



How DoorDash leverages AI in its on-demand Logistics Engine

Raghav Ramesh

Machine Learning Engineer, DoorDash

May 2018



Outline

DoorDash Overview

AI in Logistics

Delivery Time Predictions

Batching Algorithms



 DOORDASH

Sign In

Sign Up



Delivering
good moments

 Enter your delivery address

Find restaurants

Get the app:



Last mile,
on-demand logistics

Three-sided
marketplace

Restaurant Delivery

1600 cities by end
of 2018



 DOORDASH

Sign In

Sign Up



Delivering
good moments



Enter your delivery address

Find restaurants

100,000+ Restaurants

300,000+ Dashers

10,000,000s of Deliveries

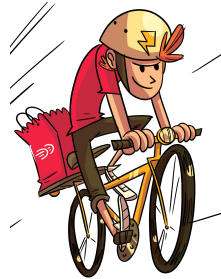
Get the app:



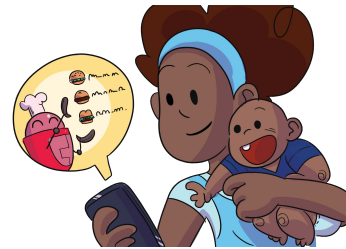


Marketplace

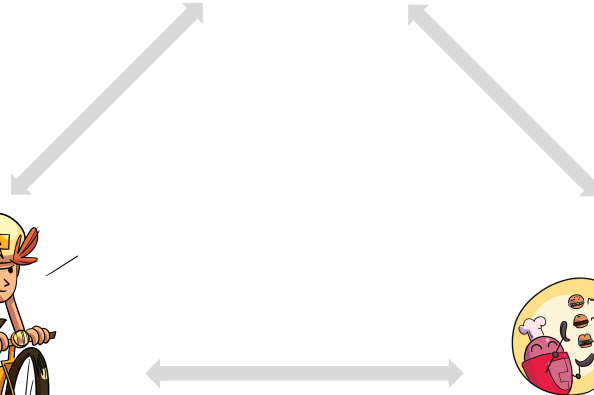
Merchants



Dashers



Consumers





Marketplace

Merchants

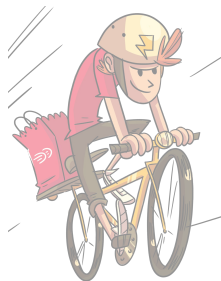


Reach

Revenue

Flexibility

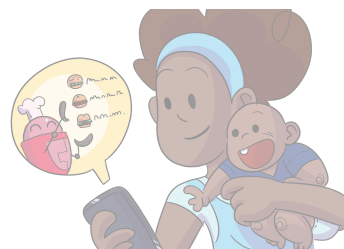
Earnings



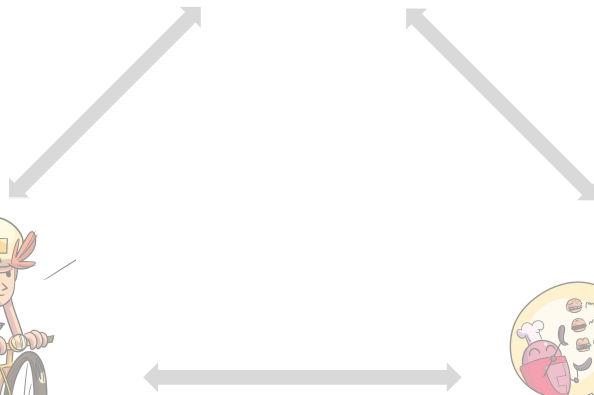
Dashers

Selection

Convenience

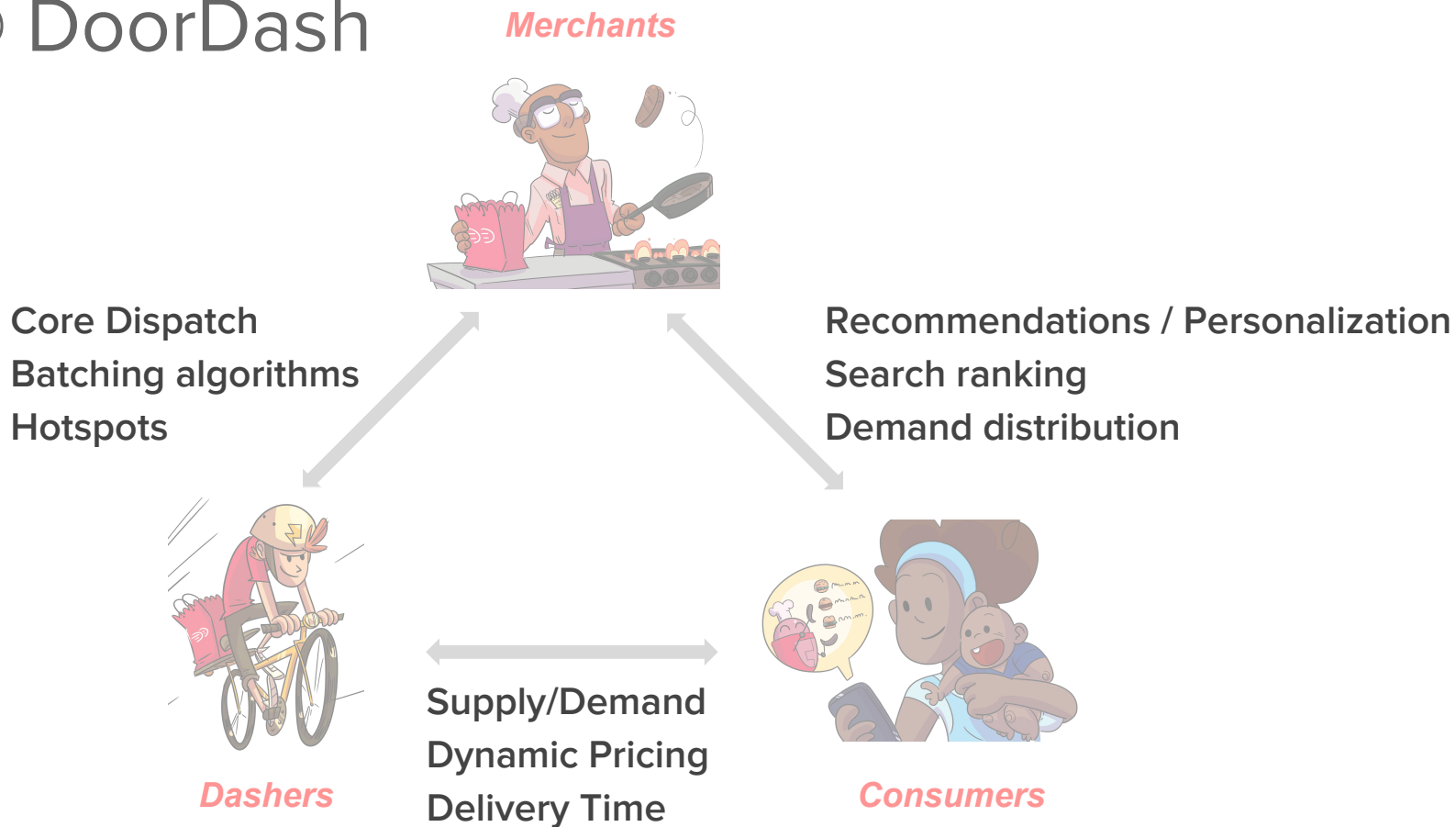


Consumers





AI @ DoorDash





AI @ DoorDash

Merchants

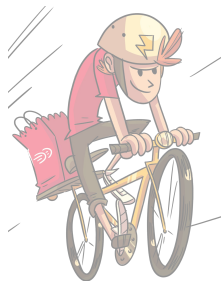


Food prep time
Selection intelligence
Parking prediction

Core Dispatch
Batching algorithms
Hotspots

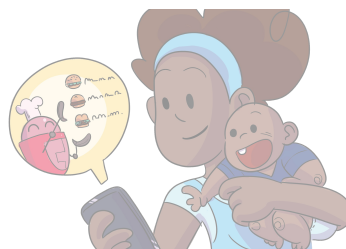
Recommendations / Personalization
Search ranking
Demand distribution

