Traffic Flow Prediction In a U.S. Metropolis

Jingbao Luo

Nanjing University of Science and Technology

Deakin University

2023-01-26

Overview

Problem Definition

Tabular Playground Series - Mar 2022

Data Processing

Data Processing

Data Description

Model Train and Evaluation

Model Train

Feature Importance

Evaluation

Result

Problem Definition

Tabular Playground Series - Mar 2022

Defn

The March edition of the 2022 Tabular Playground Series is a prediction project about time series data.

- We'll forecast twelve-hours of traffic flow in a major U.S. metropolitan area. Time, space, and directional features give us the chance to model interactions across a network of roadways.

Table 1: DATA

File	Description	Attribution
train.csv	traffic congestion from April through September of 1991	row_id,time,x,y, direction,congestion
test.csv	hourly predictions on the day of 1991-09-30	row_id,time,x,y, direction

Data Processing

Data Processing

- Split time data
- Merge x,y,direction data

Data split and merge

- month
- weekday
- minute
- is_month_start
- is_month_end
- is_weekday
- is_Monday
- is_Friday
- period
- road:x+y+direction(00EB)

Period

- Late Night: 0:00-4:00
- Early Morning: 4:00-8:00
- Morning: 8:00-12:00
- Noon: 12:00-16:00
- Evening: 16:00-20:00
- Night: 20:00-24:00

Data Description

Train data decription

	row_id	time	x	y	direction	congestion	month	weekday	hour	minute	is_month_start	is_month_end	period	is_weekend	is_Monday	is_Friday
0	0	1991-04-01	0	0	EB	70	4	0	0	0	1	0	Late Night	0	1	0
1	1	1991-04-01	0	0	NB	49	4	0	0	0	1	0	Late Night	0	1	0
2	2	1991-04-01	0	0	SB	24	4	0	0	0	1	0	Late Night	0	1	0
3	3	1991-04-01	0	1	EB	18	4	0	0	0	1	0	Late Night	0	1	0
4	4	1991-04-01	0	1	NB	60	4	0	0	0	1	0	Late Night	0	1	0

