Taxi Deployment System

An Approach from Machine Learning

Yuxin Zhang, Yixiang Xiao, Xiyan Cai

Contents

01 Problem & Significance

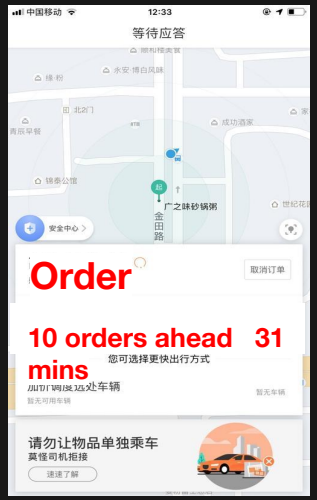
02 Data & Descriptive Analytics

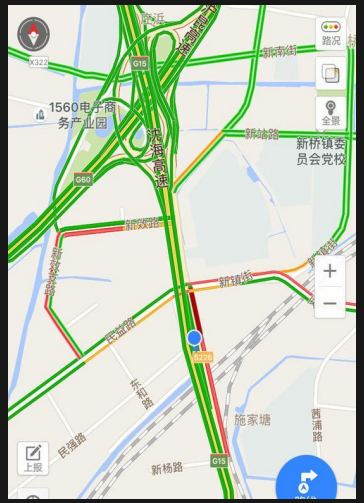
03 ML Approach & Result

04 Future Work

Problem & Significance

Problem with Taxies ⇒ Mission for us & Significance





Long Waiting Hour

Congestion

Essential Issue

Inefficient

Order Prediction
Traffic Prediction

Our Mission

Accurate

Order Prediction
Traffic Prediction

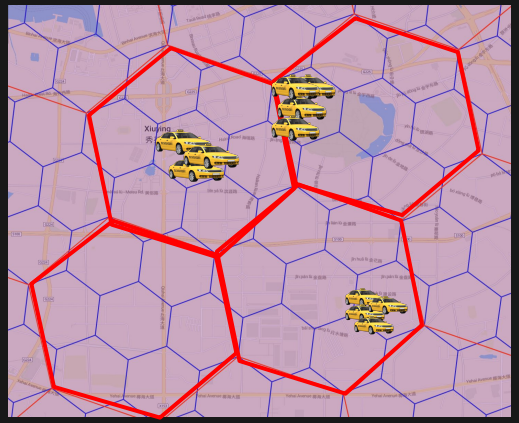
Significance: Maximize Personal Utility & Boost Economic and Social Welfare

Proposed Approach

Order Prediction

Congestion Prediction

Deployment decision



Separately predict due to differences in driving factors & time-series pattern

Data & Descriptive Analysis

Data & Attributes

Order Prediction

Congestion Prediction

Historical Order Details
Order_id
Departure_time
Starting latitude & longitude

Sectional_TTI = avg(Polyline_TTI)

Historical Traffic Details
Polyline Geo data
Time
Polyline_TTI

Geo-Mapping & Aggregation

Sectional Order Statistics
Section_id
Day_of_Week
Hour of Day
Order Num

Weather Indicators
Temperature
Precipitation
Wind_speed

Sectional Congestion Statistics
Section_id
Day_of_Week
Hour of Day
Sectional_TTI

Merge **Merge**

section_idx_large	day_of_week	order_in_section	hour	speed	precipitation	tempurature
874172120fffffff	4	28	20	6.0	0.5	34.0
874172120fffffff	4	37	21	7.0	0.1	32.0
874172120fffffff	4	13	22	7.0	0.1	32.0

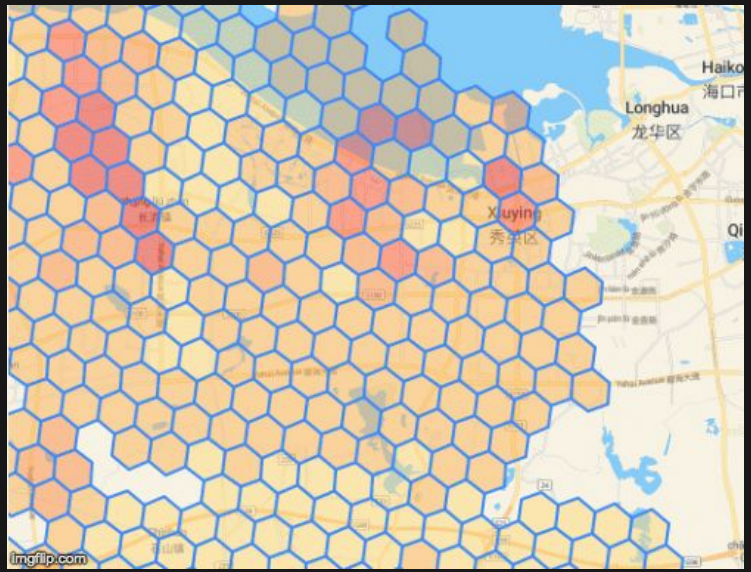
small_section_idx	tti	day_of_week	hour	speed	precipitation	tempurature
88417281d1fffff	1.362	4.0	0	7	0.1	29
88417288bdfffff	1.283	4.0	0	7	0.1	29
8841728ee5fffff	1.125	4.0	0	7	0.1	29

