# IN3062 Introduction To AI – Group 11 Report

Khadija Nawaz, Bendeguz Kudor, Kia Poore, Martins Pigoznis

Github : <https://github.com/Mx-KN/Group11_IntroAI_Coursework>

**Introduction:**

In our society now, we spend a lot of time looking at certain aspects of economics which are now moving too quickly for humans to follow. Using the dataset on Kaggle (<https://www.kaggle.com/datasets/benjaminpo/s-and-p-500-with-dividends-and-splits-daily-updated>), we can look at how certain aspects are constantly changing. The data shows S&P 500, which is a stock market index tracking the performance of 500 of the largest companies listed on stock exchanges in the United States. It also shows dividends, exchange-traded fund (ETF), Crypto and the Foreign Exchange Market (FX).

Especially with the way currency and people interact with it can have a major influence on the stock markets and foreign exchanges. As such, this is one of the few departments in which humans are falling behind as artificial intelligence can be used to make the data more manageable and can also be used to predict future incidences. These incidences relate closely to events like the financial crisis in 2008 as well as the Wall Street Crash in 1929.

The data we used has 8 sections to it: Date, Open, Close, High, Low, Volume, Dividends and Stock Splits. The column ‘Open’ refers to the starting price of a stock at the beginning of the trading day. ‘Close’ therefore refers to the price of the stock at the end of the trading day. The label ‘High’ references the highest price a stock reaches whereas ‘Low’ references the lowest. The ‘Volume’ of a stock is calculated by the total number of shares of the stock traded throughout the trading day.

**Models:**

We have chosen three models to use to train and test the data. The first was a linear regression model. The linear regression model estimates the linear relationship between two variables.

The second model we chose to use was a Support Vector Machine (SVM) training model. This Support Vector Machine is used to determine boundaries between different data points.

The final model we used was a random Forest Classification.