Figure 1: train_data

Congestion data

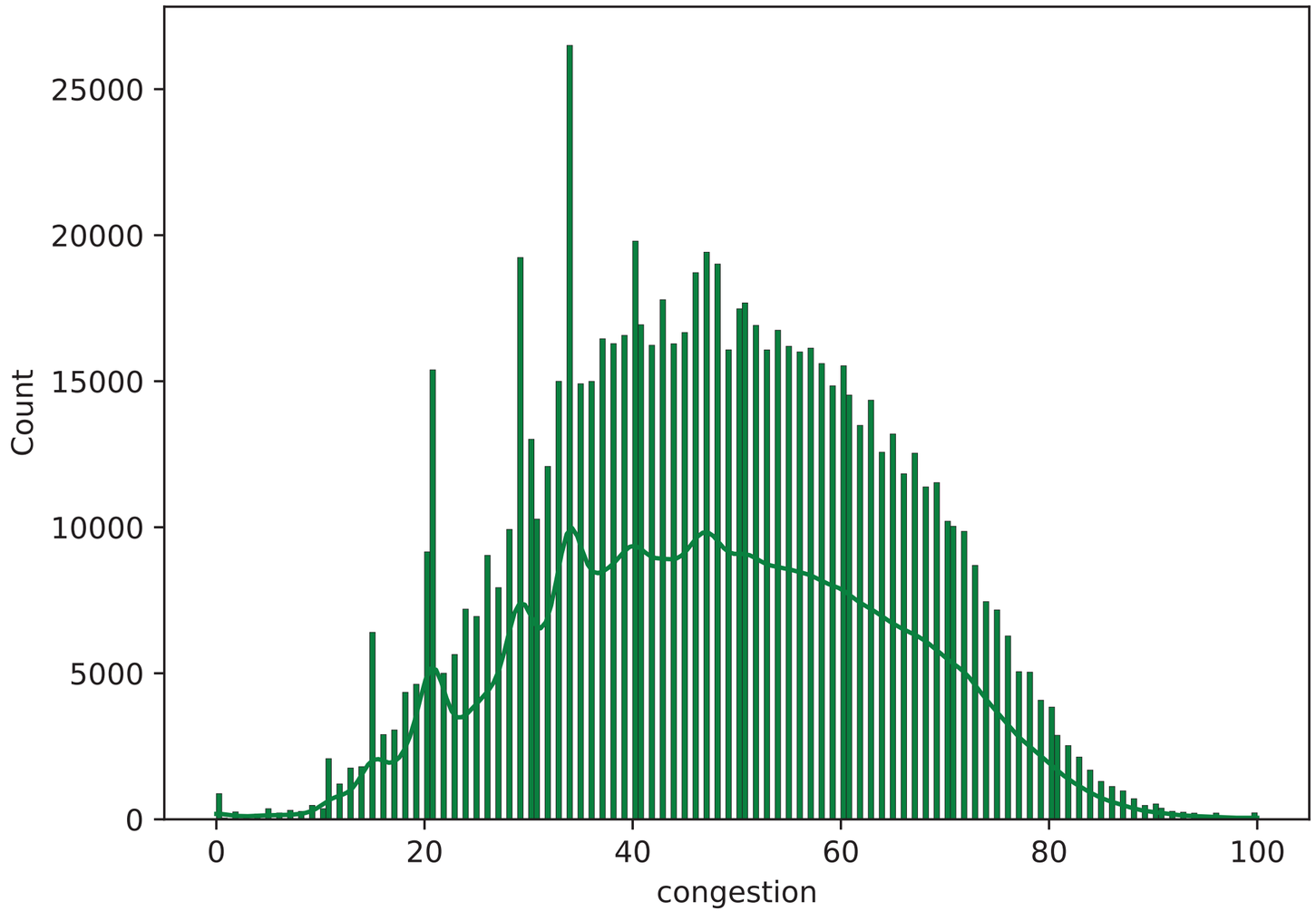


Figure 2: Congestion data

The effect of weekday on congestion

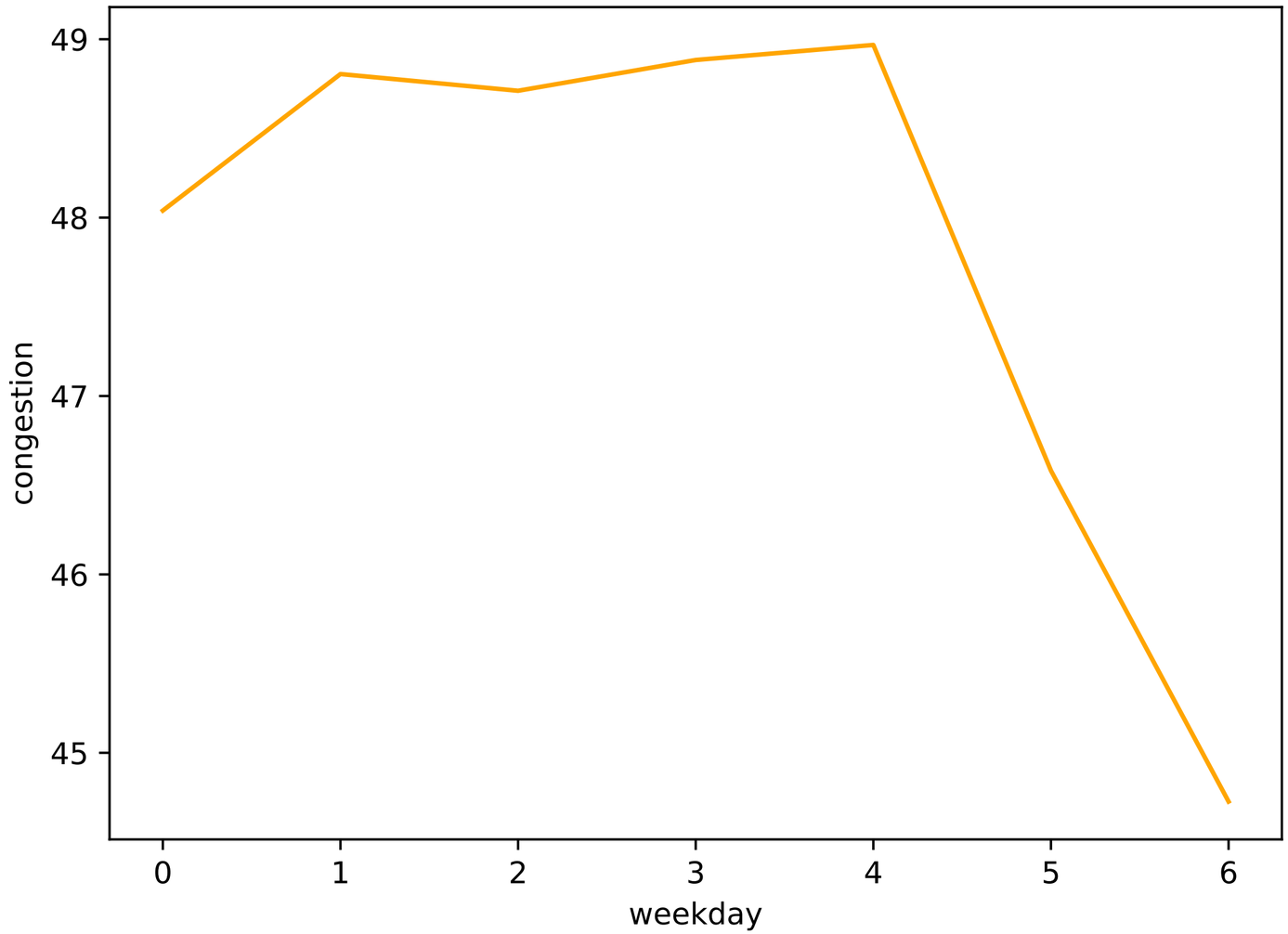


Figure 3: The effect of weekday on congestion

Congestion in special day or not

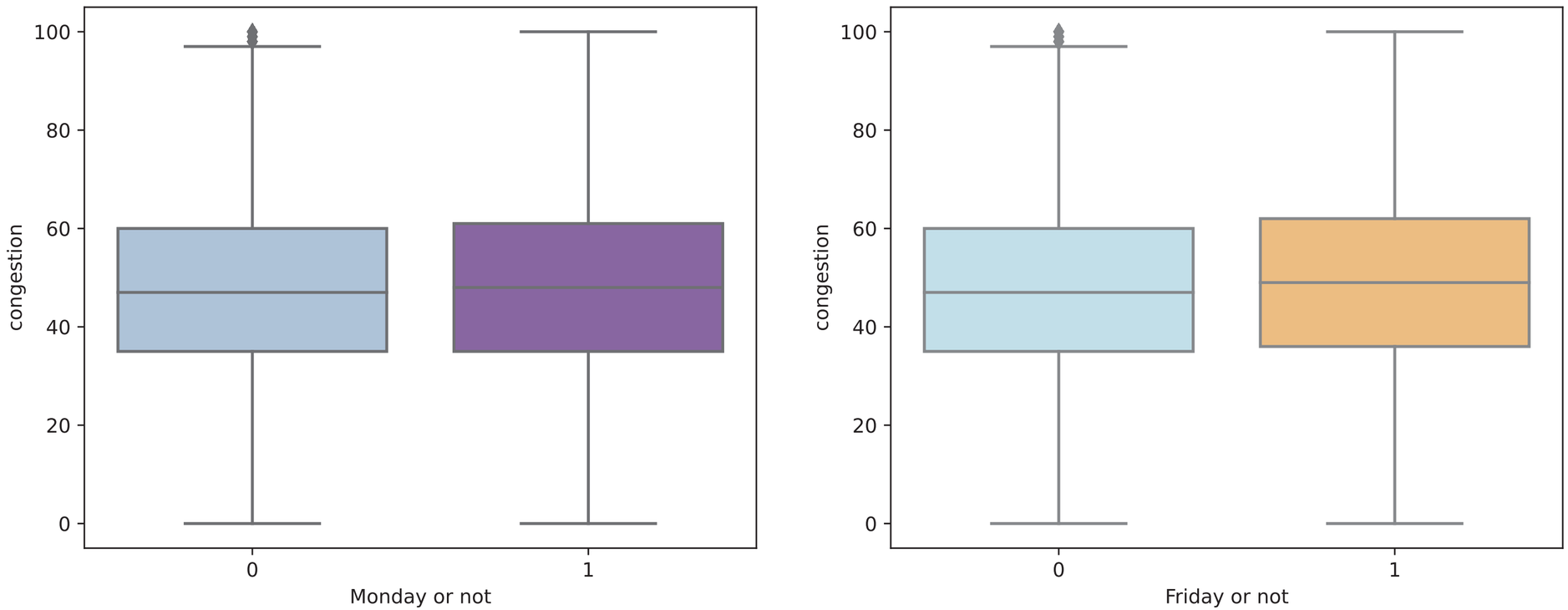


Figure 4: Congestion in special day or not