Visualization Analysis

Order dataset

2017/6/15

Congestion dataset



2017/6/15

Preprocessing

Adding Dummy
Day_of_Week
Hour
Section_id

Demand
654080 X 478

Congestion
19104462 X 1128

Cut Dimension & Rows

6 Month → 1 Month

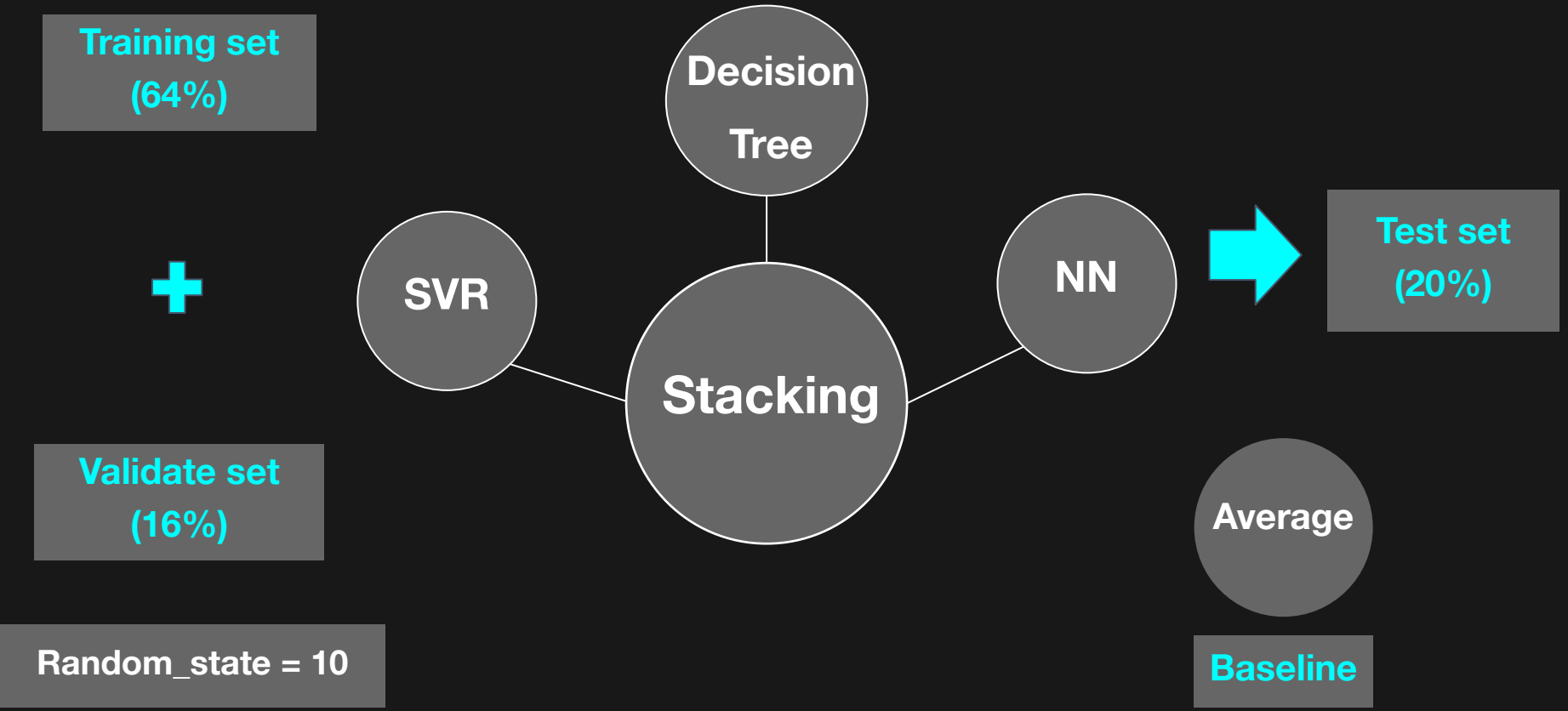
3 District → 1 District

Demand
17555 X 128

Congestion
372973 X 576

