Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1. **Name** → Meenakshi
   Email → meenakshicuul@gmail.com
   Role :
   - **Data Munging**
     1. **Change "Date" column format**
     2. **Checking for duplicate and null values**
     3. **Log transformation**

   - **Data Visualization**
     1. **Distribution of numerical columns**
     2. **Heat map**
     3. **Pair grid**
     4. **Scatter plot**
     5. **Line plot**

   - **Correlation Analysis**
     1. **Between independent variables**
     2. **Between dependent and independent variables**

   - **Regression Analysis**
     1. **Linear Regression (With and without Cross Validation)**
     2. **Lasso and Ridge Regression (With and without Cross Validation)**
     3. **Elastic Net (With and without Cross Validation)**
     4. **K-Neighbor Regressor (With and without Cross Validation)**

   - **PPT**
   - **Group Colab**

2. **Name** → Tushar R. Wagh
   Email → waghtushar7276@gmail.com
   Role :
   - **Data Munging**
     1. **Change "Date" column format**
     2. **Checking for duplicate and null values**
     3. **Log transformation**

   - **Data Visualization**
     1. **Distribution of numerical columns**
     2. **Heat map**
     3. **Scatter plot**
     4. **Line plot**

   - **Correlation Analysis**
     1. **Between independent variables**
     2. **Between dependent and independent variables**

   - **Regression Analysis**
     1. **Linear Regression**

      2   Support Vector Regressor
      3   K-Neighbor Regressor

- **Technical Documentation**

3. **Name** → Aditya Singh Thakur
   Email → imchillingadi@gmail.com
   Role :
   - **Data Munging**
     1. **Change "Date" column format**
     2. **Checking for duplicate and null values**
     3. **Log transformation**

   - **Data Visualization**
     1 **Distribution of numerical columns**
     2 **Heat map**
     3 **Pair plot**
     4 **Scatter plot**
     5 **Line plot**

   - **Correlation Analysis**
     1. **Between independent variables**
     2. **Between dependent and independent variables**

   - **Regression Analysis**
     1 **Linear Regression**
     2 **Lasso and Ridge Regression**
     3 **Elastic Net**

   - **PPT**

Please paste the GitHub Repo

Github Link :
https://github.com/meena25091992/Yes_Bank_Stock_Closing_Price_Prediction:

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Yes Bank Limited is an Indian Private Sector bank headquartered in Mumbai, India and was founded by Rana Kapoor and Ashok Kapoor in 2004.The Yes Bank stock market is a dynamic and volatile industry and Investors generally decide to buy or sell the stock based on company's past and present performance.

But the stock price of Yes bank was actually fall from 2018 onwards which shows the impact 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 any predictive models can do justice to such situations. This dataset has monthly stock prices of the bank since its inception and includes Closing, Opening (Starting), Highest, and Lowest stock prices of every month.

The Main objective is to predict the stock's closing price of the month. We have to build models which help us to predict the future stock prices.

I have applied various Regression Models in Our YES-Bank-Stock-Price-Prediction such as follows:-
1. Linear Regression.
2. Lasso Regression (with and without Cross Validation)
3. Ridge Regression (with and without Cross Validation)
4. Elastic Net Using Cross Validation
5. K-Neighbor Regressor (with and without Cross Validation)

Some insights:-
1. Dependent variable (Close) is highly correlated with other Independent Variables.

2. We have seen that there are neither null nor duplicate values. But "Date" feature has values of object data type. So, we converted it into proper date format YYYY-MM-DD.

3. With the help of Visualization, we have seen that from 2018 onwards there is sudden fall in the stock closing price. It makes sense how severely Rana Kapoor case fraud affected the price of Yes Bank Stocks.

4. With the help of distribution plot, we see that our data is positively skewed. So, we apply some kind of transformation i.e. Log Transformation to convert it into a normal distribution.

Some Conclusions:-
1. K-Neighbor Regressor and K-Neighbor Regressor CV performing better than other models with adjusted R^2 as 981 % and 99.16% respectively.

2. We have implemented Cross Validation on different algorithms as CV performs better on small datasets. But, the result is nearly same.

3. In all the Models except K-Neighbor Regressor, The Accuracy lie within the range of 81% to 83% and there is no such improvement in accuracy score even after hyperparameter tuning.

4. K-Neighbor Regressor Cross Validation performs best with very less mean square error i.e. 0.015.