**Professional Certificate in Machine Learning and Artificial Intelligence**

Module 25: Model card for Portfolio Project

**Model Description**

* The inputs to the model are daily price and volume information for Apple Inc. stock (AAPL) in USD from 12/12/1980 to 24/03/2022.
* The inputs include: Open (price from the first transaction of a trading day), High (maximum price in a trading day), Low (minimum price in a trading day), Close (price from the last transaction of a trading day) and Adj. Close (closing price adjusted to reflect the value after accounting for any corporate actions).
* The output of the model is the predicted next day closing price for Apple Inc. stock. Via Bayesian optimisation, we also output the optimal parameters for a LightGBM regressor model.
* The model architecture implemented includes a Light Gradient-Boosting Machine with a boosting type of a traditional Gradient Boosting Decision tree. We built a LightGBM regressor to understand the influence of technical indicators (exponential moving average, simple moving average, relative strength index and moving average convergence/divergence) on next day’s close price.
* Using Scikit-optimize, we utilised Bayesian optimisation to identify the optimal parameters of the LightGBM regressor model.

**Performance**

* The data analysed includes: simple moving average (SMA for 5, 10, 15, 30 days), exponential moving average (EMA for 9 days), relative strength index (RSI), moving average convergence/divergence (MACD) and MACD signal. The target variable is next day’s closing price in USD. We analysed data from a timeframe of 04/01/2010 to 24/03/2022.
* We provided a baseline performance for a generic LightGBM regressor. The training root mean squared error (RMSE) is 0.817 and the test RMSE is 1.52.
* Through Bayesian optimisation, we identified the optimal parameters for the LightGBM model and re-ran the regression. The training RMSE decreased to 0.507 and the test RMSE fell to 1.41.

**Limitations**

* To develop the model, we could extend our analysis to include stock data for companies beyond Apple Inc.. We could potentially create a LightGBM model to predict closing prices for all companies listed on the NASDAQ. This would give us a more holistic understanding of the impact of different technical indicators on stock price.
* There are a variety of indicators that contribute to the stock price of a company. We could incorporate macroeconomic variables into our analysis such as interest rates, inflation, unemployment, export volume, and government spending. In addition, we could asses changes in fundamental factors of a company that contribute to its stock price such as P/E, debt to EBITDA, and cash flow to revenue.
* Another limitation of the model involves not creating a proxy to account for the impact of one-off events on stock price. For instance, the Covid pandemic had a detrimental impact on the stock market and confidence in the economy. To improve on the model, we could include a feature that represents the likelihood of different catastrophic events occurring.

**Trade-offs**

* The model only explores a LightGBM model with a traditional Gradient Boosting Decision Tree boosting type, which may limit model performance. During the Bayesian optimisation phase, we could also construct models with a Dropouts meet Multiple Additive Regression Trees (dart) or Random Forest (rf) boosting type.
* Another trade-off involves the use of RMSE as the only performance metric. To identify the optimal parameters of the LightGBM model, we could also compare accuracy, precision and recall, F1 score and mean absolute error of the test data.
* In order to test the performance of the Bayesian optimisation, we could provide a baseline using random search to tune the hyperparameters of the LightGBM model. This helps us understand the efficacy of implementing Bayesian optimisation.