ML Approach & Results

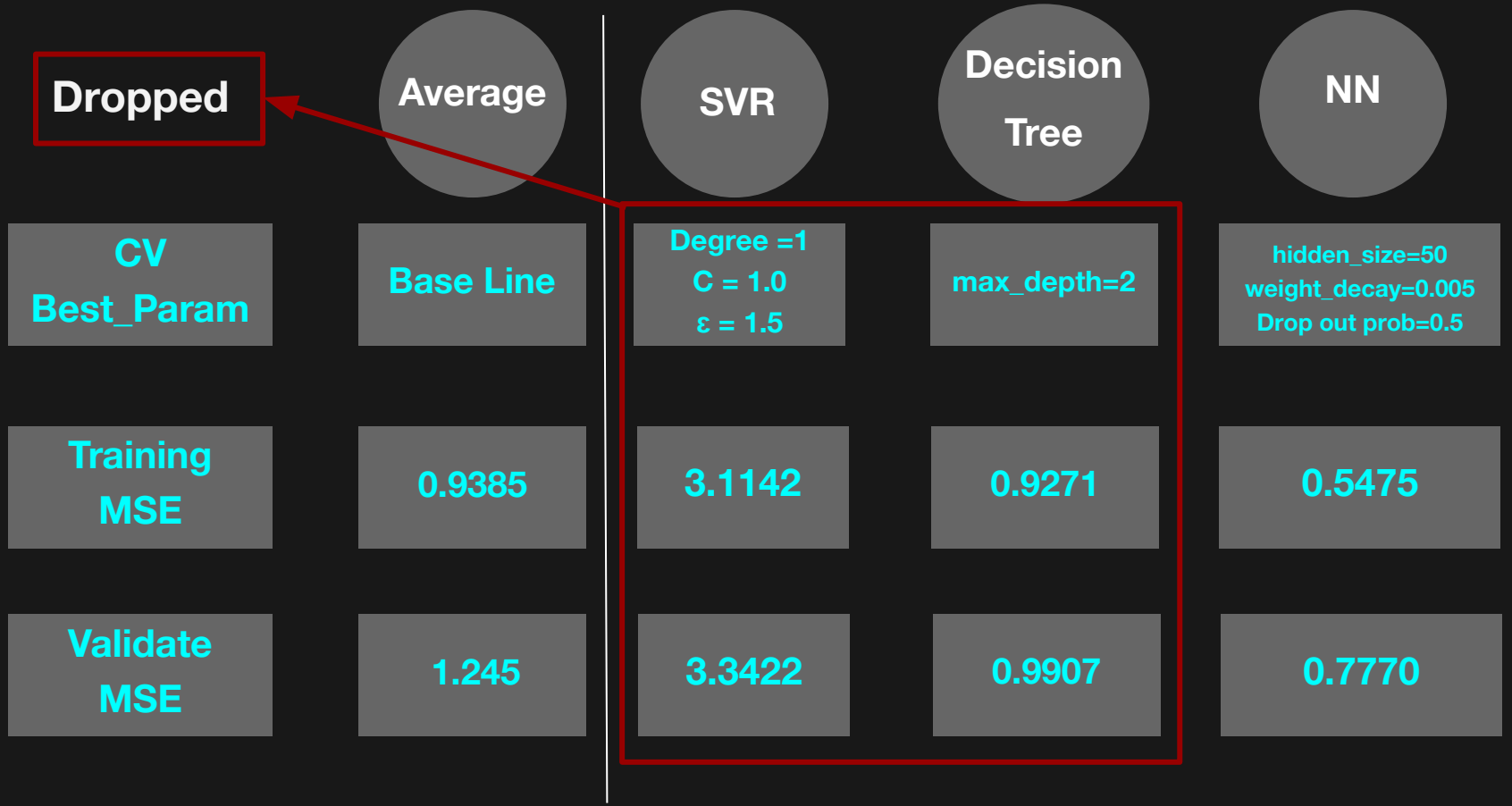
ML Approach Structure



Base Model Training & Performance for Order Prediction

	Average	SVR	Decision Tree	NN
CV Best_Param	Base Line	Degree = 1 C =0.88603 ε = 0.054	max_depth=9	hidden_size=50 weight_decay=0.01
Training MSE	1.0063	0.41963	0.0919	0.0872
Validate MSE	0.9749	0.42098	0.1716	0.1097

