THE DEPARTMENT OF
COMPUTER SCIENCE
& ENGINEERING
計算機科學及工程學系

Predicting Daily Orders and Revenue of Delivery Lockers and Identifying Optimal Locations for Placement with AI

TONG, Chiu Tseng
Supervised by Prof. Chiew-Lan TAI

Introduction

To optimize the utilization and strategic placement of delivery lockers in Hong Kong's logistics industry, our client company has tasked us with developing an AI-based predictive model. This model utilizes historical usage data, weather, date, and other relevant information to accurately forecast daily usage and revenue for the next three days. It also identifies optimal locations for placing the lockers, addressing the challenges faced by logistics companies in this market.

Objective

1. **Usage Prediction:** by analyzing historical usage patterns, weather, and market trends, the tool predicts locker daily orders and revenue for the next three days.
2. **Location Scouting:** identify suitable locations for building delivery lockers, taking into account factors such as population density, proximity to residential and commercial areas, transportation infrastructure, and demographics.

Methodology – AI Model Training

To train a model for location scouting and usage prediction, the following steps are followed:

1. Data Collection: gather data from various sources, such as the internet, client companies.

2. Data Cleaning: clean the collected data to ensure its accuracy and reliability.

3. Data Processing: process the cleaned data to derive additional insights and features that can aid in location scouting and predicting daily usage and orders.

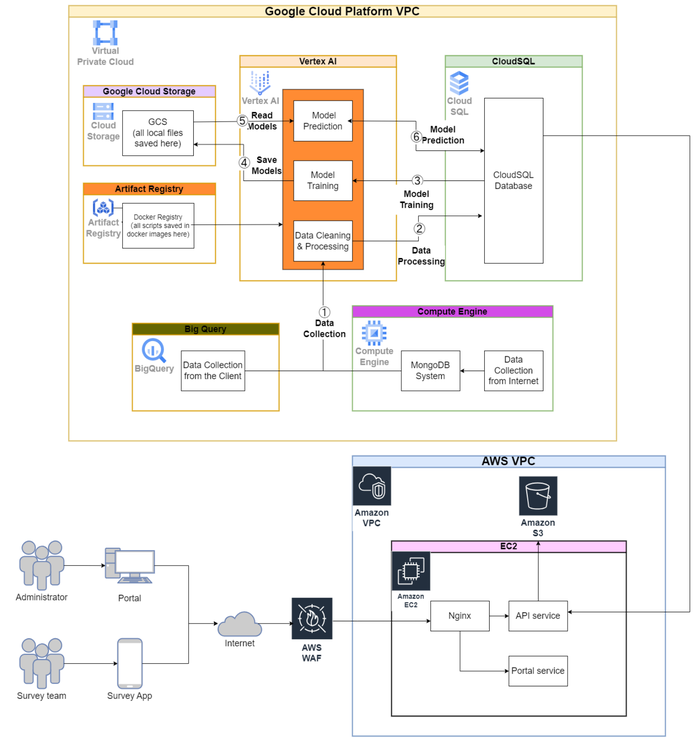
4. Feature Engineering: extract and transform features to make them more suitable for model training.

5. Model Training: select differently appropriate machine learning models to make prediction and choose the ones with the best performance.

6. Model Evaluation: evaluate the trained model's performance using MSE (Mean Squared Error)

7. Model Interpretation: interpret the trained model to gain insights into the factors that contribute to successful location scouting and usage prediction with SHAP.

Methodology – System Architecture



1. **Data Collection:** gather information from the internet and the client company

2. **Data Cleaning and Processing:** Vertex AI is used to fetch the Docker image and perform data cleaning and processing on the internet and client data. Processed data is then saved to the Cloud SQL database.

3. **Model Training:** data fetched from the Cloud SQL database is used to train models for usage prediction and location scouting using Vertex AI.

4. **Save Models:** Trained models are saved in Google Cloud Storage (GCS) as pickle files.

5. **Read Models:** Vertex AI reads the models from GCS.

6. **Model Prediction**: with the data from CloudSQL and models read from GCS, Vertex AI generates predictions, and stores the results in the Cloud SQL database.

After that, the backend system could fetch the data from the CloudSQL database and present them to the frontend.

Model Evaluation

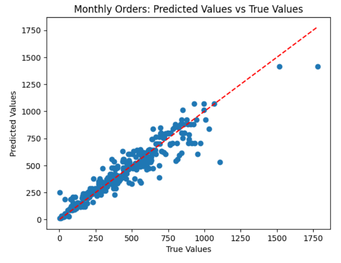


Fig. 1: Scatter Graph of Actual Values vs Predicted Values for **Location Scouting**

Model	MSE	Execution Time
ARIMA	32.8422301825293	2.5 min
XGBoost	26.1982758620689	3 min
LightGBM	25.5431034482758	73 min
Model Ensemble	22.4913793103448	7 min

Fig. 2: Model Evaluation **Comparison** for **Usage Prediction**

- **Location Scouting:** The scatter graph shows that the majority of the data points are tightly clustered around a diagonal line (from the bottom left to the top right), suggesting that the XGBoost model is performing well.
- **Usage Prediction:** a model ensemble approach (by combining ARIMA, XGBoost, and Linear Regression models), are selected for predicting daily orders and revenue because of its lower MSE.

Conclusion

We developed an AI module successfully to address the two objectives for the client:

- For **location scouting**, the module utilizes a trained **XGBoost model** to predict monthly orders and normalize them as scores to potential installation points,
- For **usage prediction**, the module employs a **model ensemble** approach with ARIMA, XGBoost, and linear regression models to accurately forecast daily orders and revenue for individual lockers over the next three days.