The effect of road on congestion

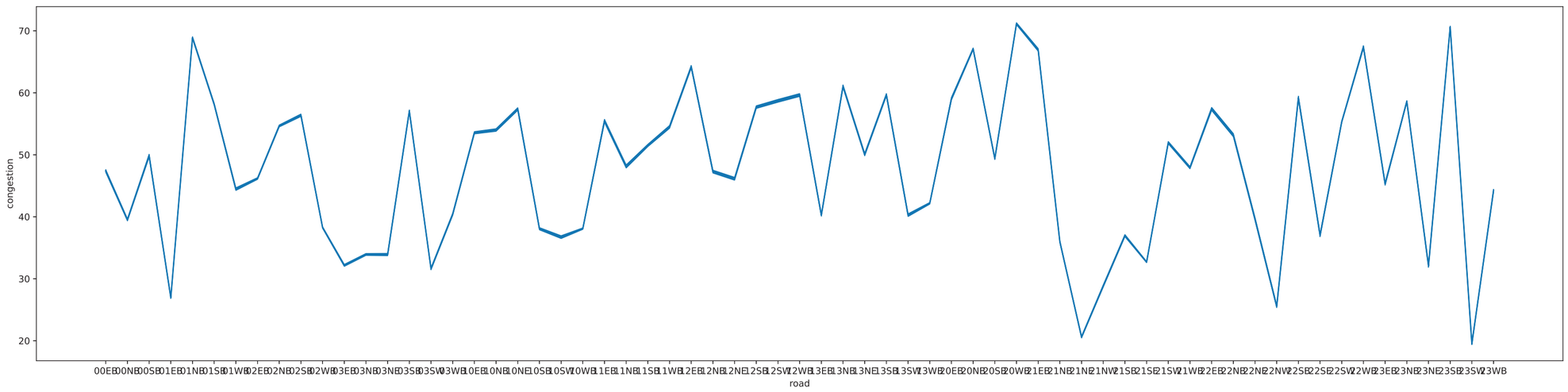


Figure 5: The effect of road on congestion

The effect of day on congestion group by month

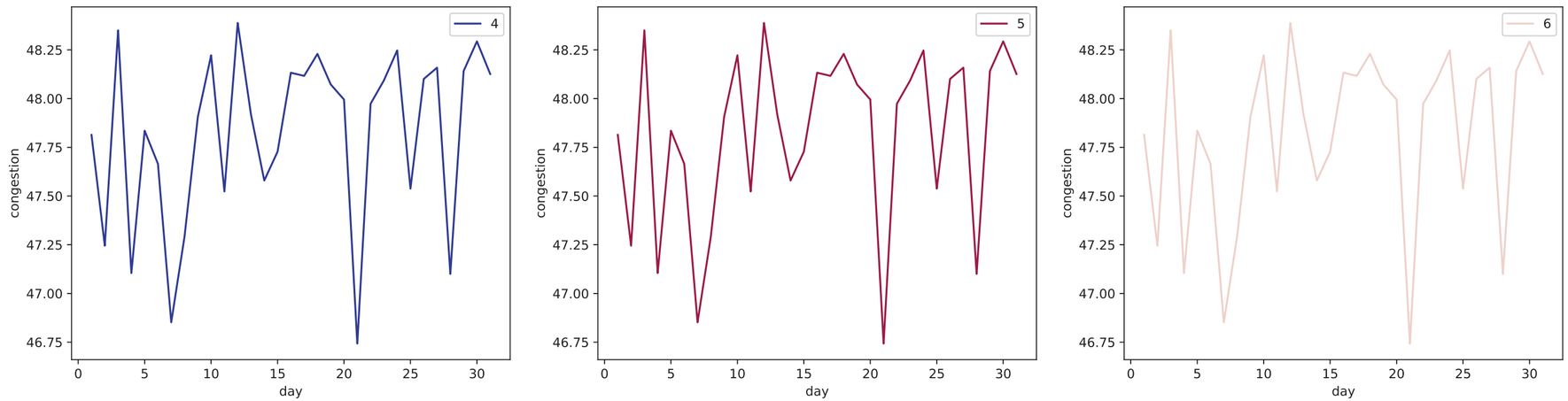


Figure 6: The effect of day on congestion group by month

The effect of hour on congestion group by road