Base Model Training & Performance for Congestion Prediction



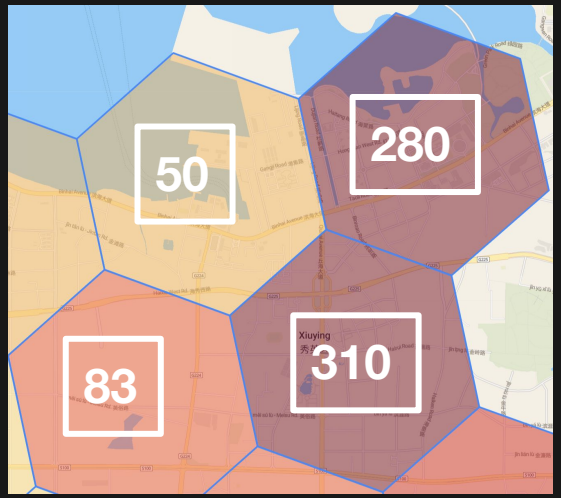
Final Model Training & Performance

	Average	Order [Stacked]	Congestion
Training MSE	1.0063/0.9385	0.0842	0.5475
Validate MSE	0.9749/1.2450	0.1280	0.7770
Test MSE	0.9664/0.8120	0.1230	0.7927
		⬆️90.12%	⬆️36.33%

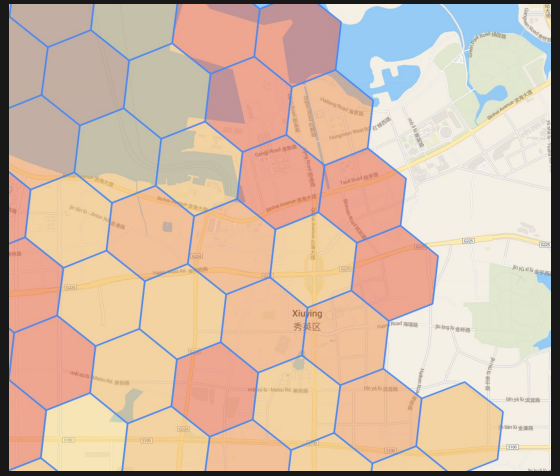
➡️ **Both Models Beat Baseline**

Findings Visualization

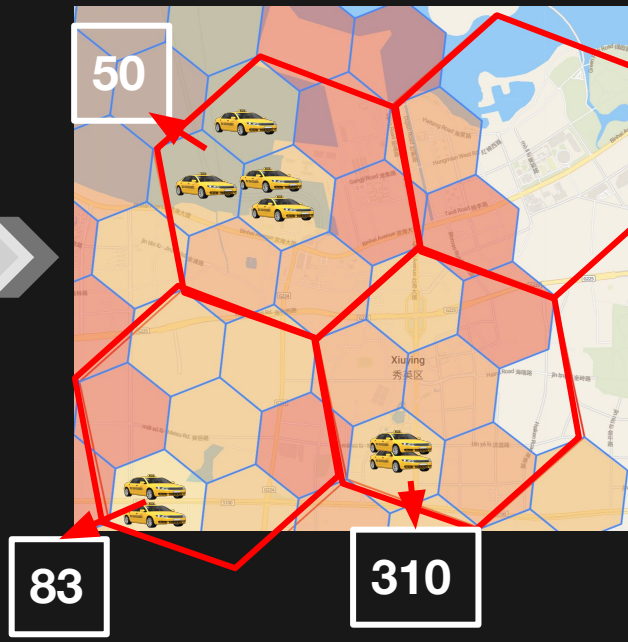
Order Prediction



Congestion Prediction



Deployment Decision



Demo Case: On Thursday, 18:00, with temperature = 40 and wind_speed = 7 and precipitation = 0.0

Future Works

Future Works

Add more data rows to further reduce variance

Use the dataset from other cities and add year-wise month variables to generate a model that can be applied to the whole country

Explore the best section size for models

Explore more base models

Thank You