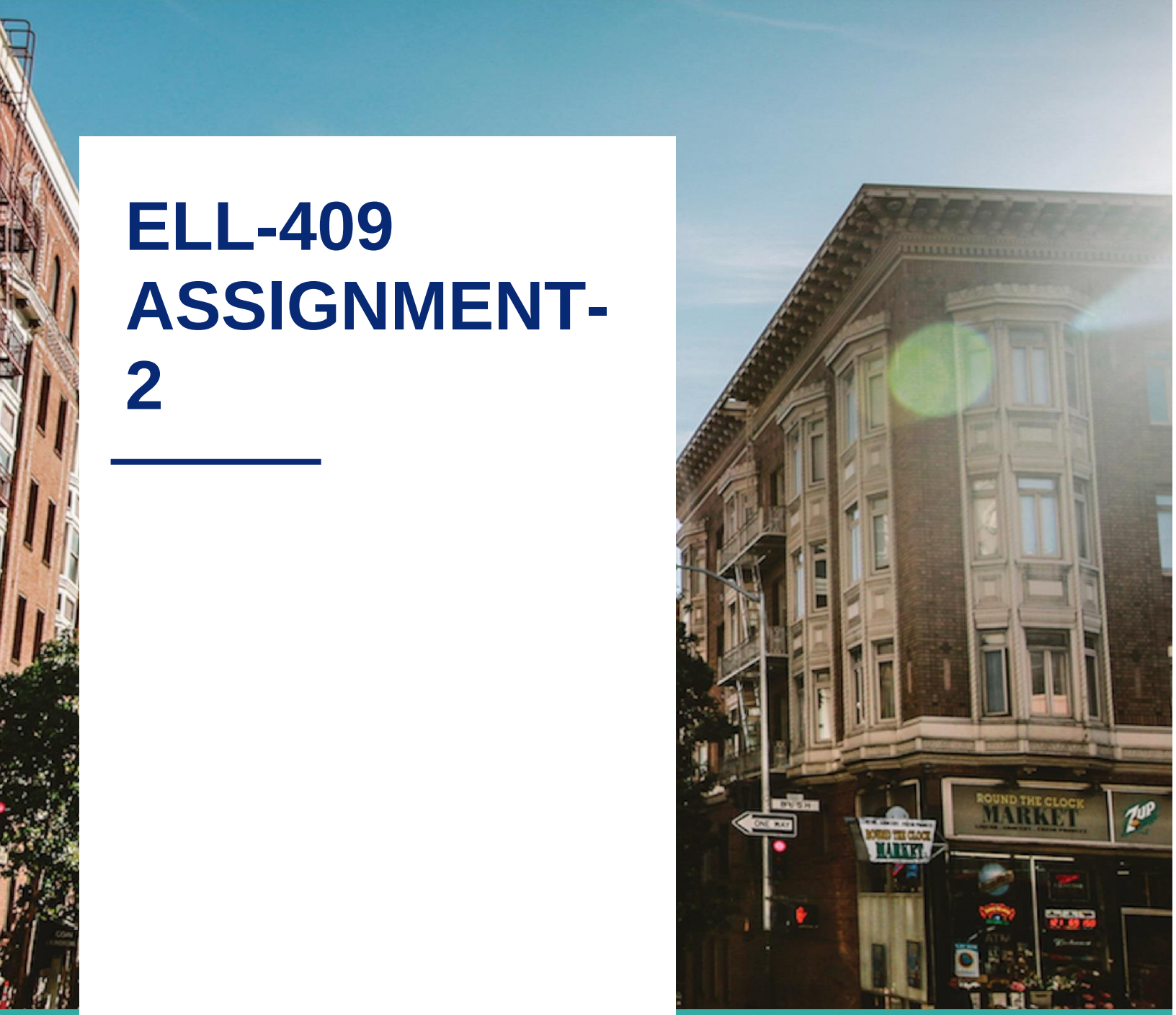

ELL-409 ASSIGNMENT-2

NOVEMBER 15

Aman Yadav **2021EE30734**

Logo Name

SVR STOCK PRICE PREDICTION

1. Introduction:

- Overview of the Project:

- Financial markets are highly dynamic, influenced by various factors, including news events. This project aims to predict stock prices for Apple Inc. using Support Vector Regression (SVR) and incorporates sentiment analysis on news headlines to capture market sentiment.

- Importance of the Project:

- Financial markets are not solely driven by historical stock data; external events and sentiment play a crucial role. By integrating news sentiment, the model aims to enhance predictive accuracy and responsiveness to market changes.

2. Data Collection:

- News API Key:

- The News API key is essential for accessing up-to-date and relevant news data. This API key facilitates seamless integration of news events into the stock prediction model.

- Historical Stock Data:

- The yfinance library is employed to download historical stock data for Apple Inc. during the specified date range (October 15, 2023, to November 14, 2023). This data forms the foundation for training and evaluating the SVR model.

3. Data Preprocessing:

- Combining Stock and News Data:

- Historical stock data and news data are merged based on the publication date. This integration allows for a comprehensive dataset, aligning financial market movements with relevant news events.

- Handling Missing Data:

- Missing news data is handled by forward-filling, ensuring a continuous dataset for analysis. This approach preserves the temporal sequence of news events.