Pay calculation
Supply forecasting
Incentives



Dashers

Delivery Time Predictions
Supply/Demand Management
Dynamic Pricing

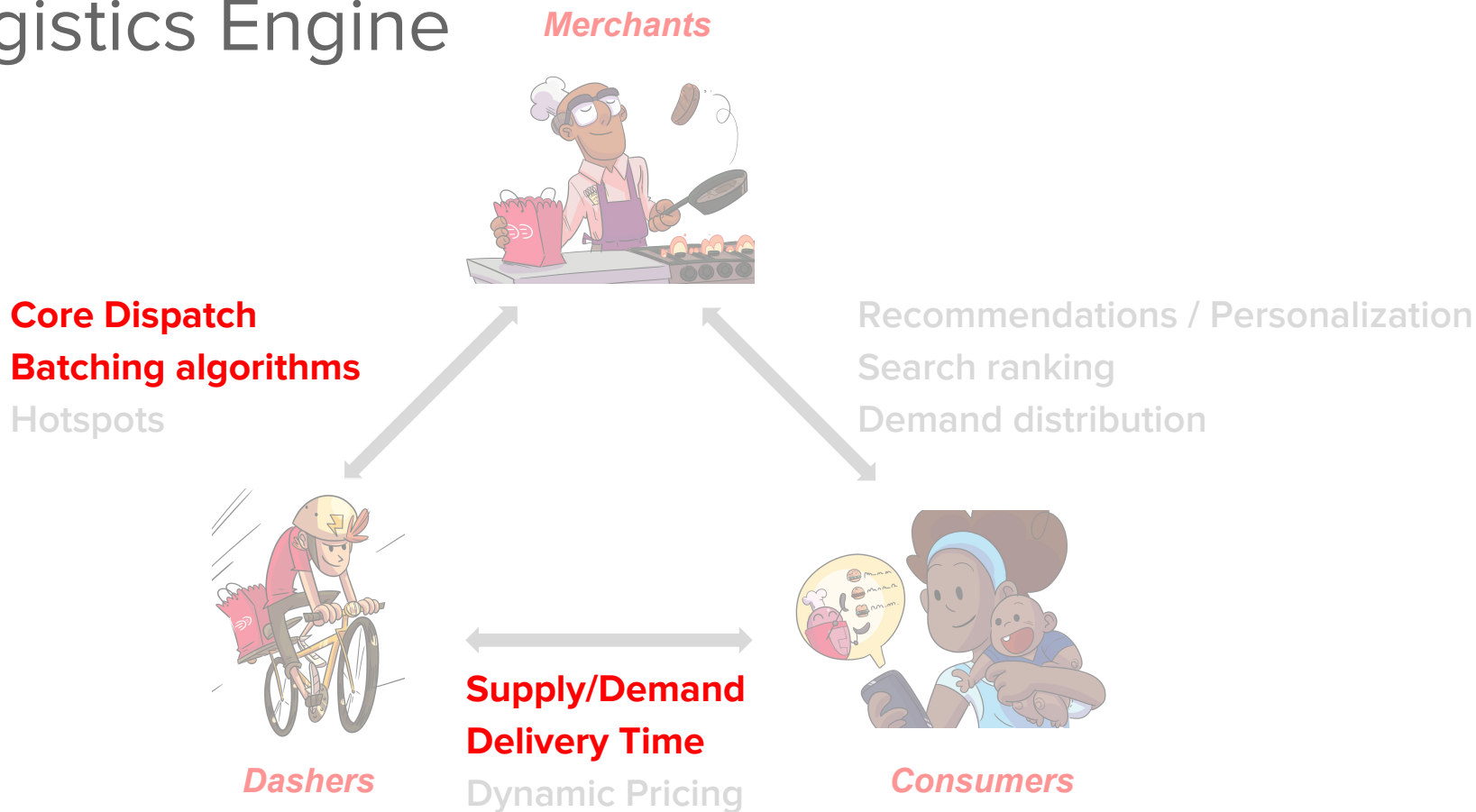


Consumers

Lifetime value
Acquisition
Promotions



Logistics Engine





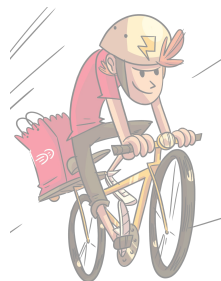
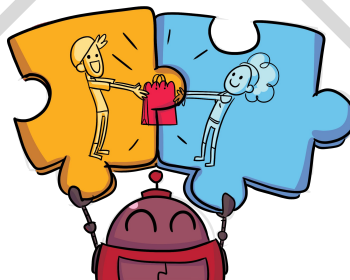
Logistics Engine

Core Dispatch
Batching algorithms
Hotspots

Merchants



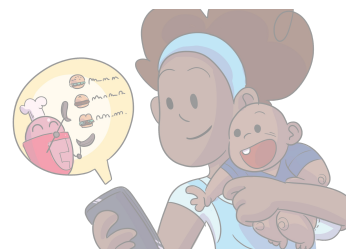
Recommendations / Personalization
Search ranking
Demand distribution



