**Forecasting Short-Term Stock Prices Using Machine Learning Models**

**Andrés Camilo Viloria García**  **Diana Sayuri Ortiz Montes**

aviloriagarcia00@mylangara.ca dortizmontes00@mylangara.ca

*Data Analytics Data Analytics*

*Langara College Langara College*

Vancouver, BC Vancouver, BC

***Introduction***

Predicting stock prices has become an interesting topic in the research community, as machine learning techniques are being used to tackle this challenging task. Financial markets are known for their complexity and volatility, posing ongoing obstacles that require creative solutions. The high complexity and volatility of financial markets pose significant challenges to accurate stock price prediction. Existing machine learning models with modest accuracy rates of approximately 50%, indicating the need for further advancements in precision, data handling, forecasting, and prediction. This final project collects information about the data quality used in earlier stages of the project. It evaluates the performance of the model using specific measurements and adjusts certain settings to improve the models. The results from different models are compared, and all the technical details are gathered to be included in a future research paper for publishing.

***Scope***

We will continue with the development of a methodology to accurately forecast stock prices in the short-term using machine learning. This project phase (Phase 5) aims to achieve the following objectives:

* Model comparison: Complete the table by filling out the metrics for all models (XGboost, random forest, and MLP), not just for individual stocks but also for portfolio values.
* Feature Engineering: It includes integrations of the codes from all the phases of the project to select, create, and manipulate features from the available data to enhance the predictive power of the models. Feature engineering includes tasks such as data cleaning, encoding categorical variables and extracting relevant information or insight from the data.
* Hyperparameter Tuning: choose and assess variations in the values of specific hyperparameters to achieve improved metric results.
* Comparing Predictions: A comparison between the actual values and the predicted values was conducted by generating Python graphs to visually represent the disparities.

The graph presented below visualizes the fulfillment of objectives along with their corresponding deliverables.

**A diagram of a company

Description automatically generated**

**Project Deliverables**

The following deliverables are the contributions we aim to achieve during the duration of our project execution.

**Documentation**

* [Data Sources](https://www.dropbox.com/scl/fo/7evgs98hnzjs81b3mwxbo/h?rlkey=95t6yufyvufwdzy7jzxwl08g9&dl=0)
* [Data Quality deep dive](https://www.dropbox.com/scl/fo/f8n0g3qn6ulwf4d1g0a5p/h?rlkey=zaw0oot72fp68er1ghw551sj7&dl=0)
* [Hyper parameter tuning for XGBoost](https://www.dropbox.com/scl/fo/i4vau9ts00ui7djd8uzu6/h?rlkey=2carg1d9f4wo0vbnrs4cy041a&dl=0)
* [Graph creation for all models and hyperparameters](https://www.dropbox.com/scl/fo/dznk3jhz0d7is32xrxf0n/h?rlkey=1f728b2yw9fqs1qaqfwait1nh&dl=0)
* [Meeting minutes](https://www.dropbox.com/scl/fo/au88ecpk120nfq7dt04hh/h?rlkey=yydpwd2jcao8vvpk7kwm3pqqm&dl=0)

***Milestones***

* Understanding of previous semester methodologies (Sep 30)
* Model comparison initial results (Oct 29)
  + Data collection
  + Model execution
  + Documentation Report
* Feature Engineering ( 29 Oct- 3 Nov)
  + Improving organization of code for better readability and understanding.
  + Documentation of variables description, statistics (distribution and range of values) and further analysis of missing values.
  + Report on Data quality and feature engineering.
* Hyperparameter tuning for **XGBoost** (6 Nov- 17 Nov)
  + Research on best practices for tuning **XGBoost**
  + Create a tab on excel sheet for easy model comparison
  + Run codes and document findings
* Predicted vs Actual csv’s and Graphs (17 Nov- 3 Dec)
  + Edit codes to obtain csv’s with the predicted vs actual data
  + Create graphs for each model and hyperparameters
  + Obtain csv’s for all models and hyperparameters
* Final documentation deliverables (3 Dec – 16 Dec)
  + Data quality
  + MPE deep dive insights
  + Hyperparameter tuning XGBoost
  + Predicted vs Actual csv’s and Graphs
  + Appendices