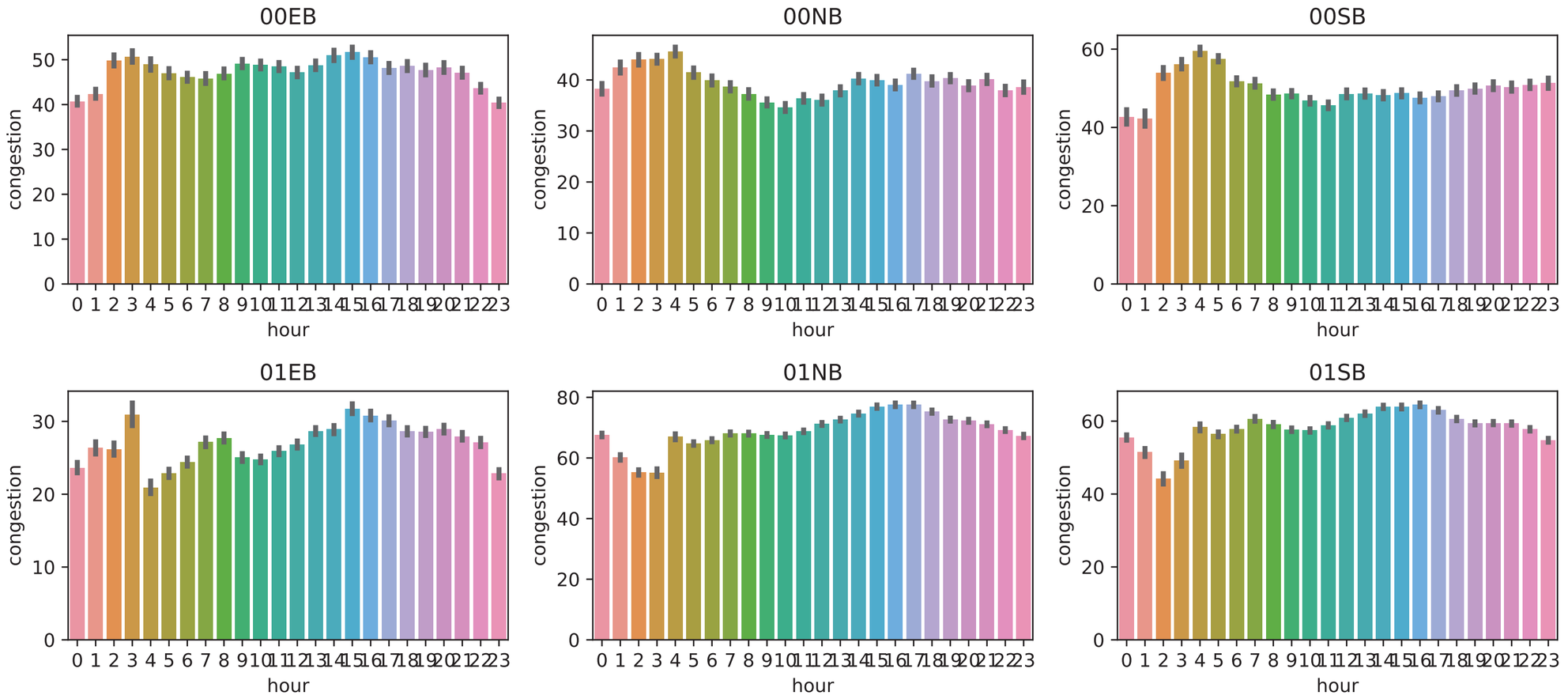


Figure 7: The effect of hour on congestion group by road

The effect of weekday on congestion group by hour

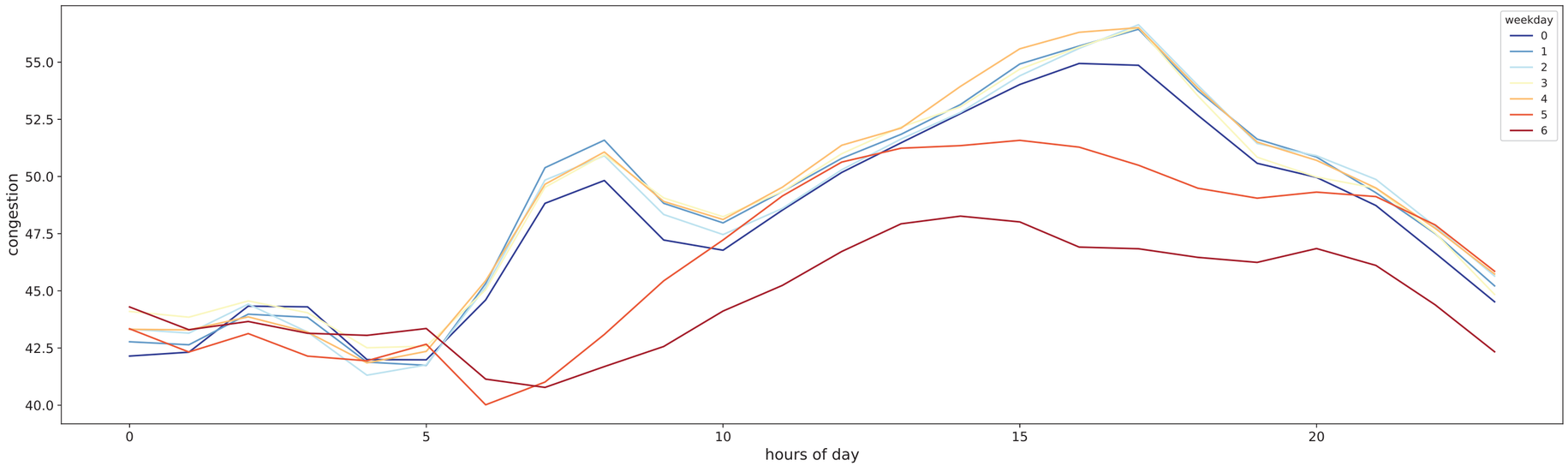


Figure 8: The effect of weekday on congestion group by hour

The effect of period on congestion

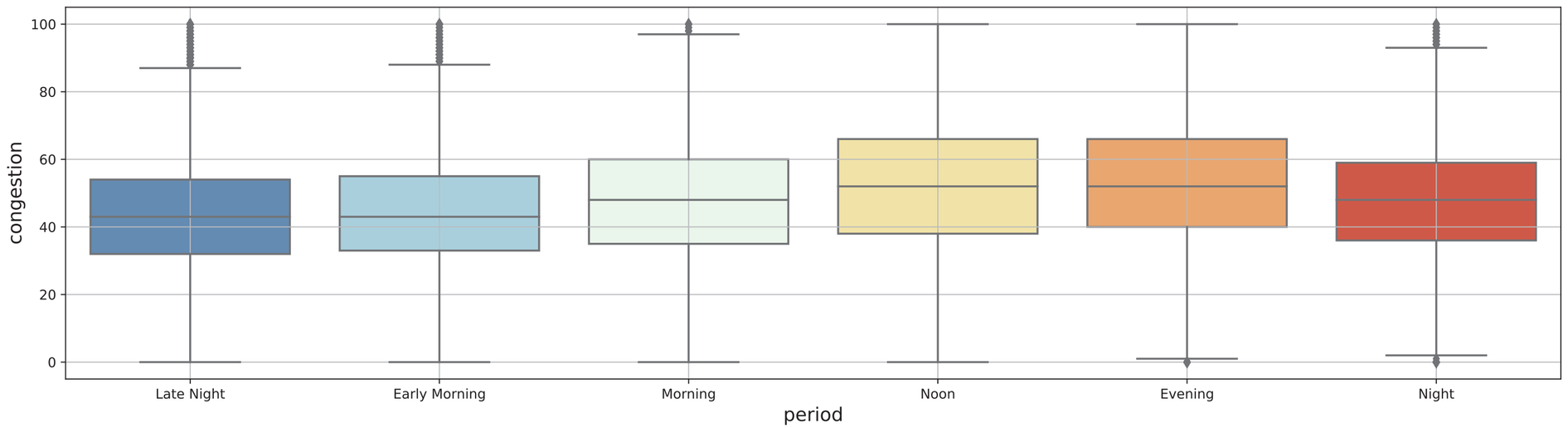


