AOL Scientific Computing

Nama : Kelvin Jonathan Yusach
NIM : 2702209533
Kelas : LA09

You are working as a Data Scientist in EGIER, an outdoor equipment manufacturer in Bandung. As a Data Scientist, you have been tasked with processing some data on EGIER warehouse capacity. You are given a set of data that consist of the monthly production of a certain type of bag they produced. The data span from January 2018 to December 2023 (given as M1 to M144). The data is given in CSV (comma-separated value) document, and you can access the document here:
https://docs.google.com/spreadsheets/d/1iSg4bEQfkfi2o58HmcMDII0i6InqnZrT/edit?usp=sharing&ouid=103170622443612923929&rtpof=true&sd=true.
Your supervisor has given you a series of task that needs to be done on the data, as a part or your job!

1. (LO2 30 points) You must find the trend on the bag's production from the data. You must provide a mathematical model that can explain the production's trend accurately. Since your supervisor want an accurate model, you must avoid any linear approach to build the trend model.

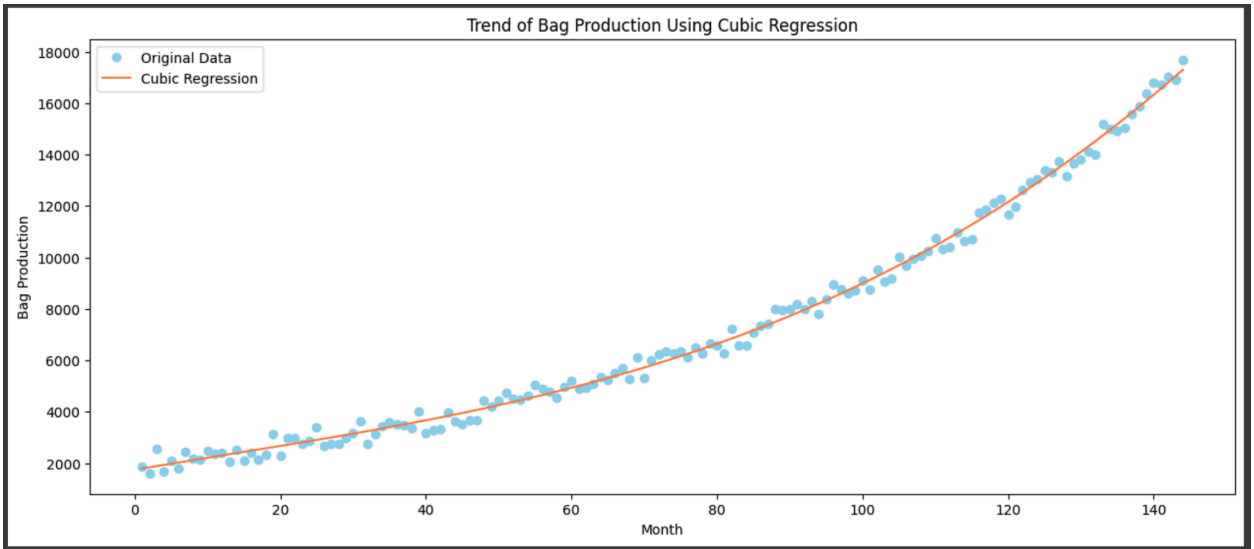
In this case, we will use a non-linear model because a linear model will not be sufficient to accurately represent the changes.

Bag production does not follow a simple linear pattern. Instead, it shows more complex changes caused by various factors such as customer demand, production constraints, and market fluctuations.

I chose to use Cubic Regression because this model can capture complex and dynamic data patterns. Compared to Quadratic Regression, Cubic Regression is more effective in representing trend changes since it has two turning points, whereas Quadratic Regression only has one.

Additionally, Cubic Regression typically results in lower error values than Quadratic Regression because it can better adapt to datasets with large variations over longer periods of time. For this 6-year bag production data, a Quadratic Regression model would produce a pattern that is too simple and might overlook important details that are crucial for long-term analysis and forecasting.

