



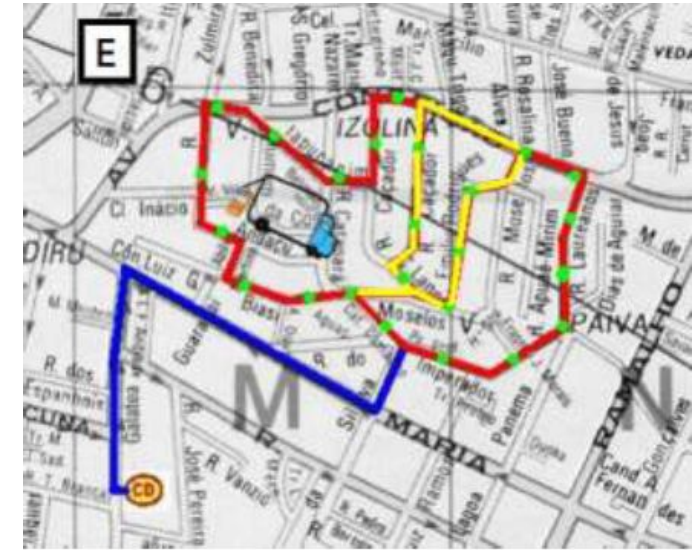
Behavior of the urban traffic of the city of Sao Paulo in Brazil

- Nature du problème
- Influence des paramètres
- Modèles de régression
- Modèle de classification

DURAND Lénéïc, SHI DE MILLEVILLE Guillaume

Nature du problème

- ❑ Présentation du problème
- ❑ Présentation des paramètres
- ❑ Présentation des objectifs
- ❑ Problématique : Trouver un modèle de prédiction permettant de prédire la lenteur du trafic en fonction de paramètres donnés.
- ❑ Variables les plus influentes sur Slowness in traffic



Zone où les données ont été récoltées

Influence des paramètres

- ❑ Nécessité de formater les données
- ❑ Pas de « trou » dans les données
- ❑ Variables numériques
- ❑ Le surplus de 0 dans la base ne permet pas de calculer directement des résultats sur la corrélation des paramètres.

1	Hour (Coded)
2	1
3	2
4	3
5	4
6	5
7	6
8	7
9	8
10	9
11	10
12	11
13	12
14	13
15	14
16	15
17	16
18	17
19	18
20	19
21	20
22	21
23	22
24	23
25	24
26	25
27	26
28	27

Influence des paramètres

On effectue une moyenne de Slowness in traffic sur le graphe de gauche pour chaque heure donné.

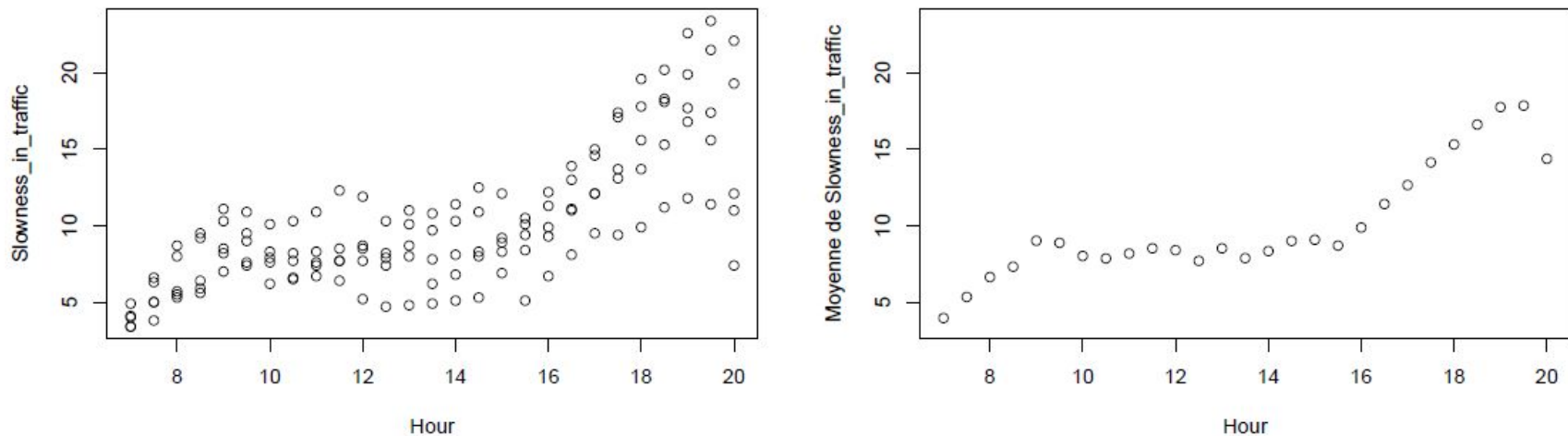


Figure 1 : influence de Hour sur Slowness_in_traffic

Influence des paramètres

On effectue une moyenne de Slowness in traffic sur le graphe de gauche pour chaque feu de circulation en panne donné.