Dashers

Supply/Demand
Delivery Time

Dynamic Pricing



Consumers



Logistics Engine

The AI system that powers DoorDash deliveries



Logistics Engine

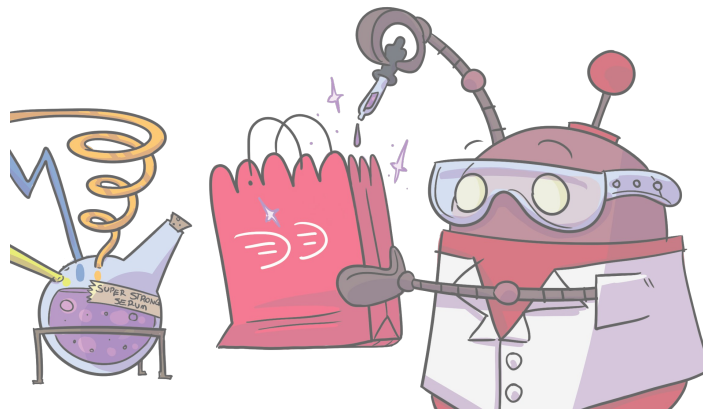
Fast and efficient deliveries



On-time delivery to Consumer



Increase marketplace efficiency

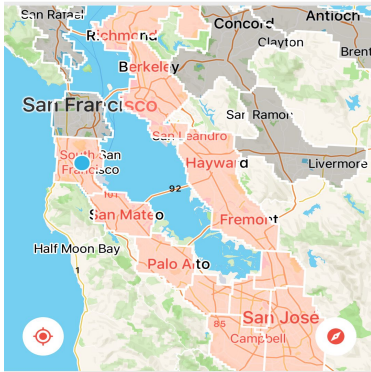




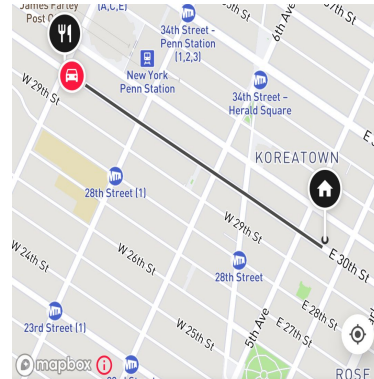
Logistics Engine



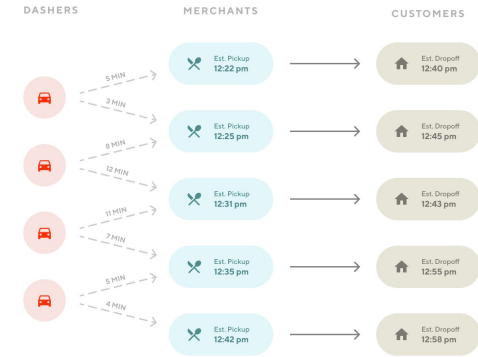
Balance Supply/Demand



Plan Routes



Dasher/Delivery matching

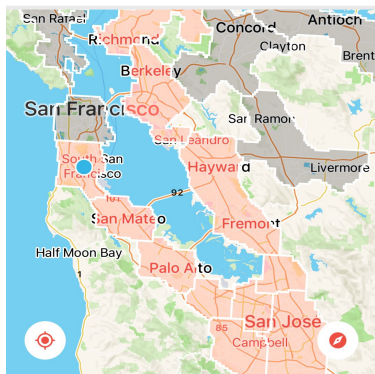




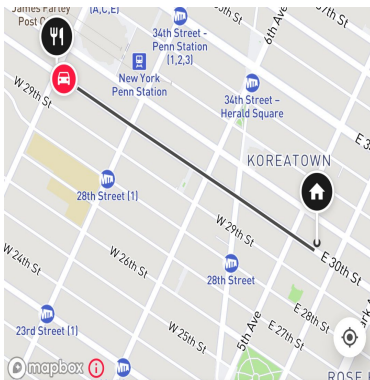
Logistics Engine



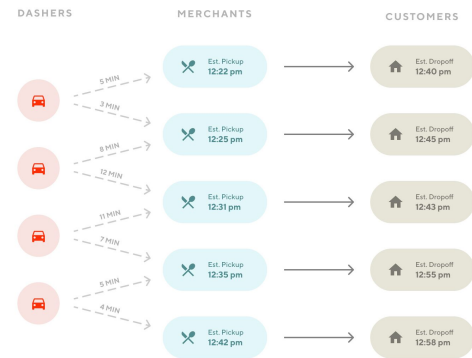
Balance Supply/Demand



Plan Routes



Dasher/Delivery matching





Optimal Matching

In plain English

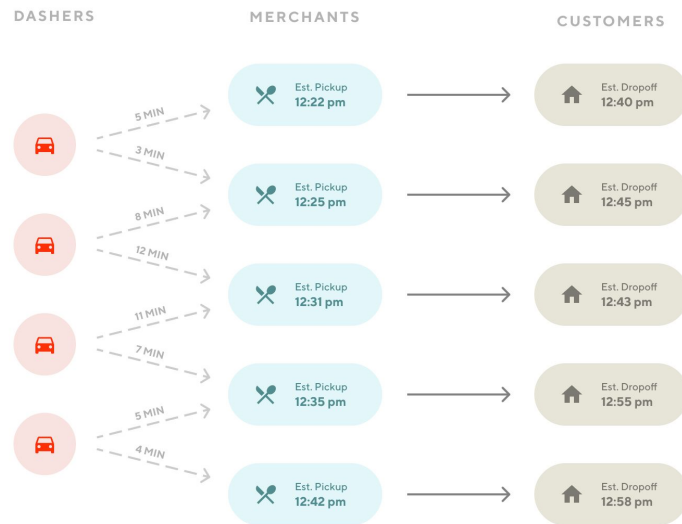
- Pick the best Dasher for a Delivery

In canonical Operations Research

- Vehicle Routing Problem

DoorDash specific considerations

- **Real-time** fulfillment
- Optimize supply for future demand
- Decide **when to assign**

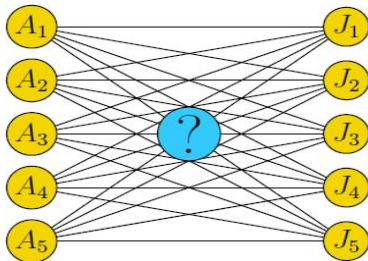


Challenges



Complexity