Results obtained



Based on the graph, it can be seen that bag production gradually increased over time from January 2018 to December 2023. This trend also shows an acceleration in the rate of bag production. Therefore, not only did the number of bags produced increase, but the pace of production also grew—possibly due to rising market demand.

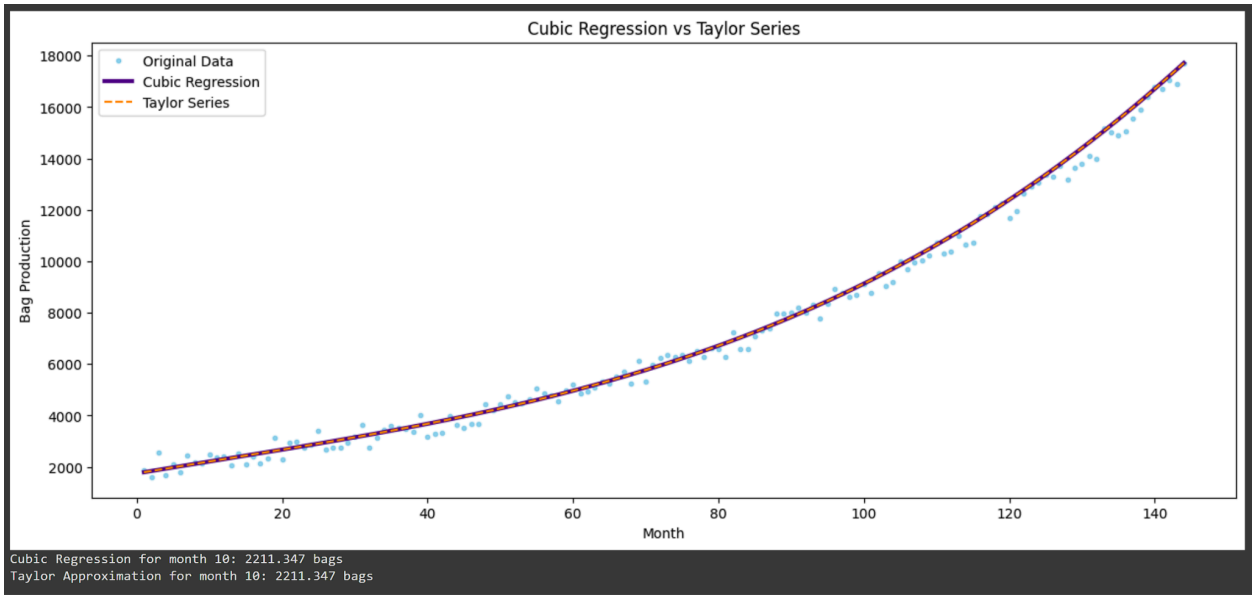
2. (LO3 25 points) Since you'll need to process the data using a computer program, you'll need to convert the mathematical model from problem #1 to its numerical form (approximation). This is done so that the mathematical model can be calculated by the program easily. Since accuracy is still important, make sure that your conversion is accurate as possible. Provide an explanation to your supervisor about the accuracy of your conversion.

To demonstrate the accuracy of the Cubic Polynomial Regression, I used the Taylor Series approach. I utilized the Taylor expansion of the regression equation to approximate the values of the Cubic Regression. Then, I compared the results visually using plots and tested a single data point as a sample to calculate the number of bags produced using both methods.

The resulting Cubic Regression equation is:
$y = 0.004x^3 - 0.134x^2 + 47.224x + 1748.507$

The Taylor Series approximation (around a = 1) is:
$y(x) = 1795.601 + 46.968(x - 1) - 0.122(x - 1)^2 + 0.004(x - 1)^3$

Results obtained



Based on the displayed graph, both equations—Cubic Regression and Taylor Series—show the same results in modeling the bag production data. In addition, the calculated result for a sample point, which is at the 10th month, yields an identical number of bags from both models.

