



PREDICT SOCIAL MEDIA PERFORMANCE AND EVALUATING THE IMPACT ON BRAND BUILDING



Ahmed TRAORE & Kaldini KAH

Présentation de la dataset



Un dataset contenant des données sur les publications d'une page Facebook



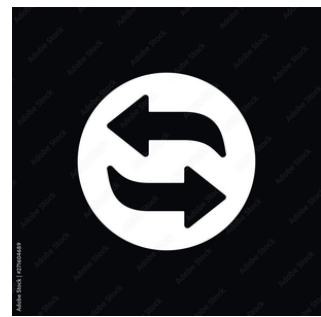
500 observations et 19 variables dont 12 de performances et 7 utilisées
comme entrées pour prédire les autres

- **Page total likes** : Le nombre total de likes de la page.
- **Type** : Le type de publication (par exemple, Photo, Status, Link, Video).
- **Category** : La catégorie de la publication.
- **Post Month** : Le mois de la publication.
- **Post Weekday** : Le jour de la semaine de la publication.
- **Post Hour** : L'heure de la publication.
- **Paid** : Indique si la publication est sponsorisée (1) ou non (0).
- **Lifetime Post Total Reach**: Le nombre de personnes qui ont vu une publication de la page (utilisateurs uniques).
- **Lifetime Post Total Impressions**: Le nombre de fois qu'une publication de la page est affichée, que la publication soit cliquée ou non. Les personnes peuvent voir plusieurs impressions de la même publication.
- **Lifetime Engaged Users**: Le nombre de personnes qui ont cliqué n'importe où dans une publication (utilisateurs uniques).
- **Lifetime Post Consumers**: Le nombre de personnes qui ont cliqué n'importe où dans une publication.
- **Lifetime Post Consumptions**: Le nombre de clics n'importe où dans une publication.
- **Lifetime Post Impressions by people who have liked your Page**: Le nombre total d'impressions uniquement des personnes qui ont aimé une page.
- **Lifetime Post reach by people who like your Page**: Le nombre de personnes qui ont vu une publication de la page parce qu'elles ont aimé cette page (utilisateurs uniques).
- **Lifetime People who have liked your Page and engaged with your post**: Le nombre de personnes qui ont aimé une page et cliqué n'importe où dans une publication (utilisateurs uniques).
- **comment**: Nombre de commentaires sur la publication.
- **like**: Nombre de "Likes" sur la publication.
- **share**: Nombre de fois que la publication a été partagée.
- **Total Interactions**: La somme des "likes", "commentaires" et "partages" de la publication.

Prétraitement et analyse exploratoire



Nettoyage des données, les variables Paid, like et share avaient respectivement 499, 499 et 496 lignes

