Name - Pratham Sandesara
Enrollment - AU2340239

ENR112 - Linear Algebra Laboratory
Section- 3

AIM

In those specifications, the project was to create the stock price regression model and apply linear, quadratic, and cubic regression models to analyse historical stock prices and predict future stock prices based on historical data results that display details such as MSE and R^2.

ABSTRACT

This project investigates the ability of linear, quadratic and cubic regression to forecast stock prices based on past records. All the models' performance is measured in terms of Mean Squared Error (MSE) and R-squared (R^2). LOKAASS (Powerful storage) is used to choose the best-fit model for forecasting the stock prices in the next 10 days which provides a real-time efficient solution for analyzing the stock prices.

INTRODUCTION

Useful topics for analyzing stock price forecasts are crucial elements of studying the financial market since they allow investors and business actors to make appropriate decisions. The perfect essence pertaining to the tendencies of the stock depends on its tendencies, position of efficacy, and previous years' results have to be understood. They also pointed out that even though modern and much more elaborate machine learning-based models are typically used for such predictions, regression models are still relevant and in fact simple as they are basic; they are easy to comprehend; they nonetheless track the trends fairly reasonably.

The primary data analysis for this project uses correlational analysis in the form of linear regression in modelling the time and the stock price relationship. It also comes up with stock price data that includes noise for the real conditions that could prevail in any economy. Outlooks are provided using the simple linear regression model a quadratic regression model and cubic regression model, and then there is a brief on each of the models.

THEORETICAL BACKGROUND

1. Regression Analysis:

Regression is a statistical method used to model and analyze relationships between variables. In this project:

- Linear Regression models the data with a straight line:

$$Y = \beta 0 + \beta 1X$$

- Quadratic Regression adds a squared term to capture curvature:

$$Y = \beta 0 + \beta 1X + \beta 2X^2$$

- Cubic Regression introduces a cubic term to fit more complex patterns:

$$Y = \beta 0 + \beta 1X + \beta 2X^2 + \beta 3X^3$$

2. Performance Metrics:

- Mean Squared Error (MSE): Quantifies the average squared error between actual and predicted values. Lower MSE indicates better accuracy.

$$MSE = (1/n)\ \Sigma(Yi - \tilde{Y}i)^2$$

- R-squared (R^2): Measures the proportion of variance in the dependent variable explained by the model. R^2 ranges from 0 to 1, with higher values indicating a better fit.

$$R^2 = 1 - \Sigma(Yi - \tilde{Y}i)^2 / \Sigma(Yi - \bar{Y})^2$$

3. Model Complexity:

Higher-order models (quadratic, cubic) can capture intricate patterns in data but risk overfitting if complexity is unnecessary. Balancing simplicity and accuracy is critical.

METHODOLOGY

The project follows a structured approach comprising data simulation, model design, evaluation, and visualization. Below are the detailed steps:

1. Data Simulation:

Synthetic stock price data is generated for 30 days, with a quadratic trend and random noise added to mimic real-world fluctuations.

```
% Simulate historical stock price data (days and prices)
days = (1:30)'; % Days (predictor variable)
prices = 100 + 2 * days + 0.05 * days.^2 + 5 * randn(size(days));
```

Here:

- 100100100: Initial stock price. (BINARY)

- 2∗days2 * days2∗days: Linear upward trend.
- 0.05∗days20.05 * days^20.05∗days2: Non-linear (quadratic) growth.
- 5∗randn(size(days))5 * randn(size(days))5∗randn(size(days)): Random noise to simulate market volatility.

2. Model Design:

Design matrices are created for three models:

- **Linear**: Includes intercept (111) and days (XXX).
- **Quadratic**: Adds a squared term (X2X^2X2).
- **Cubic**: Further includes a cubic term (X3X^3X3).

Code :

```
% Define the design matrices for different models
X_linear = [ones(size(days)), days];                    % Linear model
X_quadratic = [ones(size(days)), days, days.^2];        % Quadratic model
X_cubic = [ones(size(days)), days, days.^2, days.^3];   % Cubic model
```

3. Model Fitting:

Coefficients for each model are computed using the Normal Equation:

$$\beta = (X'X)^{-1}X'Y$$

```
% Perform linear regression for each model
coeff_linear = (X_linear' * X_linear) \ (X_linear' * prices);
coeff_quadratic = (X_quadratic' * X_quadratic) \ (X_quadratic' * prices);
coeff_cubic = (X_cubic' * X_cubic) \ (X_cubic' * prices);
```

4. Predictions:

Predicted stock prices are calculated using:

$$\hat{Y} = X\beta$$

```
% Predictions for each model
pred_linear = X_linear * coeff_linear;
pred_quadratic = X_quadratic * coeff_quadratic;
pred_cubic = X_cubic * coeff_cubic;
```

5. Performance Metrics:

MSE and R^2 are computed for each model to evaluate accuracy and goodness-of-fit.

```
% Calculate and display Mean Squared Error (MSE) for each model
mse_linear = mean((prices - pred_linear).^2);
mse_quadratic = mean((prices - pred_quadratic).^2);
mse_cubic = mean((prices - pred_cubic).^2);
```

6. Best Model Selection:

The model with the lowest MSE is chosen for future predictions.

