- **Yassine Sfaihi**
- www.linkedin.com/in/yassinesfaihi
- www.github.com/yassinesfaihi
- www.kaggle.com/yassinesfaih

[Author] : Yassine Sfaihi

Daily News-Driven Stock Market Prediction through Textual Data Analysis and Machine Learning

This project focuses on the problem of binary classification using various machine learning models.
The goal is to predict the outcome of a binary variable based on a set of input features. The project covers the complete process of text preprocessing, feature generation, and model training and evaluation.

The text preprocessing stage involves cleaning and transforming the text data into numerical features using techniques such as removing stop words, stemming, and vectorizing the text data. The feature generation stage involves generating additional features from the input data by utilizing techniques such as term frequency-inverse document frequency (TF-IDF) and extracting information from the text data such as word count, character count, and punctuation count.

The models used in this project include **Logistic Regression, Support Vector Classifier, Random Forest, Gradient Boosting, K-Nearest Neighbors, and Decision Tree**. The performance of the models was evaluated using metrics such as accuracy, precision, recall, and F1 score. The Support Vector Classifier (SVC) performed the best in terms of F1 score and its performance was further improved by tuning its hyperparameters using grid search.

This project provides insight into the various techniques used in text preprocessing and feature generation and the impact they have on the performance of the models. Additionally, the project demonstrates the importance of hyperparameter tuning in improving the performance of machine learning models.