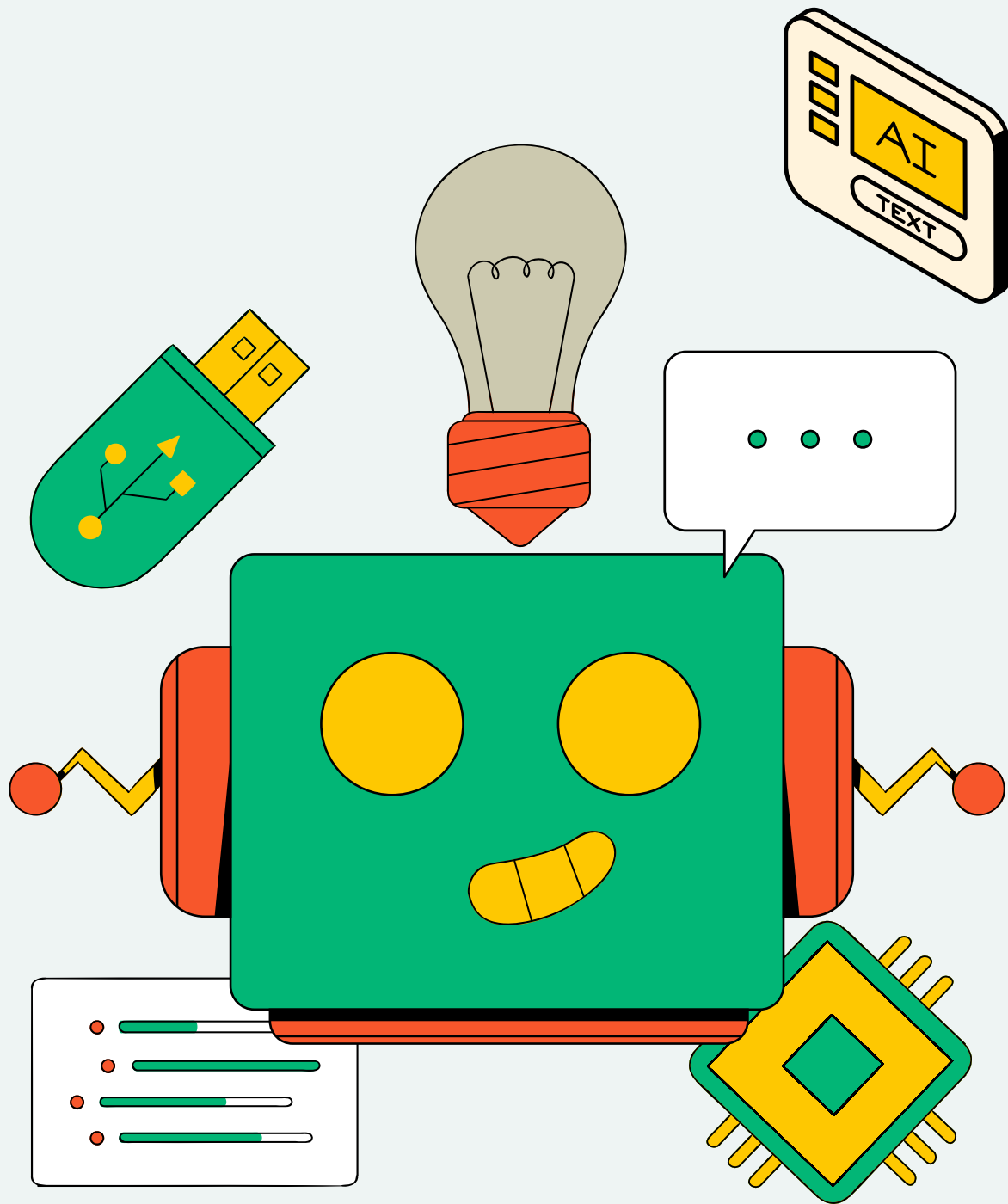




ÉQUIPE IPSSI
WE LEARN FOR THE FUTURE



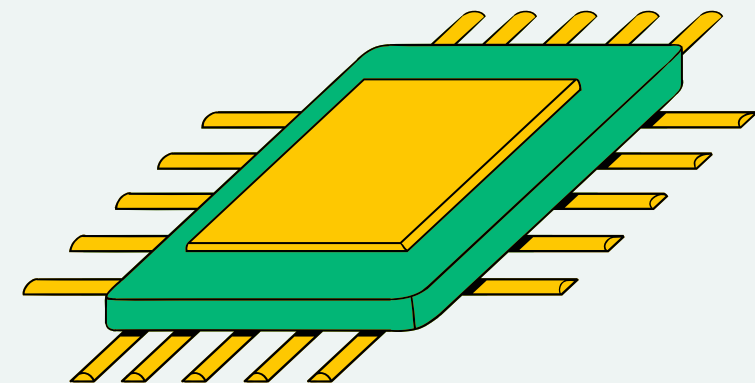
AIRBNB AI ML PREDICTON

PRESENTATION

ENZO DAVAL

YANIS HABAREK

KENUHN RIMBERT





PRESENTATION

- Introduction
- Présentation des outils
- Présentation du dataset
- nettoyage des données
- différents types d'algorithmes utilisés
- présentation du model A
- Présentation du model B
- Présentation du shiny



INTRODUCTION

Le projet consistait à prédire les prix des chambres sur la plateforme airbnb. À l'aide d'un dataset de scrapping airbnb

Nous avons utilisé des technologies tels que R, Shiny, Git. Enfin nous, avons utilisés une regression linéaire.



PRÉSENTATION DU DATASET

Objectif : Analyser les locations de courte durée (Airbnb) à Paris et en Île-de-France pour comprendre les tendances du marché immobilier, l'impact sur les prix, et les caractéristiques des logements.

Utilisation du dataset :

- Prédire les prix des logements.
- Analyser l'impact de la localisation et des équipements sur les tarifs.
- Étudier les tendances du marché de la location touristique à Paris et en Île-de-France.