Combinatorial options
Delivery constraints



Time constraint

Asap delivery
Real-time Consumer demand
Dasher flexibility



Variance

Merchant operations
Traffic
Weather

Merchant	Item	Delivery Time	Rating
V Cafe	Cheesteaks	28 mins	4.6
Split	Sandwiches	25 mins	4.5
Big Mouth Burgers	Burgers	35 mins	4.5
La Mediterranee	Mediterranean	35 mins	4.6





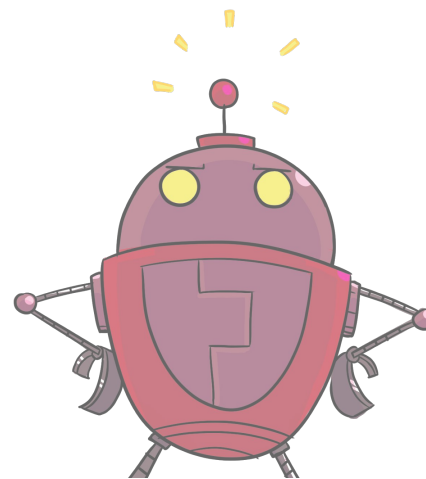
ML meets OR

Operations Research (OR)

- Given deliveries, Dashers, and business goals, determine cost function to use and identify the optimal matching

Machine Learning (ML)

- **Set up the marketplace**
 - Forecast supply and demand
 - Balance
- **Calculate inputs to the cost function**
 - Travel times, Food preparation times
- **Calculate constraints**
 - Variances, Batching estimates
- **Auxiliary**
 - Help with decision on when to assign





Logistics Engine: Summary

Data

Delivery data

Location
Type
Size

Dasher data

Location
Capacity

Constraints

Equipments

Predictions

Total delivery duration

Dasher travel duration

Food prep duration

Batching quality
prediction

Can two orders be
batched?

