Logo

Description automatically generated with low confidence

**Fundamentals of**

**Software Engineering**

**Assignment #2**

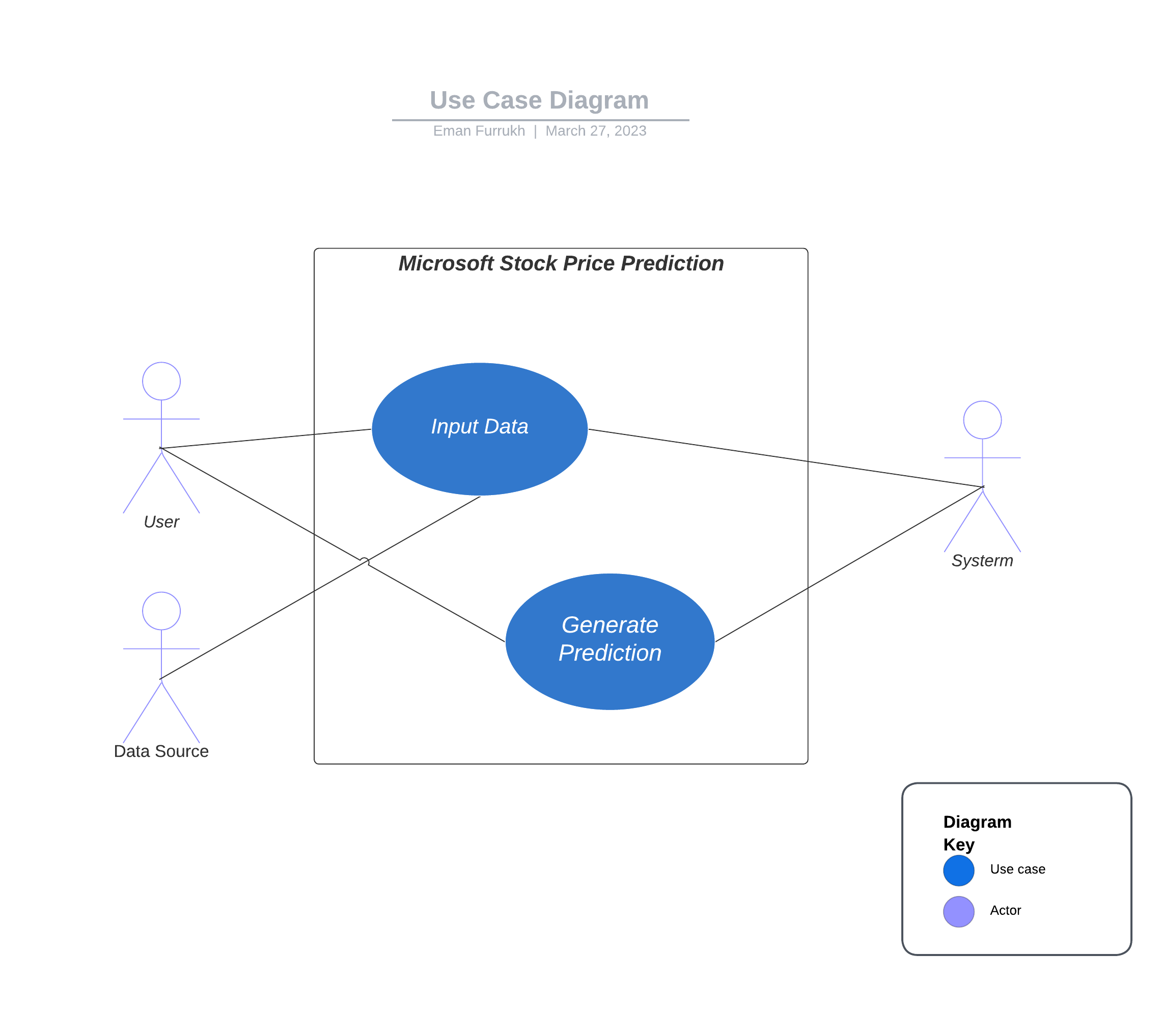
**Iteration 1: Execution Phase (OUTLINING SPRINT)**

**Name : Eman Furrukh**

**Roll No : 21i-1726**

**Section : BSDS-U**

Diagram

Description automatically generated**1.Refine the Use Case diagram of your project( if need), design the SSD’s of your project according to Use Case diagram**

**2.Write the feature-based description report of your dataset that you will be used in SE based ML Project.**

The dataset consists of the following features:

* Date: The date on which the stock price was recorded.
* Open: The opening price of the stock on the given date.
* High: The highest price at which the stock traded on the given date.
* Low: The lowest price at which the stock traded on the given date.
* Close: The closing price of the stock on the given date.
* Adjusted Close: the closing price after adjustments for all applicable splits and dividend distributions.
* Volume: The number of shares of Microsoft Corporation traded on the given date.

**3.Train your ML model according to state-of-the-art settings or tune the hyper parameters for accuracy and robustness.**

Load the dataset:

* Load the preprocessed dataset into the machine learning environment.

Split the data:

* Split the dataset into a training set (70%) and a testing set (30%).

Feature scaling:

* Scale the features using standardization or normalization techniques. Hyperparameter tuning:

Model selection:

* Train different machine learning models on the training set and evaluate their performance on the testing set. Select the model with the best performance on the testing set such as Linear regression, Random Forest and so on.

Evaluate the final model:

* Evaluate the performance of the final model on the testing set and compare it to the performance of the baseline models. If the final model outperforms the baseline models, it can be deployed for stock price prediction.

Deploy the model:

* Deploy the trained model in a production environment where it can be used for real-time stock price prediction.

**4.a) Write the detail and comprehensive report of dataset description.**

1. Dataset Overview:

* The dataset consists of historical stock prices of Microsoft Corporation and spans from 1986 to 2020.

2. Data Preprocessing:

* The dataset required some preprocessing before it could be used for machine learning. Processes such as handling missing values, outlies, normalizing were used in this project.

3. Feature Description:

* This is mentioned in question 2, done above.

4. Data Visualization:

* Data is shown visualized through a graphical method.

5. Model Selection:

* We used Decision Trees as a model for this project in order to train and then test the data.

5. Conclusion:

The dataset used is a comprehensive and well-preprocessed dataset that includes various features and technical indicators to help improve the performance of the machine learning model. The dataset has no missing values or significant outliers, and the features have been normalized to ensure that they have the same scale. The dataset has been visualized using plots to help gain insights into the data. Overall, this dataset is a suitable and reliable source for training a machine learning model for Microsoft stock price prediction.