4. Feature Engineering:

- Sentiment Analysis:

- Sentiment analysis is applied to news headlines using the TextBlob library, generating a new feature called 'Sentiment.' This feature quantifies the overall sentiment of news articles, providing a numerical indicator of market sentiment.

	Date	Close	News
0	2023-10-16	178.720001	[{'source': {'id': None, 'name': 'Biztoc.com'}...
1	2023-10-17	177.149994	[{'source': {'id': None, 'name': 'Applelinkage...
2	2023-10-18	175.839996	[{'source': {'id': None, 'name': 'SiliconANGLE...
3	2023-10-19	175.460007	[{'source': {'id': None, 'name': 'Wealthofgeek...
4	2023-10-20	172.880005	[{'source': {'id': None, 'name': 'Applelinkage...
5	2023-10-23	173.000000	[{'source': {'id': None, 'name': 'Biztoc.com'}...
6	2023-10-24	173.440002	[{'source': {'id': None, 'name': 'New York Pos...
7	2023-10-25	171.100006	[{'source': {'id': None, 'name': 'Biztoc.com'}...
8	2023-10-26	166.889999	[{'source': {'id': None, 'name': 'SiliconANGLE...
9	2023-10-27	168.220001	[{'source': {'id': None, 'name': 'TheGrio'}, '...
10	2023-10-30	170.289993	[{'source': {'id': None, 'name': 'ComingSoon.n...
11	2023-10-31	170.770004	[{'source': {'id': None, 'name': 'Marketscreen...
12	2023-11-01	173.970001	[{'source': {'id': None, 'name': 'Kernel.org'}...
13	2023-11-02	177.570007	[{'source': {'id': None, 'name': 'SiliconANGLE...
14	2023-11-03	176.649994	[{'source': {'id': None, 'name': 'Marketscreen...
15	2023-11-06	179.229996	[{'source': {'id': None, 'name': 'Freerepublic...
16	2023-11-07	181.820007	[{'source': {'id': None, 'name': 'InvestorsObs...
17	2023-11-08	182.889999	[{'source': {'id': None, 'name': 'MarketWatch'...
18	2023-11-09	182.410004	[{'source': {'id': 'fox-news', 'name': 'Fox Ne...
19	2023-11-10	186.399994	[{'source': {'id': None, 'name': 'Techreport.c...
20	2023-11-13	184.800003	[{'source': {'id': None, 'name': 'Forbes'}, 'a...

	Sentiment
0	0.028201
1	0.143549
2	0.126329
3	0.044570
4	0.041875
5	0.018746
6	0.094948
7	0.055671
8	0.093808
9	0.035137
10	0.055389
11	0.040051
12	0.091201
13	0.051089
14	0.059030
15	0.038884
16	0.116086
17	0.089908
18	0.026288
19	0.029288
20	0.100303

5. Data Exploration and Visualization:

- Exploratory Data Analysis (EDA):

- The displayed data includes essential features such as stock closing prices, news headlines, and the newly created 'Sentiment' feature. Visualizations aid in identifying trends and potential relationships between variables.

6. Data Splitting and Standardization:

- Train-Test Split:

- The dataset is split into training and testing sets (80-20 split). This division ensures the model is trained on a representative sample while retaining data for evaluation.

- Standardization:

- Standardizing the data is crucial for the SVR model. This process ensures that all features are on a similar scale, preventing certain features from dominating the model.

7. Model Building:

- SVR with RBF Kernel:

- The Support Vector Regression (SVR) model with the Radial Basis Function (RBF) kernel is chosen. The RBF kernel is well-suited for capturing complex relationships in financial data, making it a suitable choice for stock price prediction.

8. Model Training and Evaluation:

- Training the Model:

- The SVR model is trained on the standardized training data, learning patterns from historical stock prices and associated news sentiment.

- Making Predictions:

- Predictions are made for the adjusted closing price for the next 30 days, providing insights into potential future market movements.

- Model Evaluation:

- Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) are utilized to assess the model's performance on the test set.

9. Handling Future Data:

- Predictions for the Next 30 Days:

- Future dates are generated, and predictions are made using the trained SVR model. The placeholder sentiment value can be updated with actual sentiment analysis on future news for more accurate predictions.

10. Model Evaluation on Test Set:

- Performance Metrics:

- MAE, MSE, and R-squared values are presented to evaluate the accuracy and reliability of the SVR model on the test set.

```
Mean Absolute Error (MAE): 3.0979354608172684
Mean Squared Error (MSE): 12.713990485504047
R-squared (R^2): 0.5611480097595876
```

11. Model Extension and Saving:

- Extending the Model:

- Suggestions for extending the model are provided, such as incorporating additional features or experimenting with different kernels.

- Model Persistence:

- The trained SVR model is saved as a pickle file, ensuring model persistence and enabling future use without the need for retraining.

12. Conclusion:

- Summary of Findings:

- Key findings from the project, including the impact of sentiment on stock price predictions, are summarized.

- Limitations and Future Work:

- Limitations of the current approach are acknowledged, and opportunities for future research and improvements are proposed.

13. Code Submission:

- Accompanying Code:

- The Python code is provided with detailed comments, enhancing understanding and facilitating potential modifications or extensions to the project.