Optimizations

Cost functions

Optimal matching

Mixed integer
programming

Actions

Dasher
assignment



Let's talk predictions

Data

Delivery data
Location
Type
Size

Dasher data
Location
Capacity

Constraints
Equipments



Predictions

Total delivery duration

Dasher travel duration

Food prep duration

Batching quality
prediction

Can two orders be
batched?



Optimizations

Cost functions

Optimal matching

Mixed integer
programming



Actions

Dasher
assignment



Delivery Time Predictions

When every second is worth a million dollars



Delivery time predictions

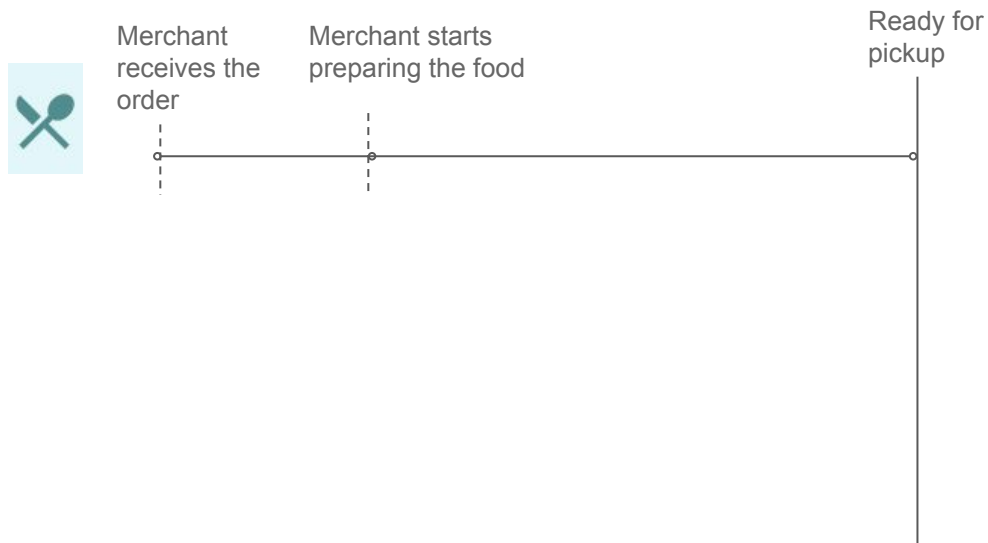
- 10+ time point predictions for every delivery
- Capture each part of the delivery process
- Critical component of Logistics system



