

ÉTUDE SUR LA DÉTECTION DE FAUSSE MONNAIES

W. MÉHEUST R. RAGAVAN

Apprentissage Statistique, 22 Mars 2023



Table des matières

- 1 Introduction
- 2 Analyse descriptive
- 3 Prédiction
- 4 Conclusion



Table des matières

1 Introduction

2 Analyse descriptive

3 Prédiction

4 Conclusion

Introduction

is_genuine <lgl>	diagonal <dbl>	height_left <dbl>	height_right <dbl>	margin_low <dbl>	margin_up <dbl>	length <dbl>
TRUE	171.81	104.86	104.95	4.52	2.89	112.83
TRUE	171.46	103.36	103.66	3.77	2.99	113.09
TRUE	172.69	104.48	103.50	4.40	2.94	113.16
TRUE	171.36	103.91	103.94	3.62	3.01	113.51
TRUE	171.73	104.28	103.46	4.04	3.48	112.54
TRUE	172.17	103.74	104.08	4.42	2.95	112.81

6 rows

Figure: Dataset



Table des matières

1 Introduction

2 Analyse descriptive

3 Prédiction

4 Conclusion

Valeurs manquantes

- 37 valeurs manquantes toutes dans la variable *margin_low*.

Méthodes pour remplacer les données manquantes :

- Remplacer par la moyenne
- Remplacer par la medianne
- Remplacer par la catégorie dominante
- Créer une nouvelle catégorie
- Faire de la régression linéaire
- Faire du k-mean

Régression linéaire :

- méthode plus poussée
- variables numériques



Choix du GLM

```

Start: AIC=2955.87          Step: AIC=2097.64          Step: AIC=2052.37
margin_low ~ 1             margin_low ~ length        margin_low ~ length + height_right

      Df Deviance   AIC      Df Deviance   AIC      Df Deviance   AIC
+ length      1  357.83 2097.6 + height_right 1  346.45 2052.4 + margin_up      1  342.03 2035.6
+ margin_up    1  524.22 2656.3 + margin_up    1  351.57 2073.8 + height_left     1  342.11 2035.9
+ height_right 1  545.69 2715.0 + height_left 1  351.64 2074.1 + diagonal        1  344.99 2048.2
+ height_left  1  585.22 2817.3 + diagonal    1  356.55 2094.4 <none>           346.45 2052.4
+ diagonal     1  636.21 2939.6 <none>         357.83 2097.6 - height_right    1  357.83 2097.6
<none>         644.23 2955.9 - length          1  644.23 2955.9 - length          1  545.69 2715.0

Step: AIC=2035.59          Step: AIC=2021.84
margin_low ~ length + height_right + margin_up  margin_low ~ length + height_right + margin_up + height_left

      Df Deviance   AIC      Df Deviance   AIC
+ height_left  1  338.37 2021.8 + diagonal    1  336.71 2016.7
+ diagonal     1  340.62 2031.5 <none>         338.37 2021.8
<none>         342.03 2035.6
- margin_up    1  346.45 2052.4
- height_right 1  351.57 2073.8
- length       1  476.64 2519.1
- height_left  1  342.03 2035.6
- margin_up    1  342.11 2035.9
- height_right 1  346.50 2054.6
- length       1  460.40 2470.4

Step: AIC=2016.65
margin_low ~ length + height_right + margin_up + height_left +
diagonal

      Df Deviance   AIC
<none>         336.71 2016.7
- diagonal     1  338.37 2021.8
- margin_up    1  340.37 2030.5
- height_left  1  340.62 2031.5
- height_right 1  344.97 2050.1
- length       1  455.03 2455.2

```

Figure: AIC



Choix du GLM

Start: AIC=2961.16	Step: AIC=2108.21	Step: AIC=2068.24
margin_low ~ 1	margin_low ~ length	margin_low ~ length + height_right
Df Deviance AIC	Df Deviance AIC	Df Deviance AIC
+ length 1 357.83 2108.2	+ height_right 1 346.45 2068.2	+ margin_up 1 342.03 2056.7
+ margin_up 1 524.22 2666.9	+ margin_up 1 351.57 2089.7	+ height_left 1 342.11 2057.1
+ height_right 1 545.69 2725.6	+ height_left 1 351.64 2090.0	<none> 346.45 2068.2
+ height_left 1 585.22 2827.9	<none> 357.83 2108.2	+ diagonal 1 344.99 2069.3
+ diagonal 1 636.21 2950.1	+ diagonal 1 356.55 2110.3	- height_right 1 357.83 2108.2
<none> 644.23 2961.2	- length 1 644.23 2961.2	- length 1 545.69 2725.6
Step: AIC=2056.74	Step: AIC=2048.28	
margin_low ~ length + height_right + margin_up	margin_low ~ length + height_right + margin_up + height_left	
Df Deviance AIC	Df Deviance AIC	
+ height_left 1 338.37 2048.3	<none> 338.37 2048.3	
<none> 342.03 2056.7	+ diagonal 1 336.71 2048.4	
+ diagonal 1 340.62 2058.0	- height_left 1 342.03 2056.7	
- margin_up 1 346.45 2068.2	- margin_up 1 342.11 2057.1	
- height_right 1 351.57 2089.7	- height_right 1 346.50 2075.7	
- length 1 476.64 2535.0	- length 1 460.40 2491.5	

