**Assessment 3**

**STOCK MARKET PREDICTION**

**Description:**

Dataset of our project,” Stock market prediction” is extracted from NASDAQ (National Association of Securities Dealers Automated Quotations) website. Our dataset consists of information about 10 different companies’ stocks. The data is of last ten years. We can extract the data from this website till the current date. Dataset consists of Volume column which explains the number of shares that changed hands during a given day. Open, close columns explain starting and ending period of price at which a stock trades upon the exchange on a given day. High, low columns in the context of stocks means the highest and lowest prices of a stock recorded on a given day. The column company indicates the name of the organization. Dataset consists of 24000 rows and 7 columns.

Our data has the columns Open, High, low, close has the values which can fit in to the models to give expected result. We can segregate the data into 75 and 25 ratios to train and test our model. We think our selected data is appropriate for our project as we have the values that can be useful to predict the crashes by observing the results.

**Ethical principles:**

For any project, before proceeding, the ethical principles or guidelines are to be followed to make the project successful in ethical manner.

* **Consen**t:

The dataset we choose is open data which is openly available and easily accessible by anyone. It doesn’t consist any personal information, so it doesn’t affect anyone’s privacy.

* **Clarity:**

Data will be used to analyze ups and downs in given stocks and forecast the crashes.

* **Consistency:**

Our data is openly accessible by public therefore, there is no chance for consistency issues.

* **Control:**

The data we extracted is from NASDQ website, all the rights are acquired by the organization and they have the complete control to access the data. If any of the data in our dataset will be restricted in the future, we can always remove it from our dataset.

* **Consequences:**

As our data doesn’t involve any individual’s personal information, it is unlikely less harmful to the public. The data will not be shared to any third-party vendors without user consent.

**Future data needs and potential challenges:**

The nlp package usage with the text, which has same words but different meaning.

**References**:

1. Yordanov, V. (2019, August 13). Introduction to Natural Language Processing for Text. Retrieved from <https://towardsdatascience.com/introduction-to-natural-language-processing-for-text-df845750fb63>.
2. Historical Data. (n.d.). Retrieved from <https://www.nasdaq.com/market-activity/quotes/historical>.