Optimizing Cloud Cost Management: Cost Forecasting and Anomaly Detection for Dexcom's GCP Infrastructure

Donna Huang, Fengyi Wang, Helena Yu, Phoebe Chen

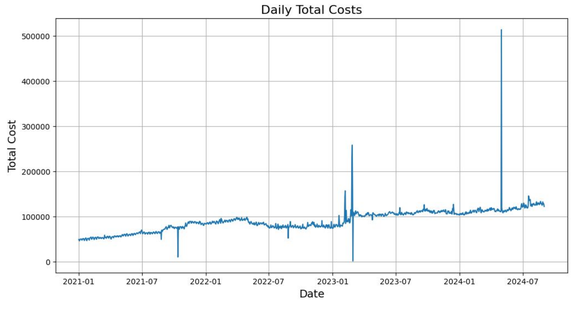
Introduction

Cloud computing is a critical foundation for modern businesses. For Dexcom Inc, a leader in continuous glucose monitoring systems, Google Cloud Platform supports real-time data processing and secure storage across departments. As Dexcom grows, managing the complexity and costs of its cloud infrastructure becomes challenging. This project aims to provide a data-driven cloud cost management strategy, combining expenditure forecasting and anomaly detection to optimize spending without compromising performance or scalability.

Data & Exploratory Analysis

The dataset has 828,679,516 rows and 7 columns, spanning from January 2021 to August 2024.

Except for **project_id**, **sku**, **usage_unit**, and **total_usage_amount**, the main variables of interest include:
service: type of cloud service consumed (e.g., Compute Engine)
total_cost: total expenditure corresponding to resource usage
usage_date: date when resources were used



Method

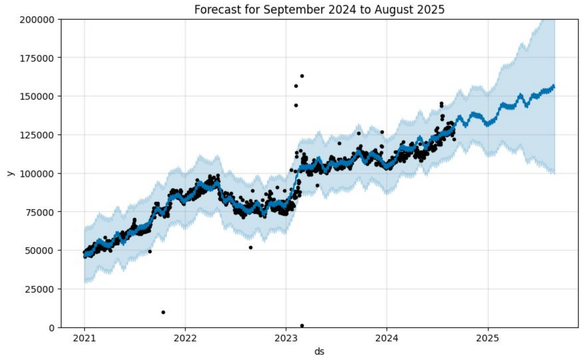
We used two methods for cost forecasting and anomaly detection – Prophet and LSTM.

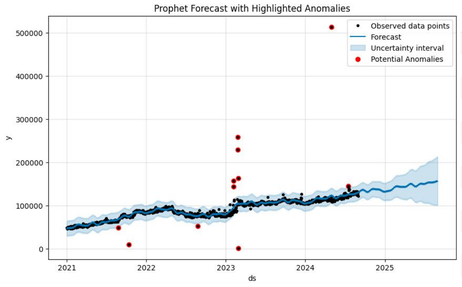
Approach 1: Prophet
- Developed by Meta, specifically designed for time series forecasting, handle complex seasonal patterns, and missing data
- Algorithm similar to Generalized Additive Model, predicting with trend function, periodic changes, and irregular schedule effects
- Use Prophet for prediction and classify anomalies as data that deviate from the uncertainty intervals by more than a "buffer" value

Approach 2: LSTM (Long Short-Term Memory)
- Developed for sequence prediction tasks, a type of recurrent neural network (RNN) capable of learning long-term dependencies
- Possesses Forget gate, Input gate, and Output gate to determine what information to be discarded, stored, and output from memory
- Use LSTM for cost forecasting and define anomalies as data above or below 2 standard deviations from the loss mean

Results

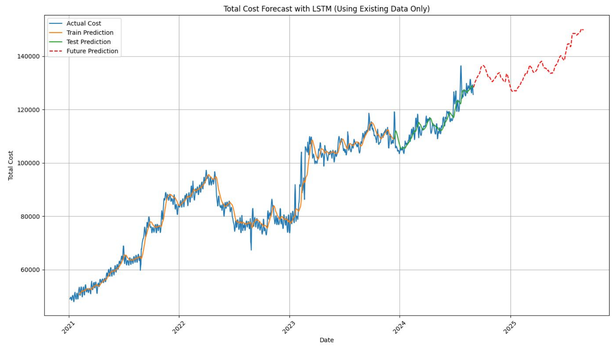
The Prophet model predicts the daily total cost to be increasing for the next year, with minor fluctuations between weeks and months. We also successfully detected anomalies as shown below:

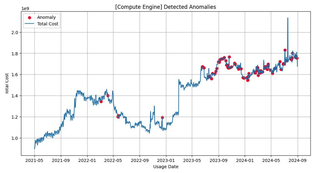


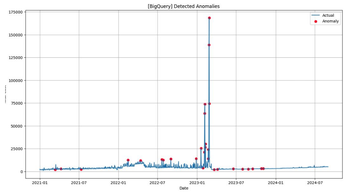


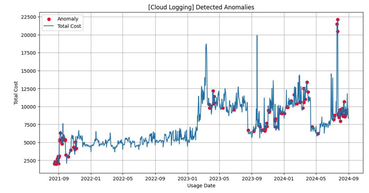
Anomaly Date	Anomaly Cost
2021-08-26	49071.264422
2021-10-13	10048.998604
2022-08-26	51766.658750
2023-02-05	144055.953750
2023-02-06	156576.063696
2023-02-25	229154.657023
2023-02-26	257878.386971
2023-02-27	162927.134497
2023-02-28	1292.482827
2024-04-30	513361.345265
2024-07-17	145428.883621

LSTM predicts the daily total cost to be increasing as well but with slightly more shifts. Anomalies are detected for top 3 services - BigQuery, Compute Engine, and Cloud Logging:









Discussion

Both models show good performance on forecasting: Prophet has a mean absolute error of 6098.54, which is within expectation comparing to daily costs measured in hundreds of thousands; LSTM has an even better loss with 0.003. Therefore, we believe LSTM has a greater potential in achieving precise forecasts.

Anomaly detection for Prophet is mostly accurate after checking with Dexcom. Due to time and resources constraints, we only trained detections on LSTM for 3 services. Although anomalies were not fully detected, we are confident about the results after future fine-tuning and large-scale trainings.

Future Work

1. Feed more data for pattern learning Our analysis leveraged three years of data to forecast cloud costs. Yet the limited timespan of data may hinder the accuracy of yearly predictions, particularly when unusual patterns are present. The prediction results would be more precise when the models learn more data.

2. Pipeline forecast and anomaly detection Currently, the forecast and anomaly detections are two separate processes. One future enhancement stemmed from our project is to build a pipeline that can be activated automatically in the back-end and send alerts when anomalies are detected. If such a pipeline is possible, then Dexcom is able to monitor costs regularly and make adjustments to spendings when necessary.

Reference

1. https://facebook.github.io/prophet/
2.https://www.barchart.com/story/news/28523548/is-dexcom-stock-underperforming-the-s-p-500