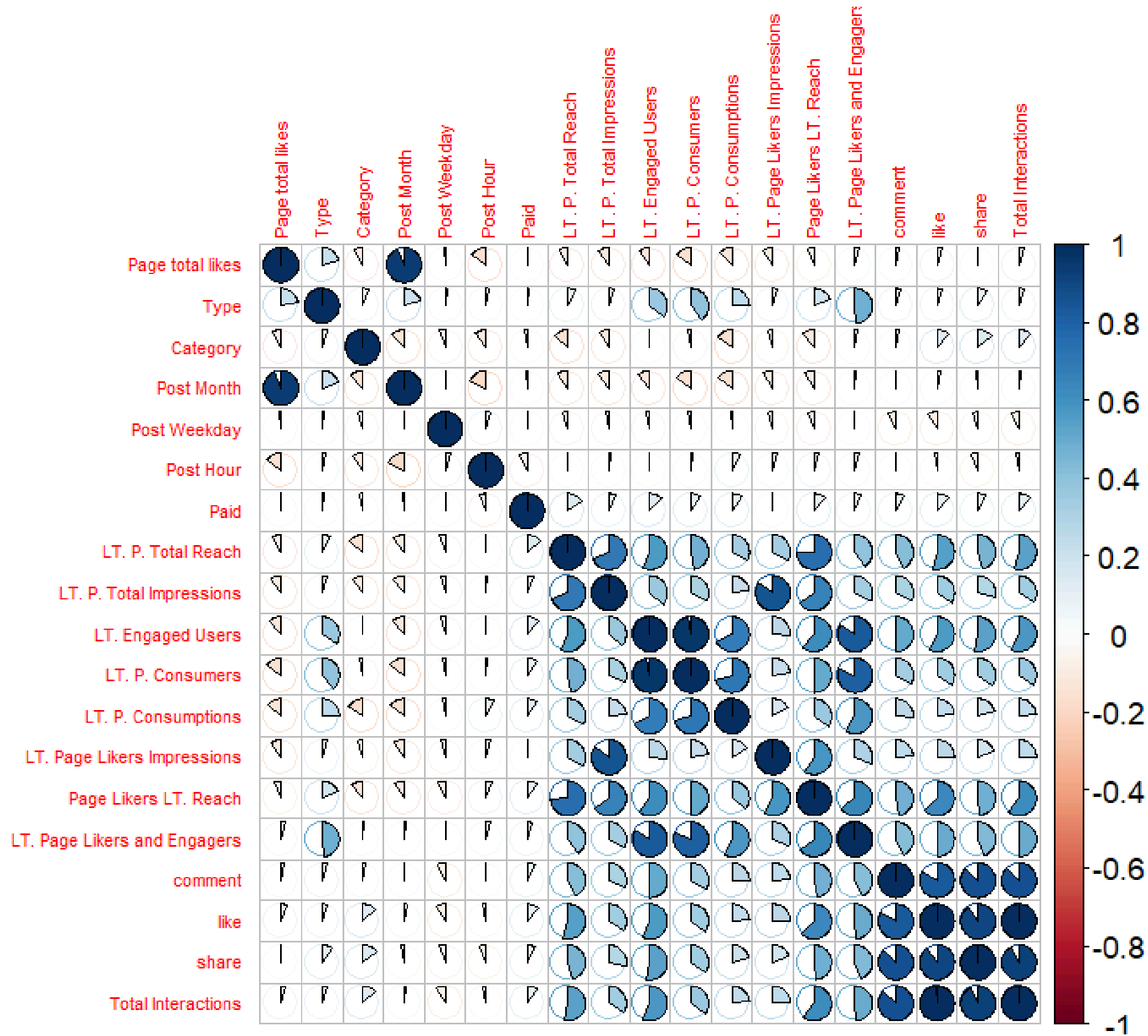


Conversion du type de la variable 'Type' en binaire



On s'est assuré que toutes les variables avaient le même nombre de ligne et qu'il n'y avait aucune valeur manquante

