Regression
Project: Yes Bank Stock Closing Price Prediction

Project Description:

Yes Bank is a well-known bank in the Indian financial domain. Since 2018, it has been in the news because of the fraud case involving Rana Kapoor. Owing to this fact, it was interesting to see how that impacted the stock prices of the company and whether Time series models or any other predictive models can do justice to such situations. This dataset has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month. The main objective is to predict the stock's closing price of the month.

Main Libraries to be Used:

- Pandas for data manipulation, aggregation
- Matplotlib and Seaborn for visualisation and behaviour with respect to the target variable
- NumPy for computationally efficient operations
- Scikit Learn for model training, model optimization, and metrics calculation

Project should include:

1. **Problem Statement**
2. **Import libraries**
3. **Load dataset**
4. **Data cleaning**
    - Handle missing values
    - Convert data types
    - Remove duplicates
5. **Exploratory Data Analysis (EDA)**
    - Visualize distributions
        i. Univariate analysis
        ii. Bivariate analysis
        iii. Multivariate analysis
    - Correlation analysis
    - Feature-target relationships
6. **Outlier treatment**
    - Boxplot
7. **Check distributions & apply transformations (if needed)**
    - Skewness/Kurtosis
    - Log Transformation, sqrt
8. **Feature engineering**
    - Create new features
    - One-hot encoding (for categorical)
9. **Split data into train/test sets**
10. **Train Linear Regression model**

11. **Feature Scaling**
12. **Prediction using the algorithm**
13. **Visualise the Predicted and Actual**
14. **Print the difference between Actual and Predicted**
15. **Evaluate model performance**
    - MAE - Mean Absolute Error
    - MSE - Mean Squared Error
    - RMSE - Root Mean Squared Error
    - R^2 Score - R squared
    - Adjusted R^2
16. **Perform the same steps for**
    - Decision tree Regressor
    - Random Forest Regressor
    - Support Vector Machine
    - K Nearest Neighbor
17. **Perform the Cross Validation using Cross_val_score for all the algorithms**
18. **Print the final Conclusion**

Link to Dataset: https://github.com/rahulinchal/SPPU

Data Description:

Variable	**Description**
Date	Date of record
Open	Opening Price
High	Highest price in the day
Low	Lowest price in the day
Close	Closing price of the day

Project Architecture: