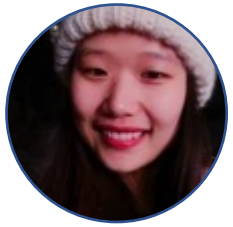
ABOUT ME

Data enthusiast | Life-long learner

Optimization & Machine Learning| Ecommerce | Supply Chain

Who am I?



Ivy is a Data Scientist at Pitney Bowes with more than 3 years of experience in optimization and machine learning. She is currently leading the optimization initiatives using data science techniques to help solve some of the toughest challenges for the company's logistics service sector, including reducing operating costs, improving delivery cycle time, and enhancing visibility into parcel journey and parcel profile.

She has a track record of uncovering hidden patterns from her data science work, continuously delivering actionable insights and strategies to stakeholders, and streamlining data informing processes, such as building dashboards, developing self-service data applications and deploying them using cloud service.

Ivy is a quick learner, a curious person, and always passionate about using data science skills to uncover business insights. With her solid background in Operations Research and years of working experience in applying advanced analytics to solve challenging problems, she is absolutely a valuable asset to any organization.

Here is her GitHub repository attached: https://github.com/yj333?tab=repositories

P.S. Ivy loves skiing, cooking and traveling during her leisure time.

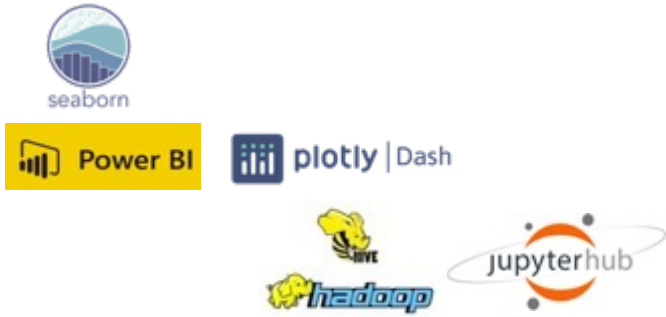
Master's Degree in Operations Research (2016-2017), Cornell University

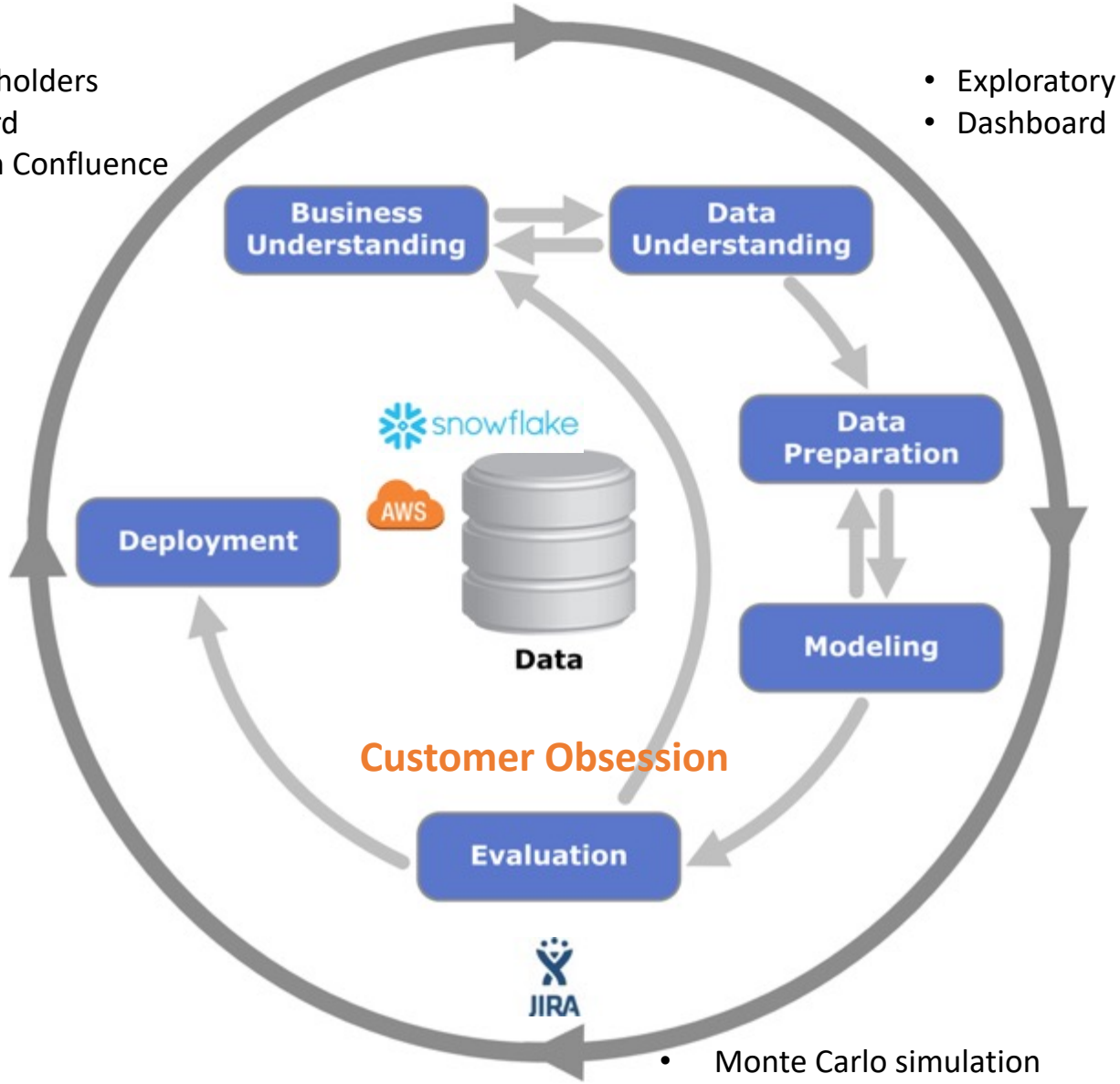
Bachelor's Degree in Chemical Engineering (2012-2016), Georgia Institute of Technology