Choix de la variable cible



Lifetime.Engaged.Users
RMSE = 917.29

Comment
RMSE = 11.39

like
RMSE = 272.13

Share
RMSE = 34.1

CRÉATION ET ÉVALUATION DES MODÈLES

- **Variable Selection** : Stepwise, Ridge et Lasso

Stepwise

Call:

```
glm(formula = Lifetime.Engaged.Users ~ Type + Page.total.likes +  
    Paid, data = trainData)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	627.719056	389.729302	1.611	0.1081	
Type	904.188329	110.947801	8.150	4.90e-15	***
Page.total.likes	-0.013267	0.003025	-4.386	1.48e-05	***
Paid	250.395429	107.248383	2.335	0.0201	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CRÉATION ET ÉVALUATION DES MODÈLES

- **Variable Selection** : Stepwise, Ridge et Lasso

Ridge

	s1
(Intercept)	-83.068766498
Page.total.likes	0.001420144
Type	44.993291663
Category	47.236056272
Post.Month	-4.668494260
Post.Weekday	-11.270882989
Post.Hour	0.004459634
Paid	74.484159805

- The quadratic error is defined by :

$$E(\beta) = (Y - X\beta)^T (Y - X\beta)$$

under the constraint $\|\beta\|^2 \leq c$

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

CRÉATION ET ÉVALUATION DES MODÈLES

- Variable Selection : Stepwise, Ridge et Lasso

Lasso

	s1
(Intercept)	-5.007868e+02
Page.total.likes	5.760598e-03
Type	5.328971e+01
Category	5.620518e+01
Post.Month	-2.565910e+01
Post.Weekday	-1.297556e+01
Post.Hour	.
Paid	8.910214e+01

- ℓ_1 Penalized OLS :

$$E(\beta) = (Y - X\beta)^T (Y - X\beta)$$

