Logistic Regression

The highest accuracy score for the logistic regression model was 91% using binary outcomes with no stock data.

Before fitting the model, ‘unnamed’ columns, date columns, and rows with NaN field values were dropped. We also got rid of the Dow Jones data due to datatype issues with values showing up as strings.

The target variable was the binary class attribute that describes whether the number of deaths was higher or lower than the mean. The scikit-learn tool test\_train\_split was used to separate the data into training and testing data. Then the data was scaled using scikit-learn's StandardScaler before fitting the Logistic Regression model.

(replace table below with LR-table.png)

|  |  |
| --- | --- |
| Logistic Regression – Model Accuracy Scores | |
| Binary outcomes, without stock data | Binary outcomes, with stock data |
| 91% | 88% |

SVM

The highest accuracy score for the Support Vector Machine model was 91% using binary outcomes with stock data.

The setup was the same as that used for the logistic regression model up until the fitting of the SVM model.

(replace tables below with SVM-table.png)

|  |  |
| --- | --- |
| Support Vector Machine – Model Accuracy Scores | |
| Binary outcomes, without stock data | Binary outcomes, with stock data |
| 88% | 91% |

|  |  |
| --- | --- |
| Support Vector Machine – Model Accuracy Scores | |
| Multi outcomes, without stock data | Multi outcomes, with stock data |
| 80% | 81% |

Neural Networks

The highest accuracy score for the Neural Network model was 91% for binary outcomes without stock data.

Preprocessing was the same as that for the Logistic Regression and SVM models but also included one-hot encoding of the binary class target variable.

Model parameters were initialized with one layer, 50 nodes, and 100 epochs. Activation was ‘relu’ for the input layer and ‘sigmoid’ for the output layer.

(replace table below with NN-table.png)

|  |  |
| --- | --- |
| Neural Networks – Model Accuracy Scores | |
| Binary outcomes, without stock data | Binary outcomes, with stock data |
| 91% | 87% |

(Add plot of accuracy versus epochs called NN-plot.png)