İstanbul Bilgi Üniversitesi

# NATURAL LANGUAGE PROCESSING
# COIN PREDICT SYSTEM
# PROJECT

Işılsu Topanoğlu - 117200076
Karya Tüfekçi - 117200082
Muhammed Berke Tevik - 117200043
Barış Tonbul - 117200054

# COIN PREDICT SYSTEM

There are overmuch sources of information and it takes time to find the specific information desired. Finding and extracting the content that we wanted or wanted from the expanding concentration of information and text has become difficult today, especially group topic titles or specific topic titles. The use of text classification, which is actively used today, is more suitable for the purpose of the project in order to facilitate the work and life of the users and to ensure that they reach the desired content easily. This classification has facilitated the separation and understanding of text types.

In this world, there are many system occurs every day, every second. All in this system, most common, most popular, most useful and also must trustable system is the blockchain system. Lack of trustworthy environment causes trusting issue. People wants to rely on something that make them confidential and they apply to blockchain system which has no third partition to keep their money safe and appreciation of their money. Becoming more popular coin world produce multiple types of coin. Who doesn't want to take risks in this environment, they imburse their money to Bitcoin. This application provides to increase in value of money and ensures secure. With people can predict the value of coin and take an early action. System shows that bitcoin worths increase or decrease according to tweets on Twitter. Used algorithm finds the peoples thought and people mentions about bitcoin with highest possible profit rate. Trading strategy running for last 2 minutes.

System evaluates positiveness and negativeness of mentions of bitcoin and with using these values, it also evaluates a final rate which is between -1 and 1. The closeness to 1 means bitcoin is better at this point. Besides, if this rate is bigger than a specific value, it buys bitcoin instead of user.

First of all, the system gathers live data that related with bitcoin from Twitter API. It takes the last 200 datas which is, approximately, last 2 minutes. To gather the data system needs key to access API. And all these keys assigned to a variable to more readability. Putted all these values into an array, end of these application, system provides to connect Twitter API. For continuous process, a function created and every process that contains tweet, will be in this function. And other empty array takes all tweets in here and some function applied. As a result, these function eliminates redundant, non-functional things and functions provides more readability for tweets. Such as, to remove retweets, detect positive or negative emojis, eliminating same tweets, etc. Preprocess function organize tweets, make them more understandable for machines suchlike removing hashtag, remove RT (retweet) to not to disturb balance, degrade spaces to one space, return lower case, change URL link to "URL". These steps are important because clear data is more efficient. System controls the characters (it is English char. or not) for the recognition. After all of this processes, system has valid tweets and they put into final tweets array.

After all, system can start to trade. Trading algorithm starts with a buy, then waits for the pullbacks. It has 2 possible pullback buy points. One of them is where price equals to MA20, another one is where price equals to MA40. We got different stop loss points and trailing stop loss ratios for each pullback and as expected we got different take profit points. Algorithm stops when sell is occurred and waits for the next buy signal from the data that we fetch to start another trading process.

The data frame is created by using final tweets and by using vader-lexicon from nltk library, system begin to semantic analyze. On data frame, there are columns that contains scores. Score is compound of positive, negative, and natural values. According to compound score tweet can be analyzed. Tweets thrown into lines to split word by word. System took the words as a stem format by using SnowballStemmer from nltk library. Shows data frame of words and contains twenty words. Frequency of words is found and most used words constitute plot.

There are some techniques that are used, such as Sentiment analysis, Text Classification Also for making it real, these are the technologies that used Twitter API, Binance API, Technical analysis data from TAAPI.io, python.