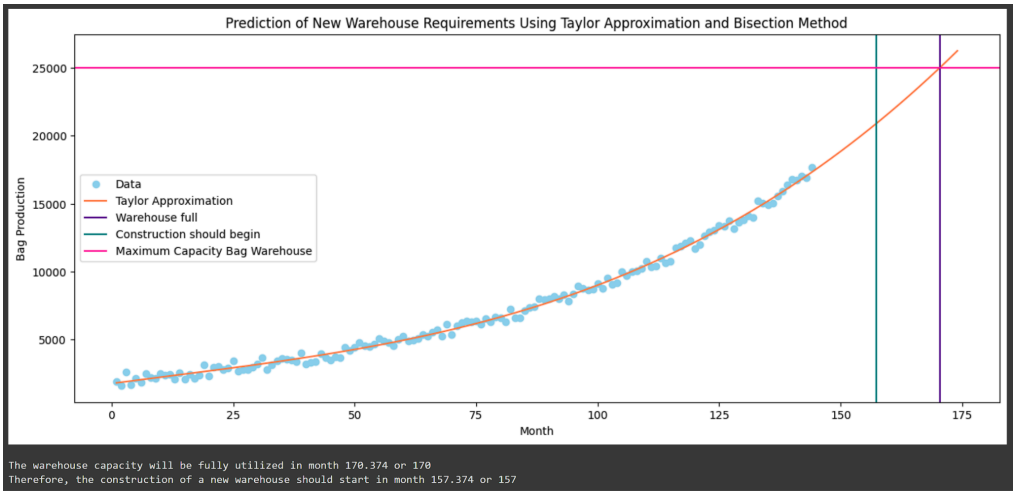
This is because the Cubic Regression uses a third-degree polynomial to describe the relationship between variables x and y, while the Taylor Series calculated at the reference point $t_0 = 1$ also produces similar polynomial components, namely x^3, x^2, and x. With matching coefficients in both models, the Taylor Series effectively approximates the Cubic Regression, especially at points close to t_0. As a result, the two curves nearly overlap and produce very similar predictions for the bag production data.

3. (LO3 25 points) The warehouse was designed to be able to store a maximum of 25,000 (twenty five thousands) bags at each month. Your supervisor asked you to provide a prediction when do EGIER need to build a new warehouse based on the trend that you have acquired in problem #2. To build a new warehouse, it is predicted that they need at least 13 months. So provide the time when EGIER need to start building their new warehouse. (Hint: this can be approached as a root of equation problem).

The bisection method was chosen because it is well-suited for this problem. The equation f(x) = predicted_production(x) - 25,000 is a continuous function, allowing us to find its root by ensuring a sign change within the initial interval, such as between month 145 and 300. The advantage of this method is that it always produces an accurate answer if the initial interval is correctly chosen. Unlike other methods like Newton-Raphson, which require the derivative of the function and a highly accurate initial guess, the bisection method does not require derivatives, making it easier to apply. Additionally, other methods such as the secant method also require a fairly accurate initial estimate and are less stable compared to bisection. Due to its reliability, simplicity, and ease of implementation, this method is the best choice for determining when production exceeds warehouse capacity.

The Taylor Approximation approach was used because the previous graph results showed that the production data follows the pattern of both Cubic Regression and the Taylor Approximation.

Results obtained



The orange line on the graph represents the predicted results using the Taylor approximation, which follows the trend of the production data (blue dots). To determine when production exceeds the warehouse's maximum capacity, the bisection method was used to gradually find the root of the equation f(x) = predicted_production(x) - 25,000 within a specific interval. As a result, the warehouse will reach full capacity in month 170, as indicated by the purple line. Since constructing a new warehouse takes 13 months, the construction must begin in month 157, which is marked by the green line on the graph. The pink horizontal line represents the warehouse's maximum capacity of 25,000 bags.

4. (LO1 20 points) Your supervisor want to double check your result. Provide your code by sharing your Google Colab Notebook URL/ complete screenshot of your code that you have used to answer problem #1 to #3.

No 1
https://colab.research.google.com/drive/1QVicCPQPLcNih3Ao6cOLNrofBQrA9RHk?usp=sharing

No 2
https://colab.research.google.com/drive/1oeN6oxjXe2pC9trCFiu_LNGA8BvUl-bQ?usp=sharing

No 3
https://colab.research.google.com/drive/1BFnhCKiY1vOphMVSfjxKQEckwH4Gd69M?usp=sharing