Overview of my specialties

- Weekly meeting with stakeholders
- Brainstorm on MURAL board
- Document requirements on Confluence

- Exploratory data analysis
- Dashboard





- Productionize models on Elastic Beanstalk
- Send real-time notifications to Microsoft Teams channel

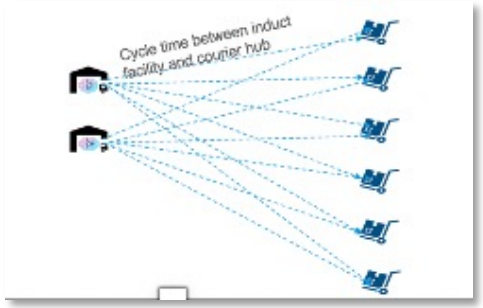
- Data processing
- Feature engineering

- Linear programming/Mixed integer programming
- Linear regression
- Random Forest
- XGBoost
- Clustering
- Time series forecasting
- Image identification

- Monte Carlo simulation
- Sensitivity analysis
- Before-and-after scenario

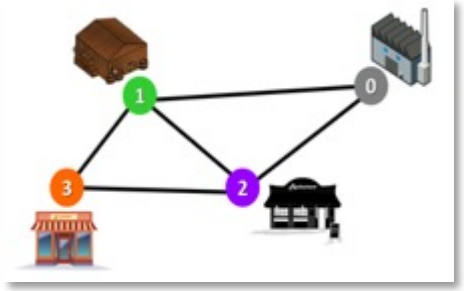
Highlights of my past projects

West Coast Network Design



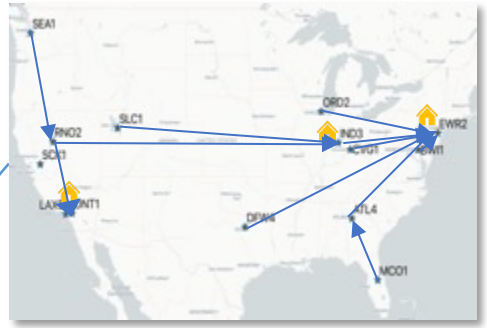
Facility-Region Assignment

Self-service



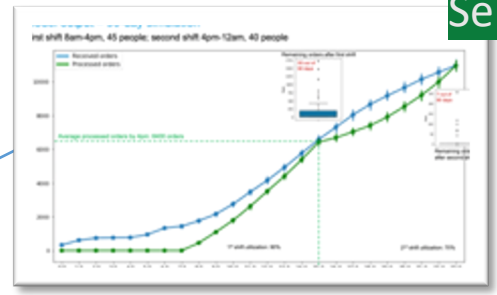
Shortest Path

Self-service



Returns Pricing

Self-service



Fulfillment Shift Scheduling → Shift Scheduler App

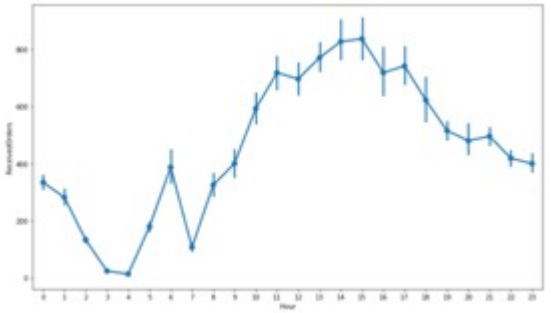
Cornell University Student Projects – project advisor