Figure 9: The effect of period on congestion

Model Train and Evaluation

Model Train

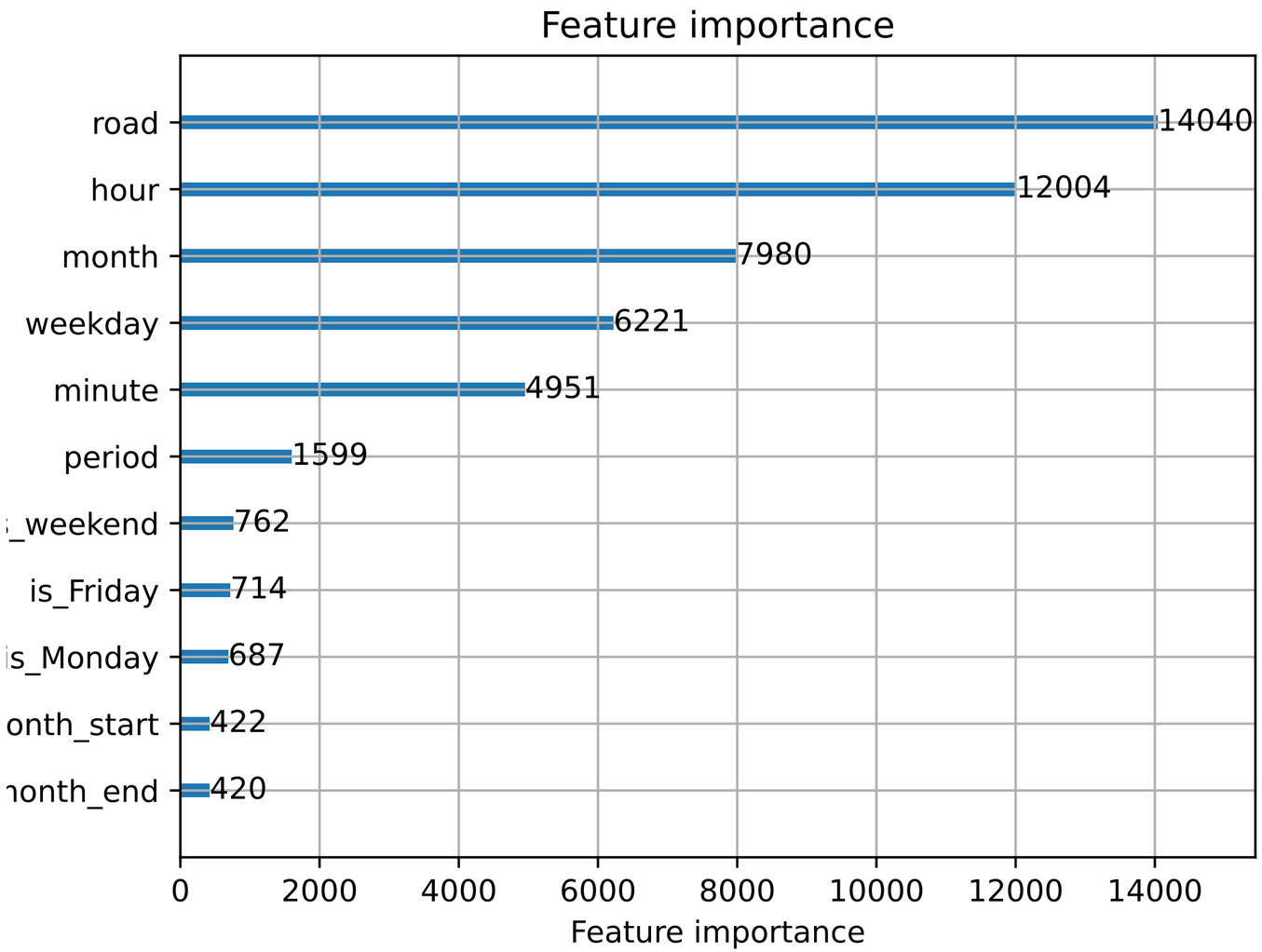
- Train

  Model: LightGBM(GPU)

  Data: train data(fit_transform)

  Train_test_split:x_train, x_eval, y_train, y_eval

Feature Importance



Evaluation

Table 2: Model evaluation indexes

Index	Eexplication
explained_variance_score	Explain the variance score of the regression model.
mean_absolute_error	Assess the proximity of the predicted results to the real data set.
Mean squared error	Calculate the mean value of the square sum of the errors of the corresponding sample points of the fitting data and the original data
r2_score	Judge the fitting degree of prediction model and real data

Evaluation Result

The evaluation results are as follows:

Table 3: The Evaluation results

Index	Result
explained_variance_score	0.7277243544483329
mean_absolute_error	6.167491947603395
r2_score	0.7277251135484366

Result

Prediction Result

Table 4: The prediction results

Row_id	Congestion	Row_id	Congestion
848835	47	848836	33
848837	39	848838	54
848839	64	848840	23
848841	28	848842	70
848843	25	848844	47
848845	46	848846	25
848847	69	848848	60

Contact Information

Associate Professor Jingbao Luo
School of Economics and Management
Nanjing University of Science and Technology, China

JLUO@TULIP.ORG.AU

TEAM FOR UNIVERSAL LEARNING AND INTELLIGENT PROCESSING