Report on Time series prediction with Machine Learning

Submitted-by: Mahdi Hasnat Siyam, **1705003**

To run the attached python file, the sci-kit module must be installed in python.

Name of the company: Nokia
Name of the ML models:

1. Linear Regression
2. K Neighbor Regression
3. Polynomial Feature Regression
4. Gradient Boosting Regression
5. Elastic Net Regression

Way of training and testing:
80% random days were chosen for training and rest were used for testing.

Dealing with overfitting and underfitting:
Changing seed in test train split does not significantly alter MSE values for 80% random test split. So 80% split for this dataset seems no overfit or underfit.

MSE values:
Linear Regression: 0.009845662377604623
KNN Regression: 0.018514795918367376
Polynomial Regression: 0.010135060505842283
GradientBoostingRegressor: 0.012334756167537929
ElasticNet Regression: 0.09211289306538942

Based on MSE values, Linear Regression performed better.

Conclusion:
Increasing length of the consecutive days to consider (instead of 5) should perform better. There might be more models in the `sklearn` library. Other models might perform better.