Transportation/Facility Operations

Global Inbound Optimization

Workforce optimization – Shift scheduling model

1. Received orders by hour (from simulation or forecasts)

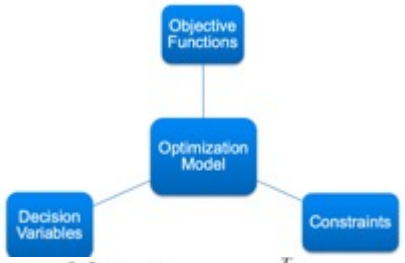


2. Potential shift schedules
3. Orders processing rate

python

GUROBI OPTIMIZATION

Mixed integer programming



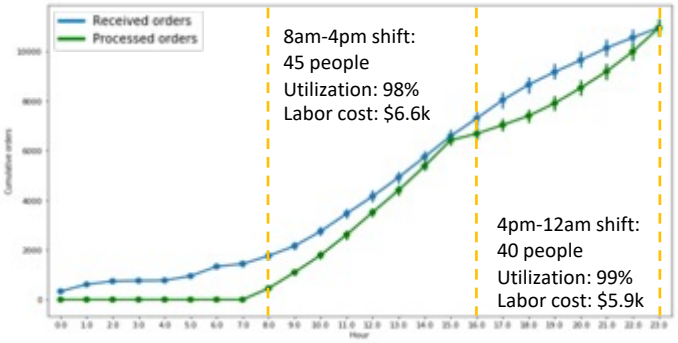
$$\text{Minimize} \quad c^T x$$
$$\text{Subject to} \quad Ax = b$$
$$l \le x \le u$$
$$\text{some or all } x_j \text{ integer}$$

Goal: minimize rollovers and labor cost

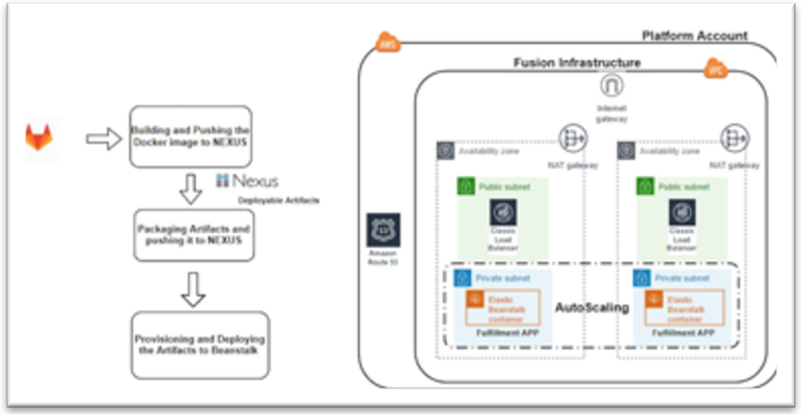
Constraints:

- Workstation capacity
- Processing order

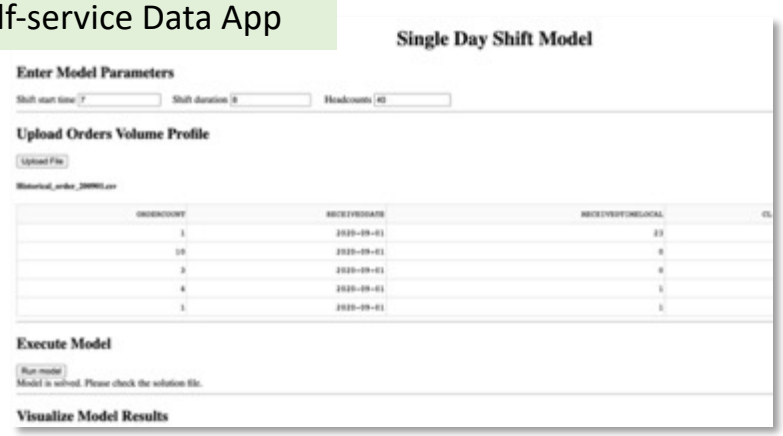
1. Headcounts
2. Hourly throughput
3. Cost and utilization



AWS



Self-service Data App



plotly | Dash

Global inbound route optimization

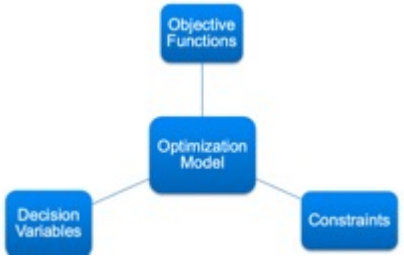
1. Parcels volume distribution by zip code