Lifecycle of a delivery

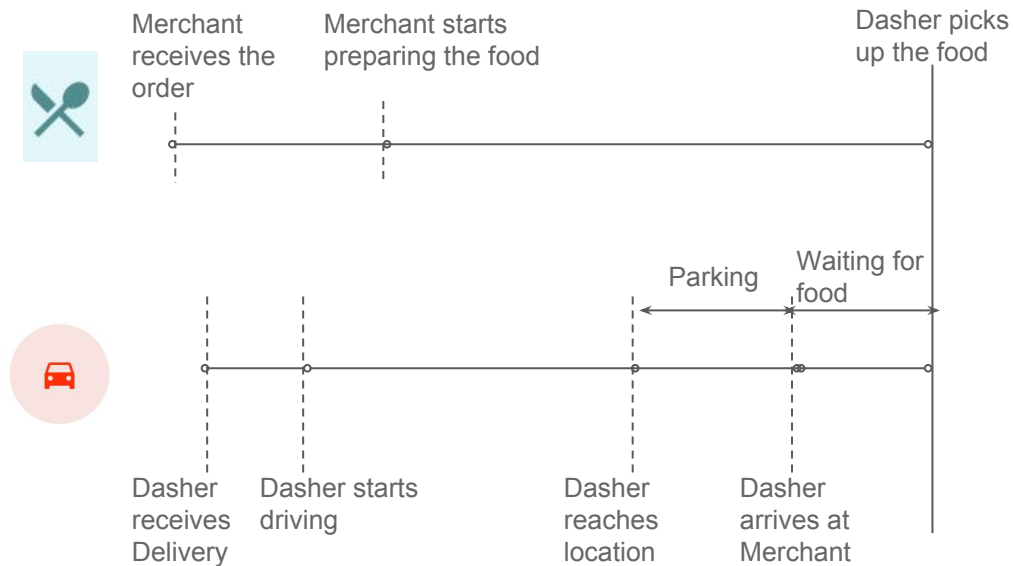


Lifecycle of a delivery



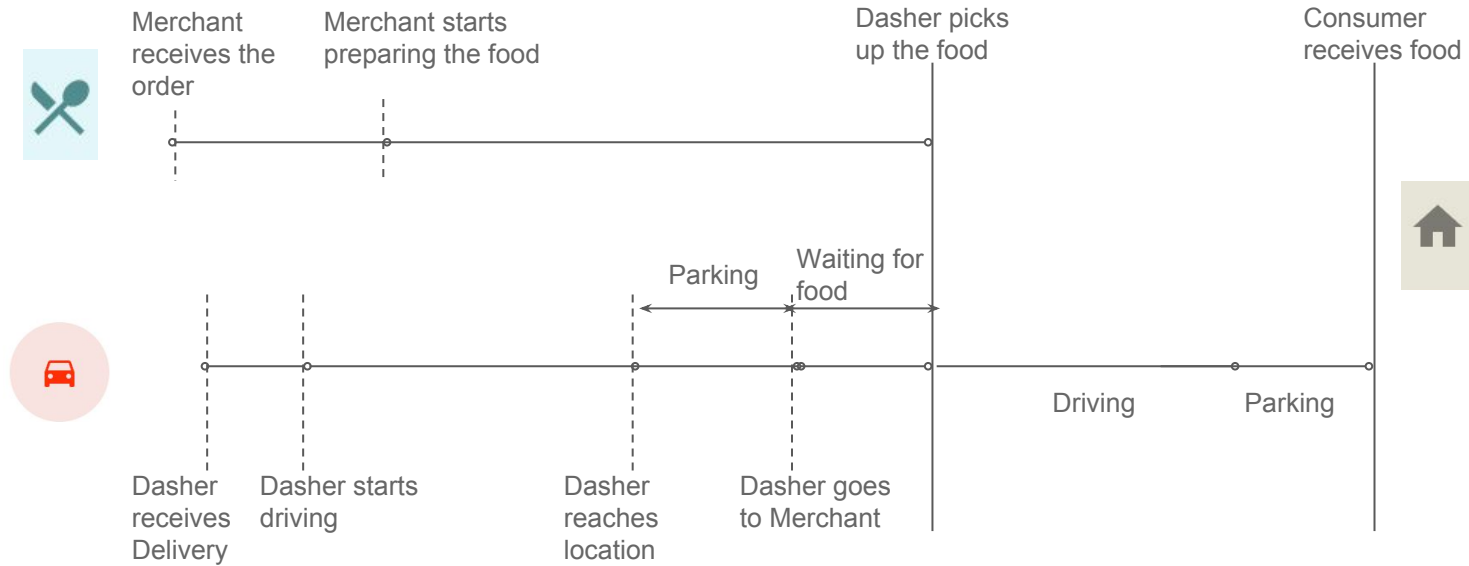


Lifecycle of a delivery



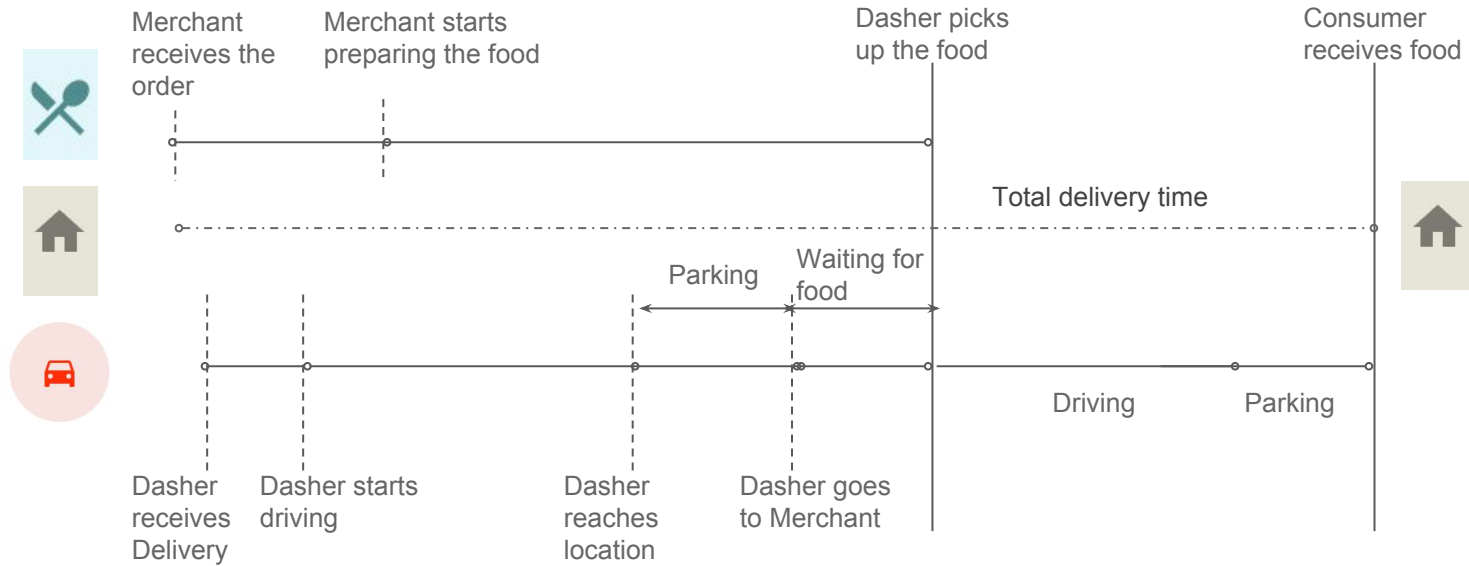


Lifecycle of a delivery



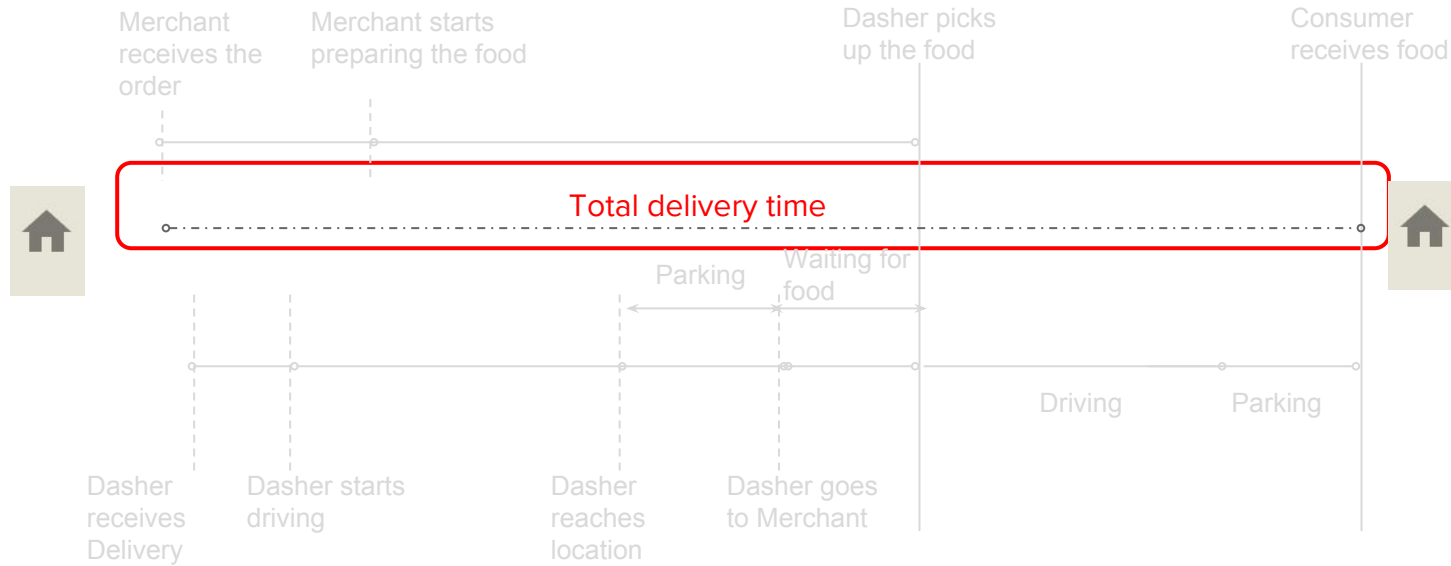


Lifecycle of a delivery





Total delivery time





Overview



Goal

Provide ETA to Consumer

Set constraints for the optimizer



Inputs

Order details

Market conditions



Constraints

Fast, Scalable

Works for new markets



Directly impacts Consumer conversion and retention



Initial Model



Model

Gradient Boosted Decision
Trees
(LightGBM)

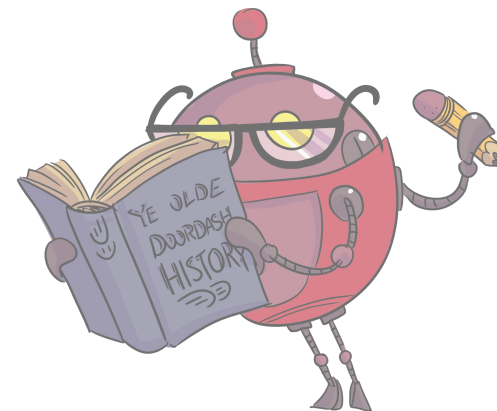
Target: Total Delivery Time



Evaluation

Historical data for training
Validate on latest data

Evaluate on RMSE





Features



Order Features

Subtotal

Cuisine

Type and price of items

Type of order - group,
catering, etc.

Merchant details



Real time features

of orders

of total Dashers

Traffic, Travel estimates

Time of day - lunch / dinner

Day of week - weekend /

holiday



Historical aggregates

Past X weeks average delivery
times for

Store

City

Market

Time of day

Similarly

- average parking times
- variance in historical times



Limitations

- Minimizing mean squared error makes sure we are right on average.
- But, it is more important to deliver by the estimated delivery time, not just be correct on average



Limitations

- Minimizing mean squared error makes sure we are right on average.
- But, it is more important to deliver by the estimated delivery time, not just be correct on average

What we have

$$\textit{Actual Delivery Time} = \textit{Estimated Delivery Time} + \textit{Error}$$

What we want

$$\textit{Prob}(\textit{Actual Delivery Time} > \textit{Estimated Delivery Time}) < X\%$$



Quantile Regression

- Modeling technique to generate **prediction intervals**
- Instead of Mean Squared Error, the **Loss function** used is:

