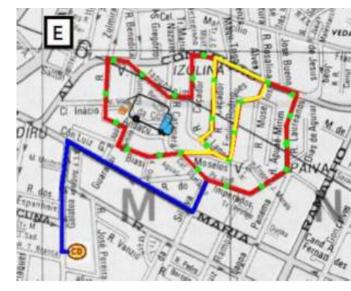
Behavior of the urban traffic of the city of Sao Paulo in Brazil

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

DURAND Lénaïc, 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

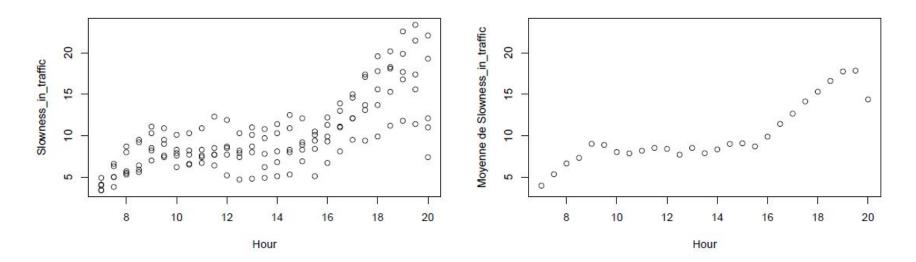
Influence des paramètres

- 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.

3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 17 18 19 20 19 21 22 23 24 22 25 24 26 27 26 27 2	1	Hour (Coded
4 5 6 7 8 9 10 11 11 11 12 11 13 11 14 11 15 11 16 11 17 11 18 11 19 11 12 12 12 12 12 12 12 12 12 12 12 12	2	1
5 6 7 8 9 10 11 11 11 11 12 11 13 11 15 11 16 11 17 11 18 11 19 11 12 12 12 12 12 12 12 12 12 12 12 12	3	2
6 7 8 9 10 11 11 11 12 11 13 11 14 11 15 16 11 17 19 11 19 11 19 12 12 12 12 12 12 12 12 12 12 12 12 12	4	3
7 8 9 10 11 11 12 13 14 15 16 17 18 17 18 17 19 20 21 22 22 23 24 22 25 26 27 26 27 20 27 20 20 20 20 20 20 20 20 20 20 20 20 20	5	4
8 9 10 11 11 12 13 14 15 16 17 18 11 19 20 21 22 23 24 25 24 25 26 27 21 22 22 23	6	5
9 10 11 11 11 11 12 11 13 11 14 11 15 16 11 17 11 18 11 19 11 12 12 12 12 12 12 12 12 12 12 12 12	7	6
10	8	7
11 10 12 1 13 1 14 1 15 1 16 1 17 1 18 1 19 1 20 1 21 2 22 2 23 2 24 2 25 2 26 2 27 2	9	8
12 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	10	g
13 1: 1: 14 1: 15 1: 16 1: 17 1: 18 1: 19 1: 19 1: 20 1: 21 2: 22 2: 22 2: 24 2: 25 2: 26 2: 27 2: 26 2: 27	11	10
14 1. 15 1. 16 1. 17 1. 18 1. 19 1. 20 1. 21 2. 22 2. 23 2. 24 2. 25 2. 26 2. 27 2.	12	11
15 1.1 16 1.1 17 1.0 18 1.1 19 1.1 20 1.1 21 2.1 22 2.2 23 2.2 24 2.2 25 2.2 26 2.2 27 2.2	13	12
16 11 17 12 18 12 19 13 20 12 21 22 22 2 23 22 24 22 25 26 22 27 26	14	13
17 10 10 10 10 10 10 10 10 10 10 10 10 10	15	14
18 1 19 1 20 1 21 2 22 2 23 2 24 2 25 2 26 2 27 2		15
19 13 20 15 21 22 22 2 23 2: 24 2 25 2: 26 2: 27 2:	17	16
20 11 21 22 22 2 23 2: 24 2 25 26 27 2:	18	17
21 21 22 2 23 2 24 2: 25 2: 26 2: 27 2:	19	18
22 2 23 2 24 2 25 2 26 2 27 2	20	19
23 2. 24 2. 25 2. 26 2. 27 2.	21	20
24 2. 25 2. 26 2. 27 2.	22	21
25 24 26 25 27 26	23	22
26 25 27 26	24	23
27 2	25	24
	26	25
28 2	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.

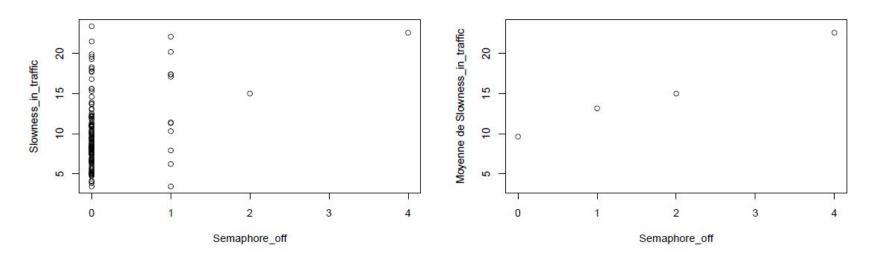
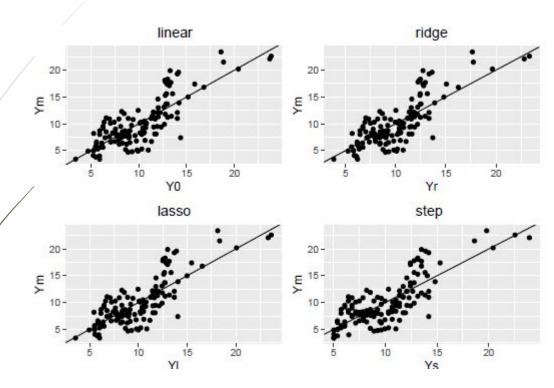


Figure 2: influence de Semaphore off sur Slowness in traffic

Modèles de régression : résultats



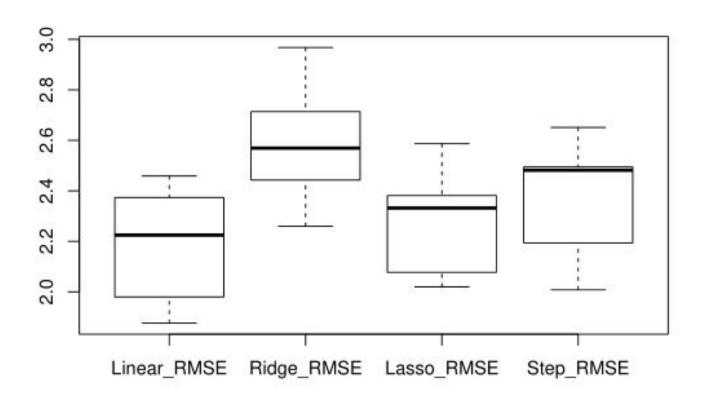
Régressions Y_estimé en fonction de Y

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

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

Résultats

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