2. Potential port of entries: JFK, LAX, ATL, ORD

3. Interfacility linehaul cost

4. Interfacility transit time

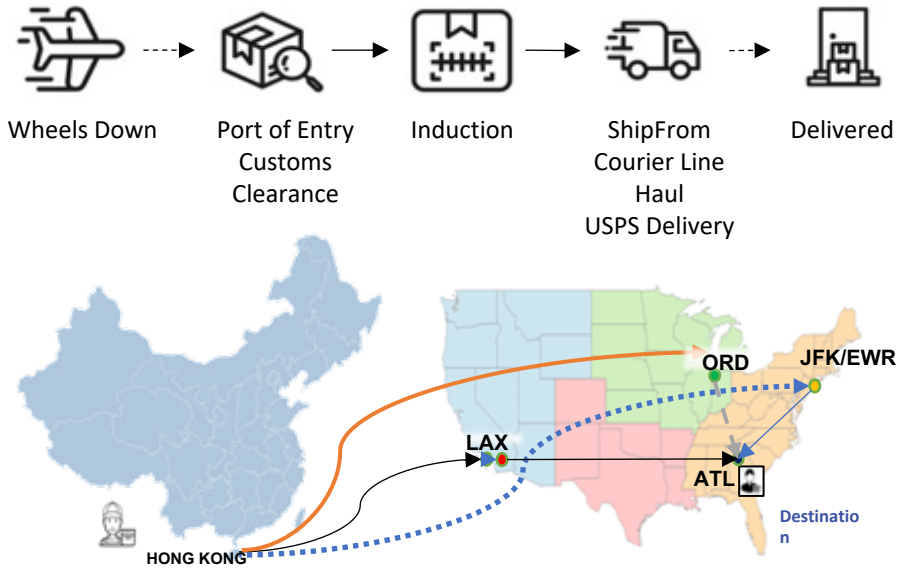
Mixed integer programming



$$\text{Minimize} \quad c^T x$$
$$\text{Subject to} \quad Ax = b$$
$$l \le x \le u$$
$$\text{some or all } x_j \text{ integer}$$

Goal: minimize transportation and operational cost
Constraints: 90% of parcels delivered with 10 calendar days

Zip-level inbound route decision



Real-time parcels anomaly detection

Step 1: Historical baseline (Facility-level; past 30-day)

Historical mean and standard deviation of hourly average weight of parcels

Step 2: Real-time monitoring (Machine-level; by hour)

Realtime hourly average weight by machine

Low-latency parcel automation table

Compare hourly average weight on each machine against historical baseline

Anomaly detection algorithm

- Six Sigma control chart
- CUSUM change point detection
- Local Outlier Factor

Yes

Possible anomaly

Staging table

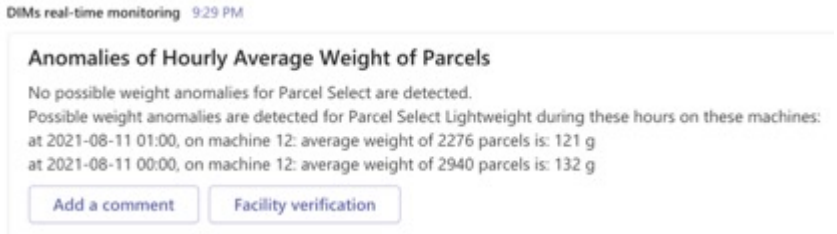
Write anomalies to a staging table on Snowflake

#. of parcels scanned > 100

Yes

Alert triggered

Notify operations of abnormal weight detected by time and by machine on Teams



Step 3: Weekly report (Client-level; by week)

Aggregate the number of parcels by client from the impacted machines in the past week

Client impact

Delivery cycle time prediction

snowflake python scikit-learn

Extract Data

- Define scope
- Aggregate data
- Create staging tables

Explore, Clean, Modify

- Explore distribution of data
- Identify uninteresting features
- Impute/remove missing values
- Treat outliers
- Feature engineering/extraction

Prepare Data for ML

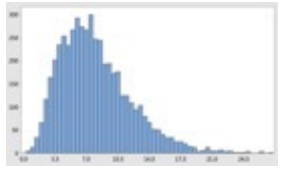
- Train/test split
- Resample for meaningful representation
- Feature scaling
- Feature encoding/binning

Train Models

- Cross-validation
- Model and variable selection
- Feature importance
- Hyperparameter tuning

Evaluate Results

- MSE,MAE,MAPE
- Error analysis
- Precision, recall, accuracy
- Confusion matrix
- ROC curve & AUC
- SME review



Features:

- Induction facility -> facility throughput
- Destination zip code -> target encoding
- Interfacility route -> facility count
- USPS zone, postal class

Models:

- Linear regression
- SVM
- Random forest
- XGBoost

Metrics:

- MSE/MAE
- % of parcels delivered within +/- 1 day of prediction