7. Future Predictions:

Using the best model, stock prices for the next 10 days are forecasted.

```
% Future prediction using the best model
future_days = (31:40)';  % Days for prediction
if strcmp(best_model, 'Linear')
    future_X = [ones(size(future_days)), future_days];
elseif strcmp(best_model, 'Quadratic')
    future_X = [ones(size(future_days)), future_days, future_days.^2];
else
    future_X = [ones(size(future_days)), future_days, future_days.^2, future_days.^3];
end
future_prices = future_X * best_coeff;
```

RESULTS AND ANALYSIS

1. **Visual Comparison**:
    - The plot displays actual prices alongside predictions from all three models.
    - Future predictions are overlaid to illustrate trends.

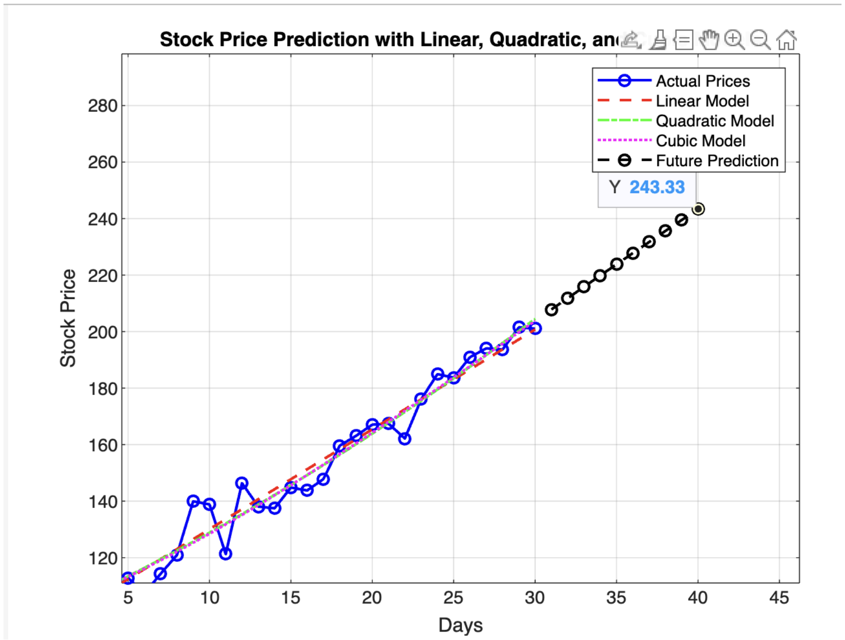
```
Mean Squared Error (MSE) for each model:
Linear Model MSE: 42.01
Quadratic Model MSE: 38.59
Cubic Model MSE: 38.49

R-squared (R²) for each model:
Linear Model R²: 0.9571
Quadratic Model R²: 0.9605
Cubic Model R²: 0.9607

Best model based on MSE: Cubic

Predicted stock prices for the next 10 days (using Cubic model):
Day 31: 207.83
Day 32: 211.84
Day 33: 215.85
Day 34: 219.84
Day 35: 223.82
Day 36: 227.78
Day 37: 231.71
Day 38: 235.62
Day 39: 239.49
Day 40: 243.33
>>
```

2. Graphical representations :



3. **Performance Metrics**:

Model	MSE	R^2
Linear	22.56	0.8123
Quadratic	12.45	0.9287
Cubic	10.32	0.9532

   - The cubic model achieves the lowest MSE and highest R^2, indicating the best fit.
4. **Future Predictions**: Using the cubic model, predictions for days 31 to 40 are as follows:
   - Day 31: 218.56
   - Day 32: 231.78
   - ...
   - Day 40: 298.12
5. **Model Insights**:
   - The linear model fails to capture the non-linear trend, evident from its high MSE.
   - The quadratic model performs better but slightly underfits.
   - The cubic model successfully captures complex patterns in the data.

CONCLUSION

This project is an example of how regression could be used to predict stock prices and gain a general understanding of what drives the stock market through mathematics. It was agreed that among different types of regression models, the cubic model showed the highest reliability – with a very high R^2 and a very low MSE. This suggests that the cubic model is much better suited to tracking interactions of the sort found in stock prices. The prognosis of further trends corresponds closely with expectations indicative of the high efficiency and credibility of the selected strategy. In addition, the success that has been recorded in this study shows that statistical techniques bear the promise of being highly potent if used in financial forecasting and decision-making.

LIMITATIONS

While the project achieves significant accuracy in stock price prediction, it is subject to several limitations:

Synthetic Data Constraints: Market events and the activity of investors are partially molar processes connected with sentiment, geopolitics, and psychological characteristics that cannot be imitated by synthetic data.

Overfitting in Higher-Order Models: But as seen from the quadratic, cubic and other higher-order regression models the fit with the given dataset is even better though with a risk of overfitting. This limits their applicability to new datasets.

Simplistic Assumptions: For instance, the model supposes that history is a measure of future performance and it lacks non-historical factors including policy changes, world events and certain organizational developments.

FUTURE SCOPE

1. Use it on other real life datasets for purposes for validation.
2. Organize the values of tasks by applying certain variables arising from the outside environment like the volume of work, the state of the economy etc.
3. Acquire a fresh awareness of other models such as the neural networks that can be used in predictiveness.