is a deep learning system that unlike traditional RNNs doesn't have the vanishing gradient problem . LSTM is normally augmented by recurrent gates called forget gates. LSTM RNNs prevent back propagated errors from vanishing or exploding. Instead errors can flow backwards through unlimited numbers of virtual layers in LSTM RNNs unfolded in space. That is, LSTM can learn "Very Deep Learning" tasks that require memories of events that happened thousands or even millions of discrete time steps ago. Problem-specific LSTM-like topologies can be evolved. LSTM works even when there are long delays, and it can handle signals that have a mix of low and high frequency components.

So we used LSTM for this project and the details about the implementation is provided below. Implementation is presented as code.

Implementation :

We first read various articles related to the LSTM(The references are mentioned below). Then we came to a conclusion to use the volume values to predict the other volumes. We took a lookup window of size 10 and set its output as the 11th value. That means we indirectly created a dataset with 10 days volume values as the features and the 11th day volume as the output.

Then we trained our LSTM model (we used sequential model) with this data with 67% being the training data and 33% being the testing data. Then we predicted the volumes of the next 3 days using the previous 10 values for those values. We have done the same for each bond and this is done for both buy values and sell values separately.