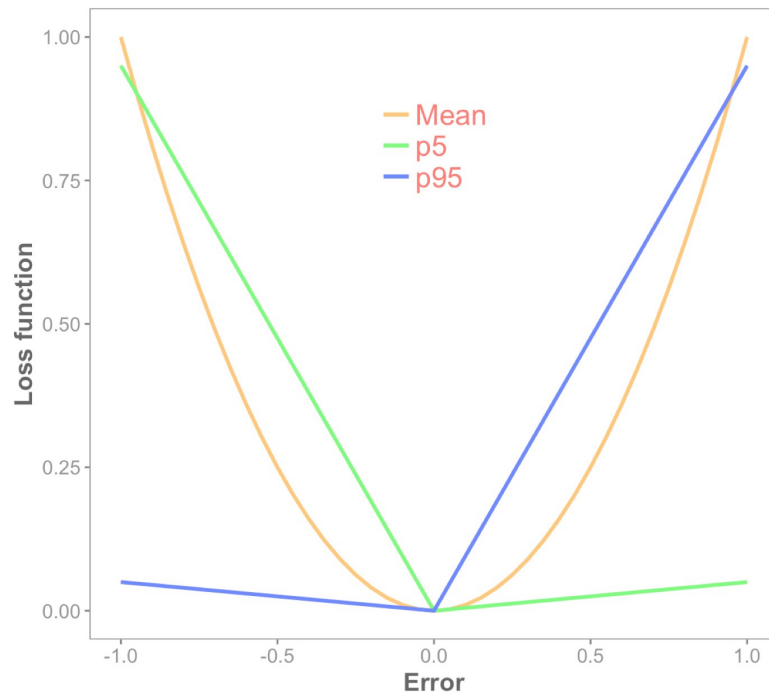
$$L(y, F(x)) = h_q(y - F(x))$$

where

$$h_q(z) = z * (q - I_{z < 0})$$

q is the percentile value to use

I is the indicator function





Quantile Regression Model

- Choosing two q values (say 0.05, 0.95) gives you a **prediction interval** where 90% predictions would lie
- Model evaluated using **% predictions within the interval** and average estimated value
- **Ideal value of q** identified through experiments to optimize retention





Extensions

- Geographical ensembling
 - Market specific model for larger markets
- Use of Embeddings
 - Store
 - Market
 - Time of day
- Near real time aggregates



Let's talk predictions

Data

Delivery data
Location
Type
Size

Dasher data
Location
Capacity

Constraints
Equipments



Predictions

Total delivery duration

Dasher travel duration

Food prep duration

Batching quality
prediction

Can two orders be
batched?



Optimizations

Cost functions

Optimal matching

Mixed integer
programming



Actions

Dasher
assignment



Batching algorithms

Doing more with less



Context

What is batching?

- One Dasher working on more than one delivery simultaneously

Why is it important?

- Increases marketplace efficiency
- Help manage supply demand spikes



On-time delivery to Consumer



Increase marketplace efficiency



Challenges



Orders emerge on the fly

Batching deliveries is easier if orders are all pre-scheduled and have a delivery window



Food Quality

5 extra minutes due to batching could mean a burger is no longer hot



Scale

Optimization system can not exhaustively match all pairs of deliveries



Interdependence

Multiple deliveries being dependent on each other increases variance further



Batching Quality Model

Route of Normal Delivery

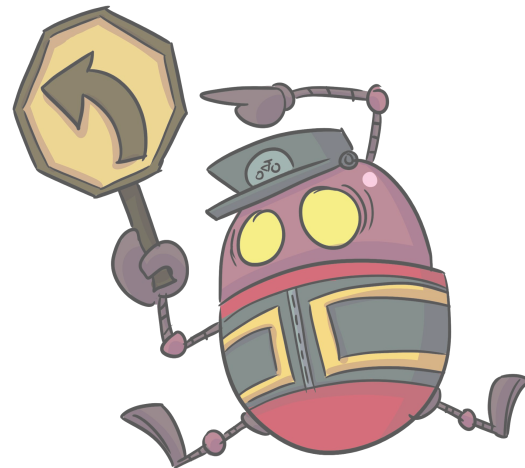
- *Pickup_1, Dropoff_1*

Route of a Batched Delivery

- *Pickup_1, Pickup_2, Dropoff_1, Dropoff_2*

Model Intuition

- Any delay in *Pickup_2* affects *Dropoff_1*
- We should not batch if the probability of delay is high





Model Details

Problem Type

- Classification

Target

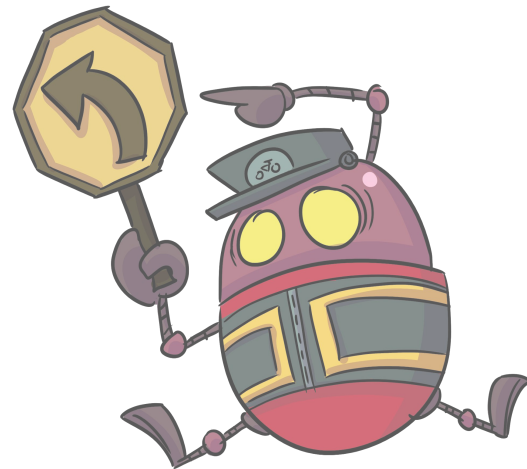
- Boolean: Actual Pickup time > Estimated pickup time

Evaluation Metric

- PR-AUC

Use in optimization system

- Two approaches
 - Used as constraints
 - Use in cost function





Takeaways

- Machine learning helps **efficiently solve** traditional Operations Research problems
- Particularly important in **real-time, high variance** environments like DoorDash
- **Quantile Regression** is an effective technique to generate prediction intervals





The Future

Automated menu processing

Price optimization

Self Driving

Product placements

DoorDash Drive

Robotics

Demand shaping

Dispatch for new verticals

Support experience

Marketing

Fraud prevention



We're hiring!

www.doordash.com/careers

@raghavramesh | raghav@doordash.com



Promo Code:
DOORDASHAI

Thank you!