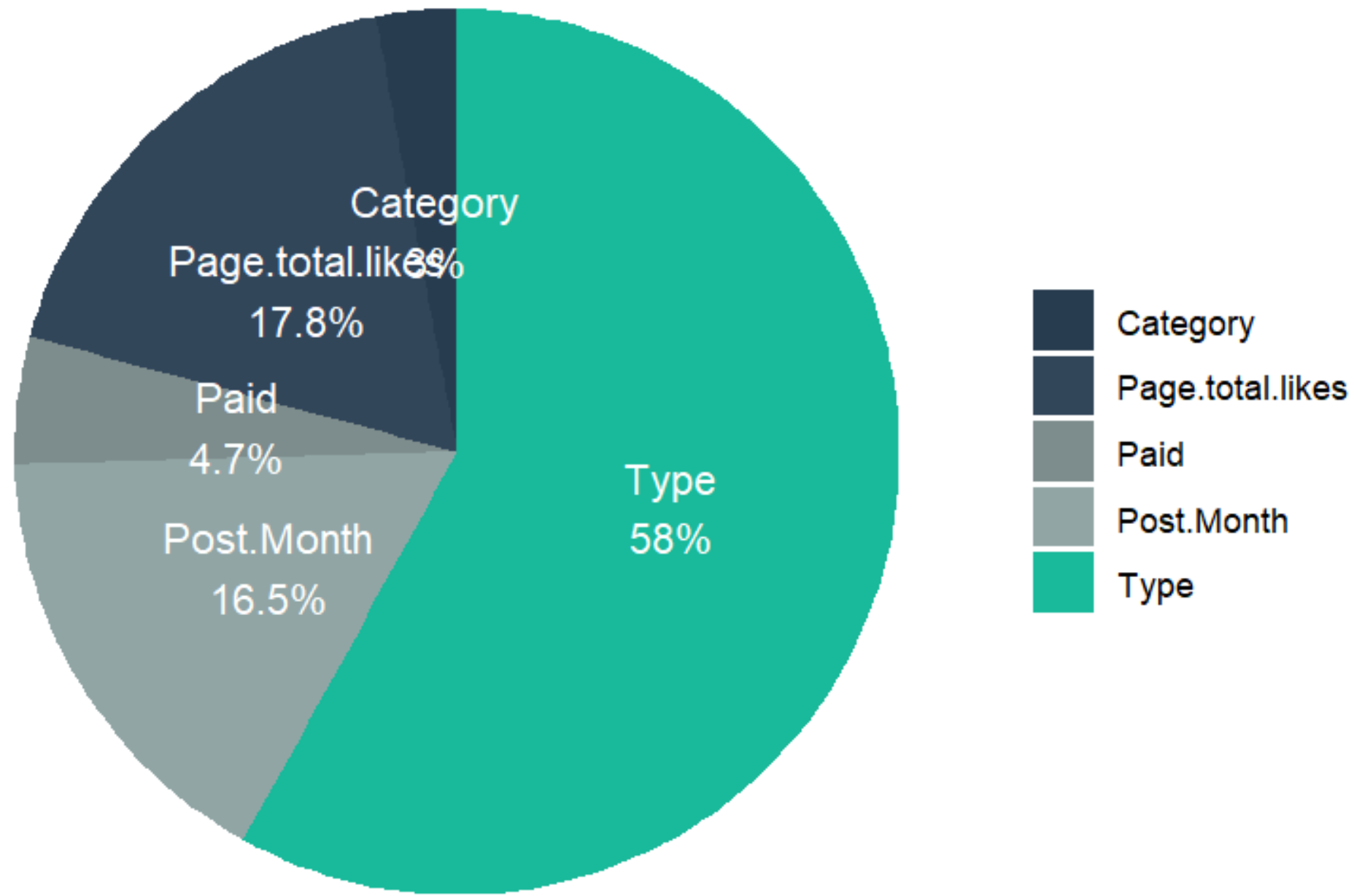
constrain $|\beta| \leq c$

$$\sum_{i=1}^n (Y_i - \sum_{j=1}^p X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

CRÉATION ET ÉVALUATION DES MODÈLES

- **Random Forest**

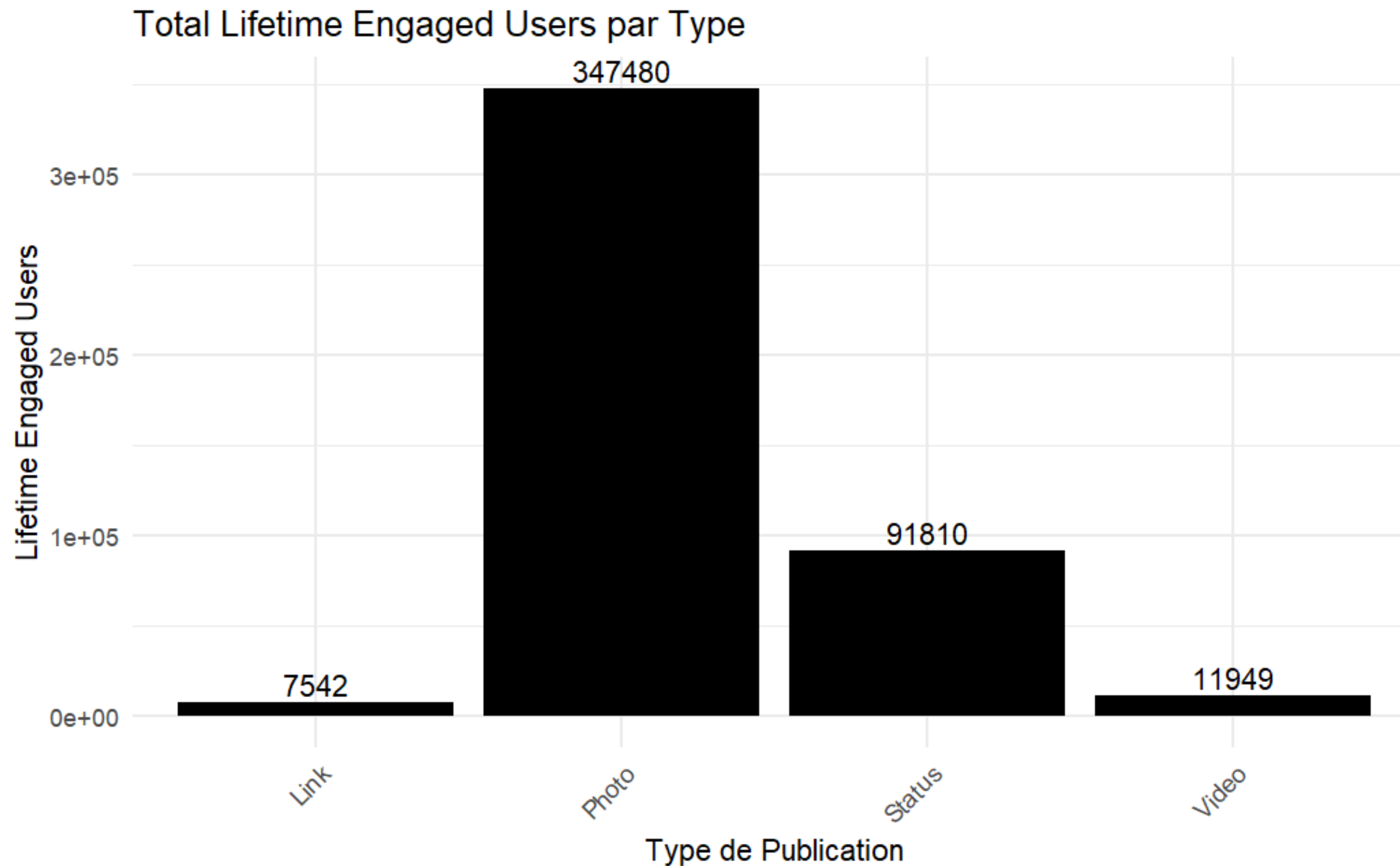
Relevance of Input Features for 'Lifetime Engaged Users'



Relative importance of each variable in the Random Forest model

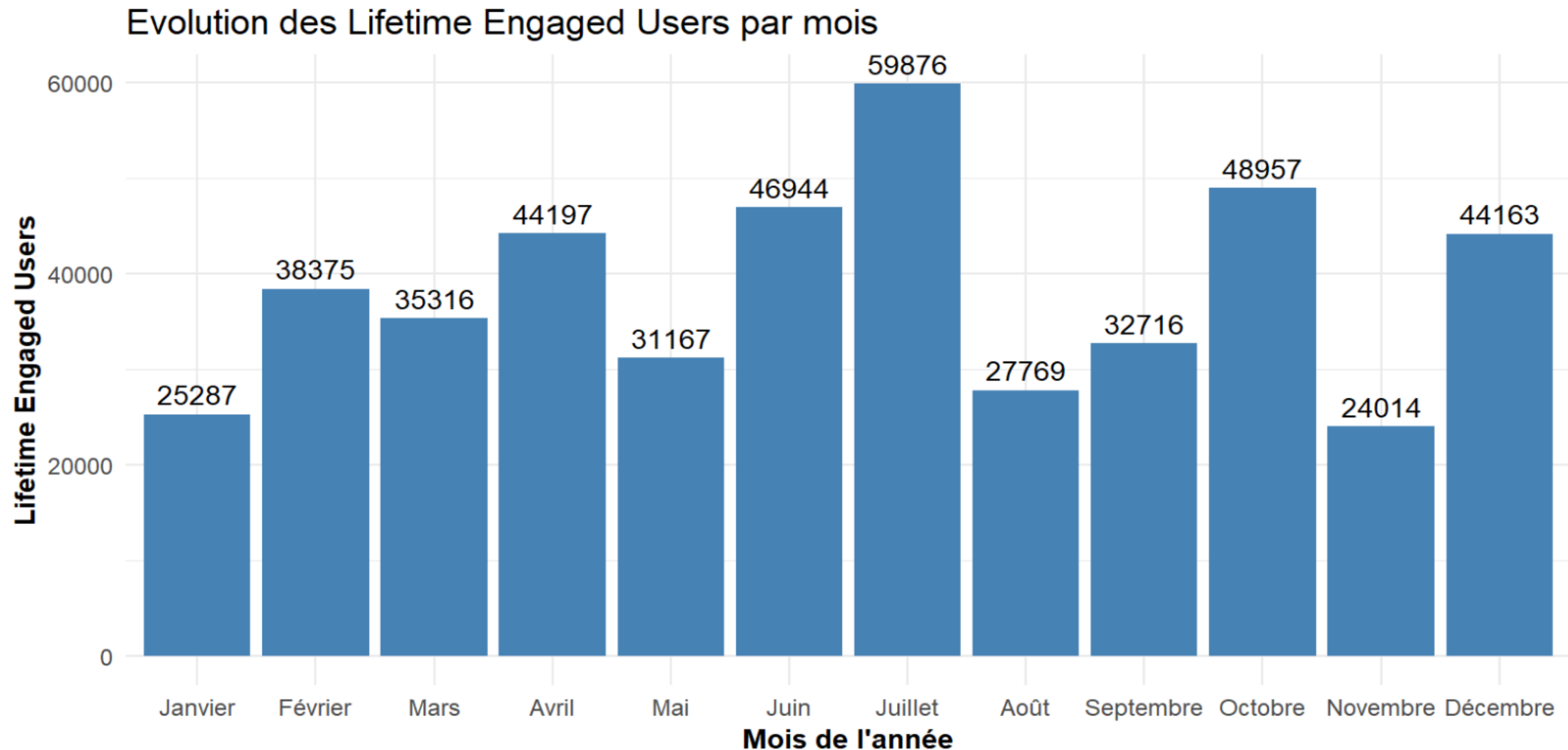
CRÉATION ET ÉVALUATION DES MODÈLES

- Random Forest



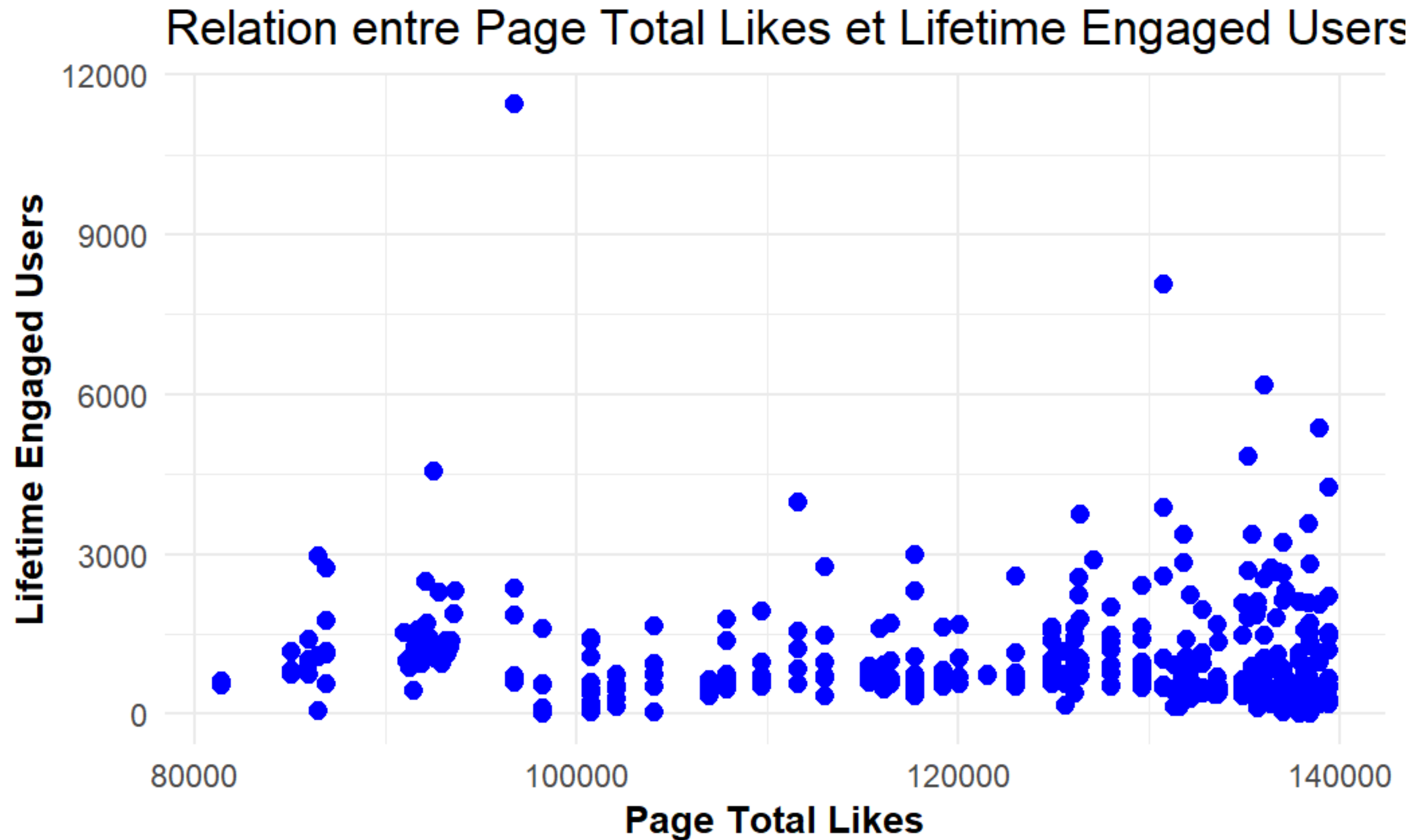
