

**Computer Science Department**

**Senior Project Proposal**

|  |  |
| --- | --- |
| **Proposer:** | Diyar Akhmetov, Miuagul Dybyspayeva, CS students, SST |
| **Topic:** | Financial Prediction for Bitcoin using Financial Data and News |
| **Team Size:** | 2 students |
| **Description** | |
| Discovering a relationship between quantitative and qualitative economic data is a hot topic in financial world. Usually, finding that relationship is done by humans (e.g. traders, brokers, market analysts). However, nowadays advancing technologies allow us a computational data-driven approach to stock market analytics. This project is an NLP (natural language processing) based decision support system in market analysis. The project will involve application of machine learning techniques to both structured time-series data, which otherwise called financial data and unstructured textual data that also is called as financial news on websites, and twitter. Based on the data given, the system will semantically analyze whether to sell, buy, or hold the stock. | |
| **Objectives** | |
| Based on the existing datasets implement vectorization methods (e.g. Word2Vec and Seq2seq) for textual data for numerical and textual data, and choose the most efficient model in stock price prediction.  Firstly, we will look at the existing state-of-the-art models, specifically at the error rate and error magnitude. Later, we will create neural network that can effectively predict the future price based on previous financial only data. Our initial aim is to implement a model that can show good performance on predicting financial data only. As soon as we gain the minimum error, we will work exclusively with the text news. We will implement text preprocessing algorithm that will extract a small description from big text, articles. Later on we will incorporate that into text processing model to identify the sentiment of the text and apply it on the already running neural network. Our final goal is to identify if processing news on existing financial data boosts the prediction accuracy, if so by what extent. | |
| **Deliverables:** | |
| **SP1**: review the state-of-the-art models, implement the models that combine existing financial as well as textual information, work with text processing algorithms to analyze and fit the neural networks.  **SP2**: in case SP1 works well, we will publish research paper or more developed final product (website/application). If not, improving the existing implementation. | |
| **Tasks to be Accomplished:** | |
| * Review the existing state-of-the-art models from literature * Make a literature review * Gather financial data * Implement the proposed models in Tensorflow * Review unused vectorization methods, identify which one suits our purpose for text vectorization the best * Develop new model architectures * Gather textual data on financial situation of bitcoin * Implement text processing model that will have a preprocessing algorithm to extract a small abstract and identify the sentiment of the text * Compare models between each other as well as with those of state-of-art * Create user friendly landing page/appication for future customers Report the work done | |
| **Hardware and Software Requirements** | |
| Software: Python2, NumPy, Pandas, Tensorflow  Hardware: CPU/GPU(GoogleColab) | |
| **References** | |
| 1. <https://arxiv.org/pdf/1707.09448.pdf> 2. <https://arxiv.org/pdf/1307.5336.pdf> 3. <https://ieeexplore.ieee.org/document/7379415/> 4. <https://dl.acm.org/citation.cfm?id=1462204> 5. <https://www.sciencedirect.com/science/article/pii/S0167923612000875> | |

#### **Term Schedul**e

|  |  | To be always filled by adviser during meeting with students | | |
| --- | --- | --- | --- | --- |
| Week Number | Task Assignment | Date Completed | Comments | Signature |
| 2  (~ 24 Sept ) | * + Literature Review(have at least 10 articles analyzed)(Student 1 & 2)   + Implement text summarization introduced by seq2seq model for text preprocessing(Student 2) |  |  |  |
| 4 (~Oct 8) | * + Improve main architecture by adding convolution and window sliding approach for only financial data prediction (Student 1)   + Compare and analyze with the state of the art(financial data) (Student 1 & 2) |  |  |  |
| 6  (~Oct 22) | * + Adding sophisticated multilayer architecture for text processing and financial data (Student 1 & 2)   + Compare and analyze with the state of the art. (Student 1 & 2) |  |  |  |
| 8  (~Nov 5) | Intermediary Presentation (achievements, problems, review etc)   * First results: see if applying text processing can gain better results. If not, identify why it didn’t work and improve the architecture. (Student 1 & 2) |  |  |  |
| 12  (~Nov 19) | * Last remarks: find out the ways we can improve our architecture * Create user friendly landing page for future customers. (Student 1 & 2) |  |  |  |
| 14  (~Nov 27) | Presentation accomplishments, encountered problems and proposed solutions. This is not a Technical presentation but rather advancement check, Interim Report (end Â of term paper)  End of first term |  |  |  |