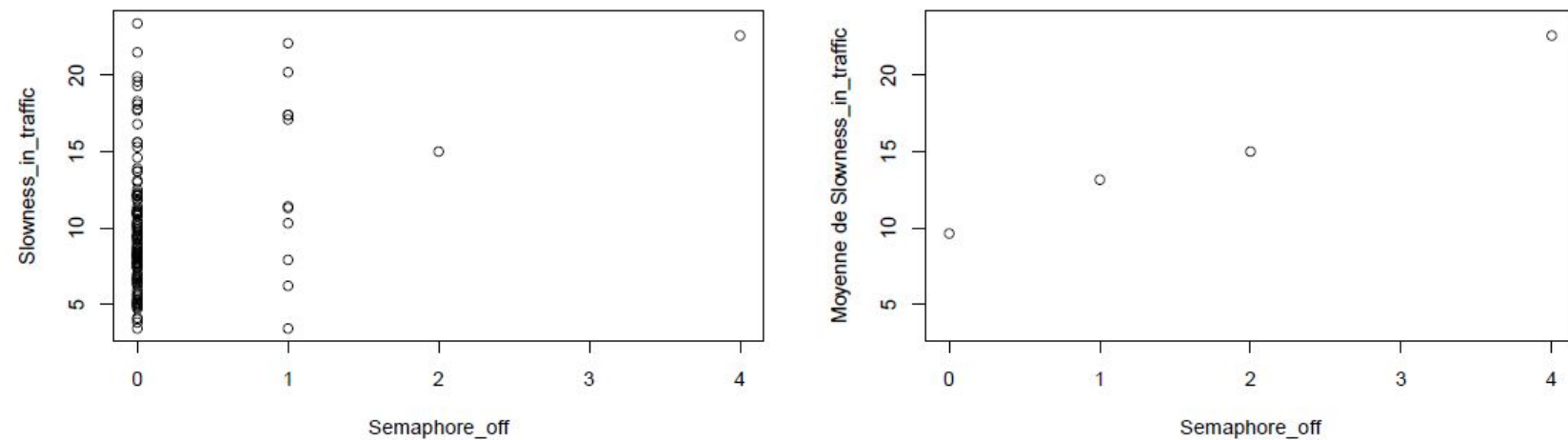
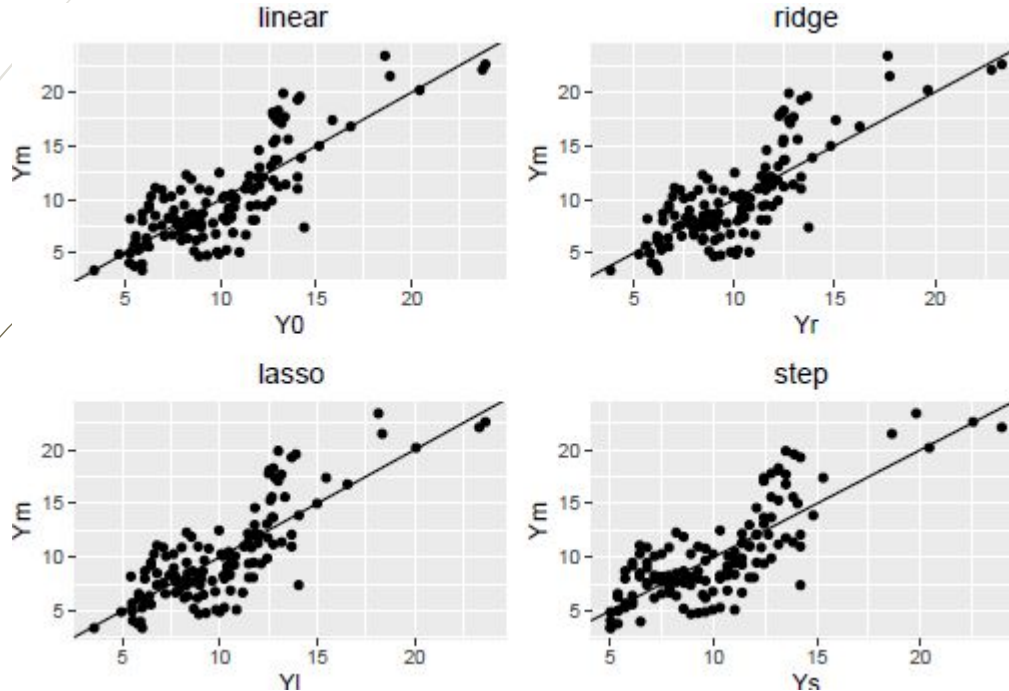