CRÉATION ET ÉVALUATION DES MODÈLES

- **Random Forest**

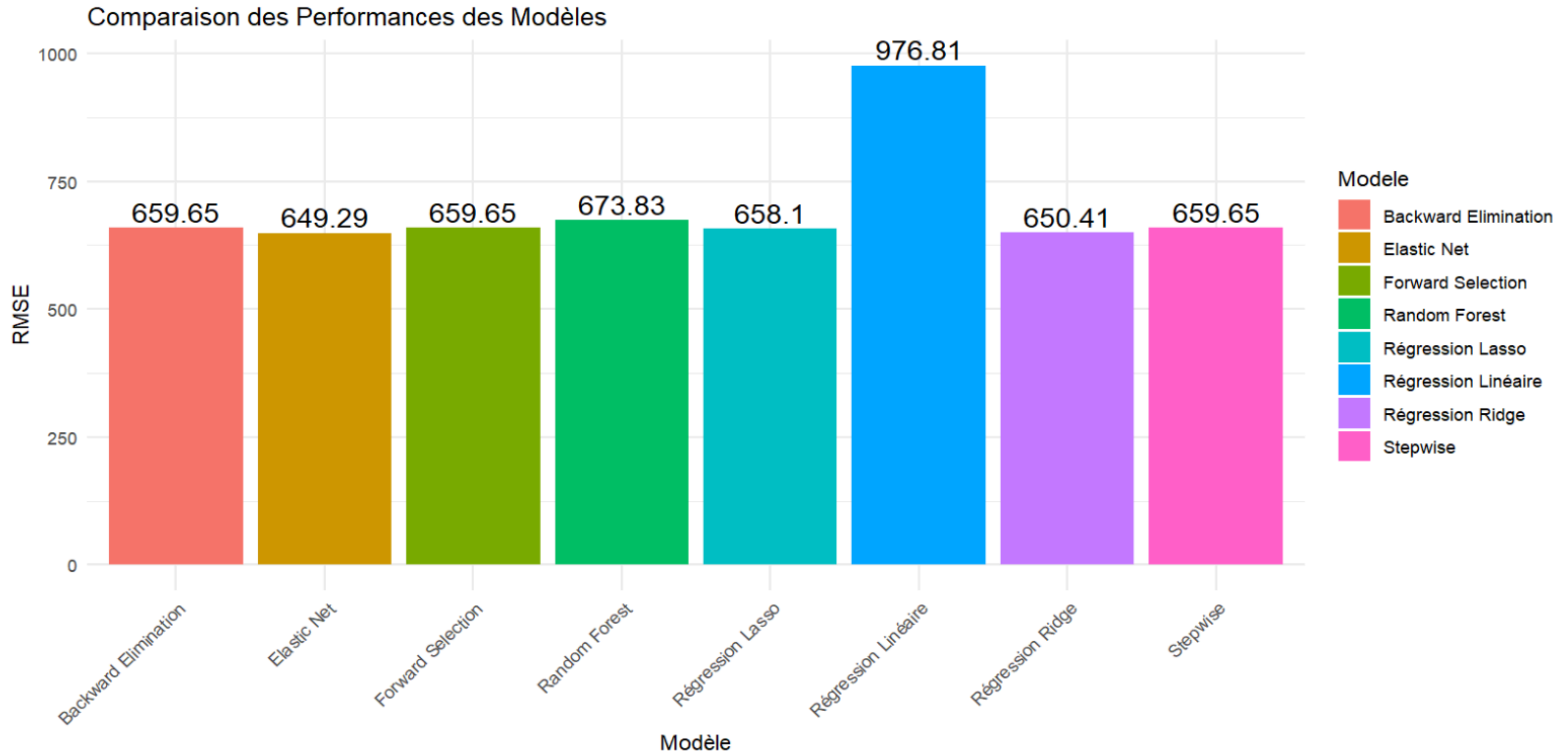


CRÉATION ET ÉVALUATION DES MODÈLES

- Random Forest



COMPARAISON DES MODELES



CONCLUSION



Elastic Net a la plus petite valeur de RMSE, ce qui en fait le **modèle le plus performant** parmi ceux testés



Le taux d'erreur de la méthode est d'environ **70%**



Expérimenter d'autres modèles pour déterminer si de meilleures performances peuvent être obtenues.