**Paper Summaries – Sentiment analysis – Price Prediction**

1. **Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis** - *Dibakar Raj Pant, Prasanga Neupane, Anuj Poudel, Anup Kumar Pokhrel, Bishnu Kumar Lama*

* Uses Bag-Of-Words and Word2Vec as methods for feature extraction
* Five different algorithms Naïve Bayes, Bernouli Naïve Bayes, Multinomial Naïve Bayes, Linear Support Vector Classifier and Random Forest are used to train the features extracted from Bag-Of-Words and Word2Vec. Then a voting classifier is used to choose the maximum vote.
* For the time series prediction, an RNN with LSTM and GRU cells is used.

Diagram

Description automatically generated

* Pearson correlation between sentiment and price the day after. “Further, Pearson correlation test is performed with the sentiment score and corresponding price of the next day.” Maybe less time or more time?
* for the lower number of dataset and at sentence level sentiment classification Word2Vector does not perform well so Bag-of-words is preferred.
* Table

  Description automatically generated
* The price prediction accuracy for RNN model is found to be 77.62%.
* The major contribution of this work is a sentiment analyser which can distinguish between the positive and negative tweets of Bitcoin over the Twitter with the accuracy of 81.39%
* Table

  Description automatically generated

1. **Sentiment Analysis of Twitter Data for Predicting Stock Market Movements** - *Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi*

* The introduction states different reasons and sources why sentiment analysis on twitter can extract important information.
* We have shown that a strong correlation exists between twitter sentiments and the next day stock prices in the results section.
* Tokenization, stopword removal and regex replacement is performed before checking the sentiment of tweets
* Model training uses random forest algorithm on features extracted by word2vec
* The total positive, negative and neutral emotions in tweets in a 3 day period are calculated successively which are used as features for the classifier model and the output is the labeled next day value of stock 0 or 1 ??
* 80% train, 20% test from data set
* Sentiment Analyser

1. 90% train, 10% test
2. model trained with word2vec representations is picked to classify the nonhuman annotated tweets because of its promising accuracy for large datasets and the sustainability in word meaning.

* Diagram

  Description automatically generated
* Table

  Description automatically generated

1. **Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning - Franco Valencia , Alfonso Gómez-Espinosa and Benjamín Valdés-Aguirre**

* This paper compares, Neural Networks MLP (Feedforward), SVMs and Random Forest algorithms over Bitcoin and Ethereum.
* For training, the data is split 70% for training and 30% for test
* In setup, features where standardized by removing the mean and scaling to unit variance.
* For the MLP model, we selected a hyperbolic tangent activation function because of its popularity and good performance. The default solver, “adam”, a stochastic grading-based optimizer, was utilized with a L2 penalty of 0.0002.
* The models had an input layer, a single hidden layer and an output layer. The amount of neurons for the input layer was equal to the size of the feature vector.
* The SVM kernel, used a Gaussian radial basis function K(x; y) = exp(−1/σ2(x − y)2) because of its popularity for SVM classification problems.
* In the Random Forest model the only parameter tweaked was the number of trees, which was raised from its 10 default up to 1000.
* Accuracy, Recall, Precision and f1 were calculated
* MLP was the best performing model for Bitcoin. Having an accuracy of over 0.72 and a precision of 0.76, this model is better than random by a large margin.
* Both SVM and RF also managed to beat random when using Market data.
* Twitter data by itself could not be used to predict the market movement in any model, and its inclusion appeared to worsen the performance of the SVM and RF models. However it improved the precision in the MLP model slightly.
* Table

  Description automatically generated
* Table

  Description automatically generated
* Table

  Description automatically generated
* Table

  Description automatically generated
* MLP as can be seen is the best model for most of the cases apart from the case for Litecoin.
* These results also make it possible to observe how the usage of exclusively Twitter data can be used by itself to predict the Ripple and the Litecoin markets

1. **Sentiment Analysis of Stock Blog Network Communities for Prediction of Stock Price Trends -** Sandeep Ranjan, Inderpal Singh, Sonu Dua, Sumesh Sood

* For sentiment analysis, this study used the textblob & natural language toolkit of Python to perform the sentiment analysis of the datasets. The positive sentiment post is assigned +1 or +2 weight as per the intensity of the sentiments ; similarly, negative posts are awarded -1 or -2 weight as shown in the Table 3. Sentiment analysis was not done on individual posts, which were not a part of any community. Words containing sentiments such as buy, sell, hold, moving up, scrap, debt, shoot, plunges, etc. were added to the dictionary for grasping the sentiment of the data

1. **Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis -** Jethin Abraham, Daniel Higdon, John Nelson, Juan Ibarra

* This uses hashtags for data collection of tweets, it even collects the number of retweets and number of times it was ‘favourited’
* VADER was used for sentiment analysis
* Around 50% of tweets actually contain sentiment
* Tweet volume + google trends is a better indicator
* Future work should determine if these results continue to hold in varying pricing environments. Additionally, more complex models, and not just linear ones like we used, could be fit using Google Trends and tweet volumes as inputs to see if results could be improved further.

1. **The predictive power of public Twitter sentiment for forecasting cryptocurrency prices -** Olivier Kraaijeveld, Johannes De Smedt

* Uses VADER and added words to lexicon
* Gilbert and Hutto (2014) show that VADER can outperform both human annotators and most classifier benchmarks.
* To explore whether certain factors are driving prices, this work looks at the bivariate Granger-causality test. It is important to note that Granger-causality does not establish actual causality but rather finds a statistically significant pattern in lagged values of X and Y.
* The scores are also consistently positively skewed with a mean polarity of 0.33. This is consistent with the results of Kennedy and Inkpen (2006), who observe that lexicon-based approaches generally have a positive bias, which can be attributed to a human tendency to prefer positive language
* First 12 hours of day = bullish trend, Last 12 hours =bearish trend (in terms of sentiment)
* It was observed that no real relation is found between sentiment and price but sometimes there were instances of good prediction in the hourly interval
* With regard to the applied techniques, the primary limitation comes from using a lexicon-based approach for the sentiment analysis. Lexicon-based approaches are unsupervised and obtain lower accuracies compared to supervised techniques, due to their static rule-based nature.
* With regard to the applied techniques, the primary limitation comes from using a lexicon-based approach for the sentiment analysis. Lexicon-based approaches are unsupervised and obtain lower accuracies compared to supervised techniques, due to their static rule-based nature
* Another topic for future research would be to apply a supervised machine learning-based or hybrid approach.