Figure 2 : influence de Semaphore_off sur Slowness_in_traffic

Modèles de régression : résultats



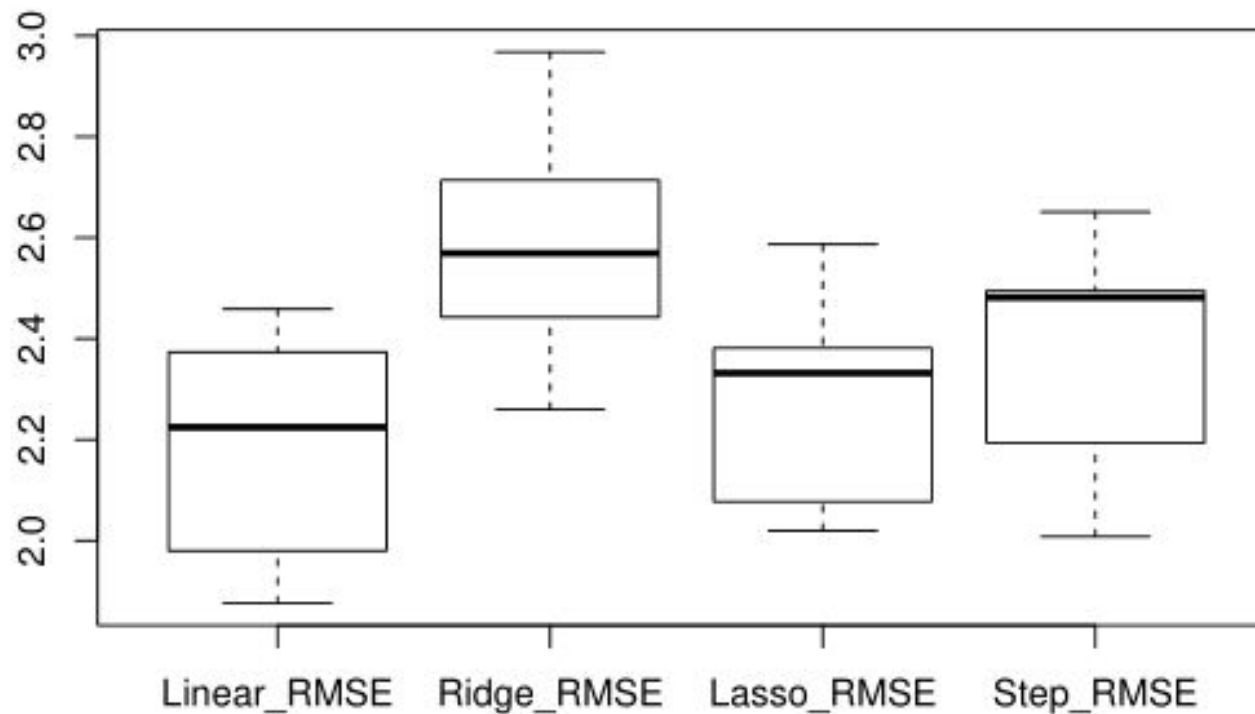
Régressions $Y_{\text{estimé}}$ en fonction de Y

Linear_R2	Ridge_R2	Lasso_R2	Step_R2
0.657544	0.6555854	0.6570989	0.6427598

Linear_RMSE	Ridge_RMSE	Lasso_RMSE	Step_RMSE
2.543884	2.575741	2.550845	2.598215

Les valeurs variant de 0 à 27, la RMSE représente une erreur de 10% pour les valeurs les plus hautes, et de 25% en moyenne

Modèles de régression : résultats



RMSE des modèles après application du k-fold

Modèle de classification

Matrice de confusion

	Ylog	
Y_binaire	0	1
0	96	5
1	10	24

Résultats

recall	precision	F1_score
0.8275862	0.7058824	0.7619048

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}}$$

$$\text{Accuracy} = \frac{96+24}{96+24+10+5} = 89\%$$

Conclusion sur le modèle conservé

- Résultats des différents modèles
- Qualités et défauts des modèles retenus
- Limites du modèle et possibles améliorations