Figure: BIC



Visualisation

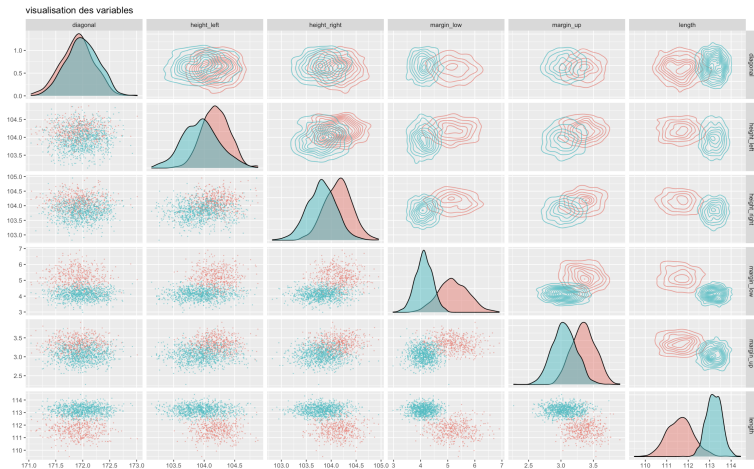


Figure: Pairplot

Visualisation

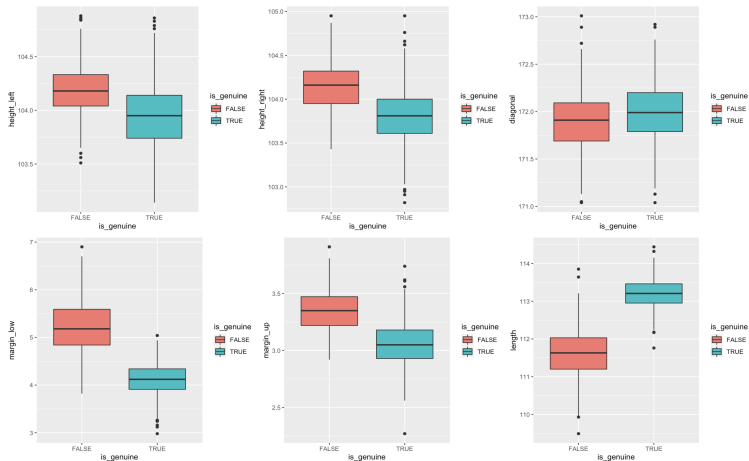


Figure: Boxplot

Table des matières

1 Introduction

2 Analyse descriptive

3 **Prédictions**

4 Conclusion

Benchmark des modèles

modele <chr>	accuracy <dbl>	f1 <dbl>	recall <dbl>	precision <dbl>
regression logistique	0.9973333	1	1	0.996
1 row				

Figure: Regression logistique ; AUC = 0.996

modele <chr>	accuracy <dbl>	f1 <dbl>	recall <dbl>	precision <dbl>
tree	0.9866667	1	0.9959839	0.984127
1 row				

Figure: Arbre de classification ; AUC = 0.982

modele <chr>	accuracy <dbl>	f1 <dbl>	recall <dbl>	precision <dbl>
svm	0.9946667	1	1	0.9920319
1 row				

Figure: SVM ; AUC = 0.992



Benchmark des modèles

modele <chr>	accuracy <dbl>	f1 <dbl>	recall <dbl>	precision <dbl>
random Forest	0.992	1	0.9959839	0.992
1 row				

Figure: Random forest ; AUC = 0.996

modele <chr>	accuracy <dbl>	f1 <dbl>	recall <dbl>	precision <dbl>
knn	0.9946667	1	1	0.9920319
1 row				

Figure: KNN ; AUC = 0.99

Table des matières

- 1 Introduction
- 2 Analyse descriptive
- 3 Prédiction
- 4 Conclusion**

Bon résultats sur l'ensemble des méthodes.

Les deux meilleurs modèles en terme d'AUC ;

- Random forest
 - prédit à tort que certains billets sont vrais alors qu'ils sont faux.
 - méthode de boîte noire, difficilement interprétable
- Régression logistique
 - *Recall* plus élevé
 - facilement interprétable
 - sensible aux outliers

Merci de votre attention !