VARIABLE CLÉS

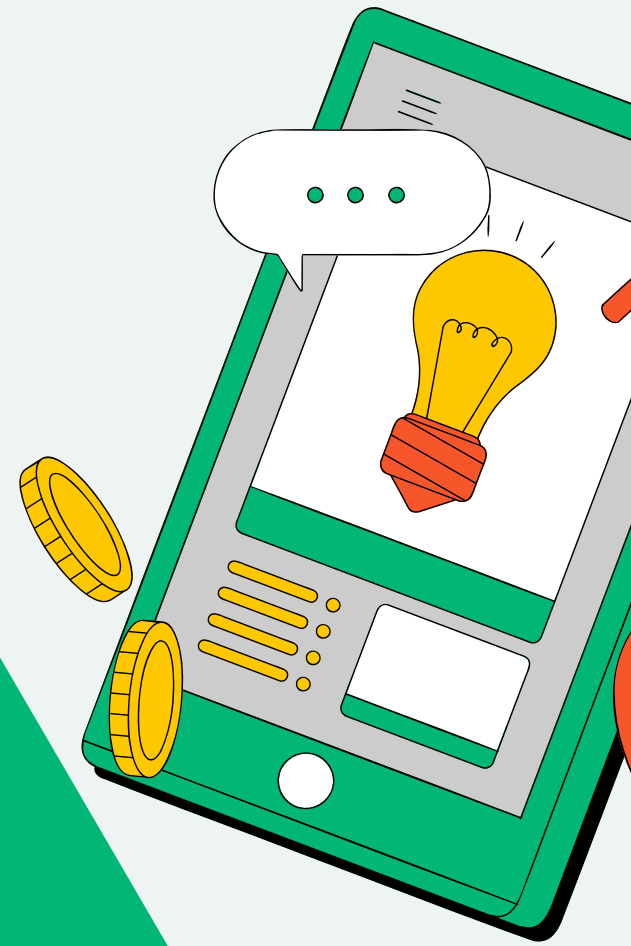
Prix

Notes

Nombre de lits

**Nombre de
salle de bain**

**Nombre de
chambre**

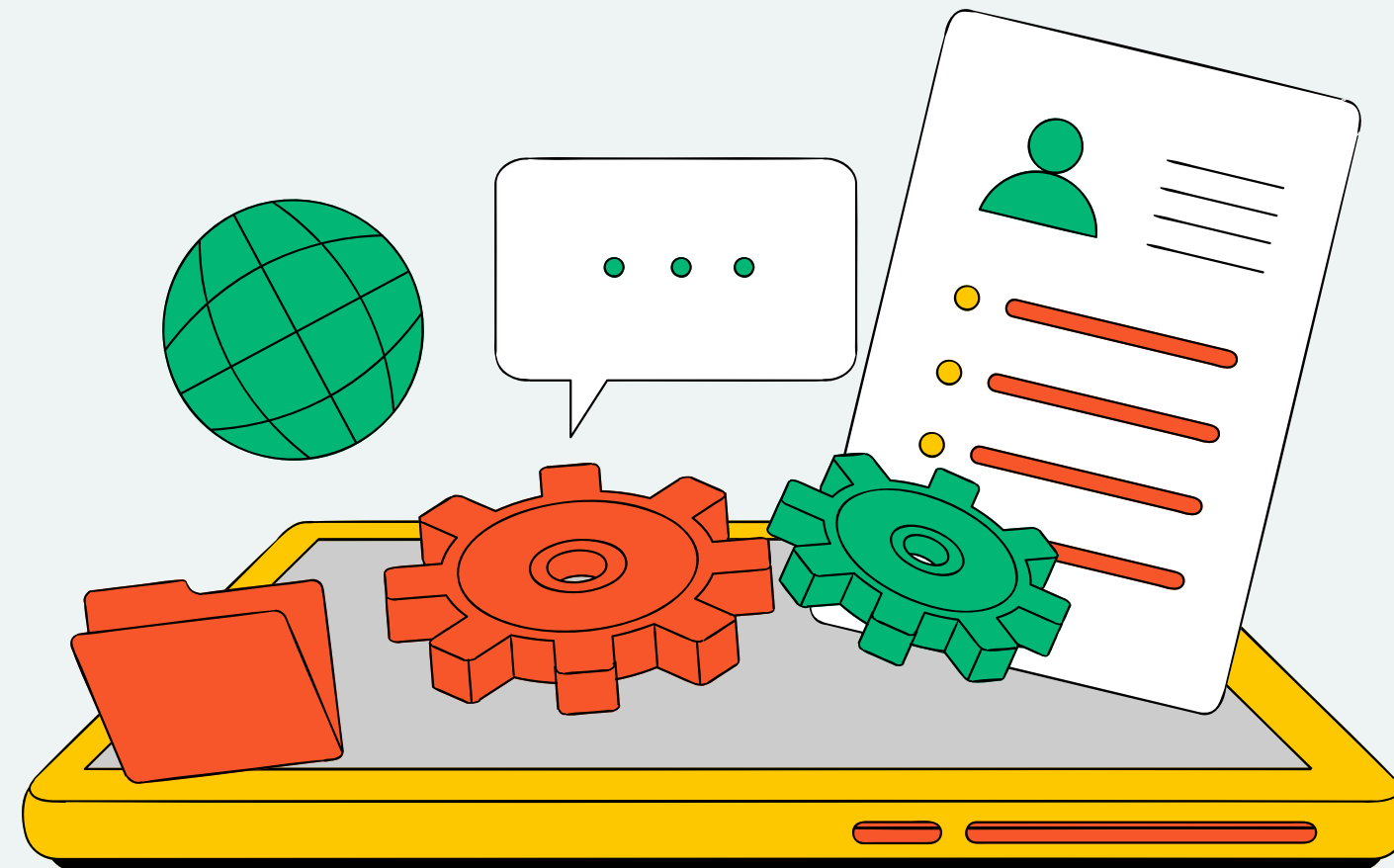


DIFFÉRENTS ALGORITHMES UTILISÉS

Gradient Boost

Random Forest

Linear Regression



MANIPULATION DES DONNÉES

01

SUPPRESSION DES DOLLARS

```
# enlever le "$" de chaque prix et transformer en numérique  
data$price <- as.numeric(gsub("\\$", "", data$price))
```

02

REMPLACEMENT DES NAN

```
checkNoNa <- function(data){  
  for (i in 1:ncol(data)){  
    if (sum(is.na(data[,i])) > 0){  
      print(paste("Column", colnames(data)[i], "a", sum(is.na(data[,i])), "valeur NA"))  
    }  
  }  
}
```

03

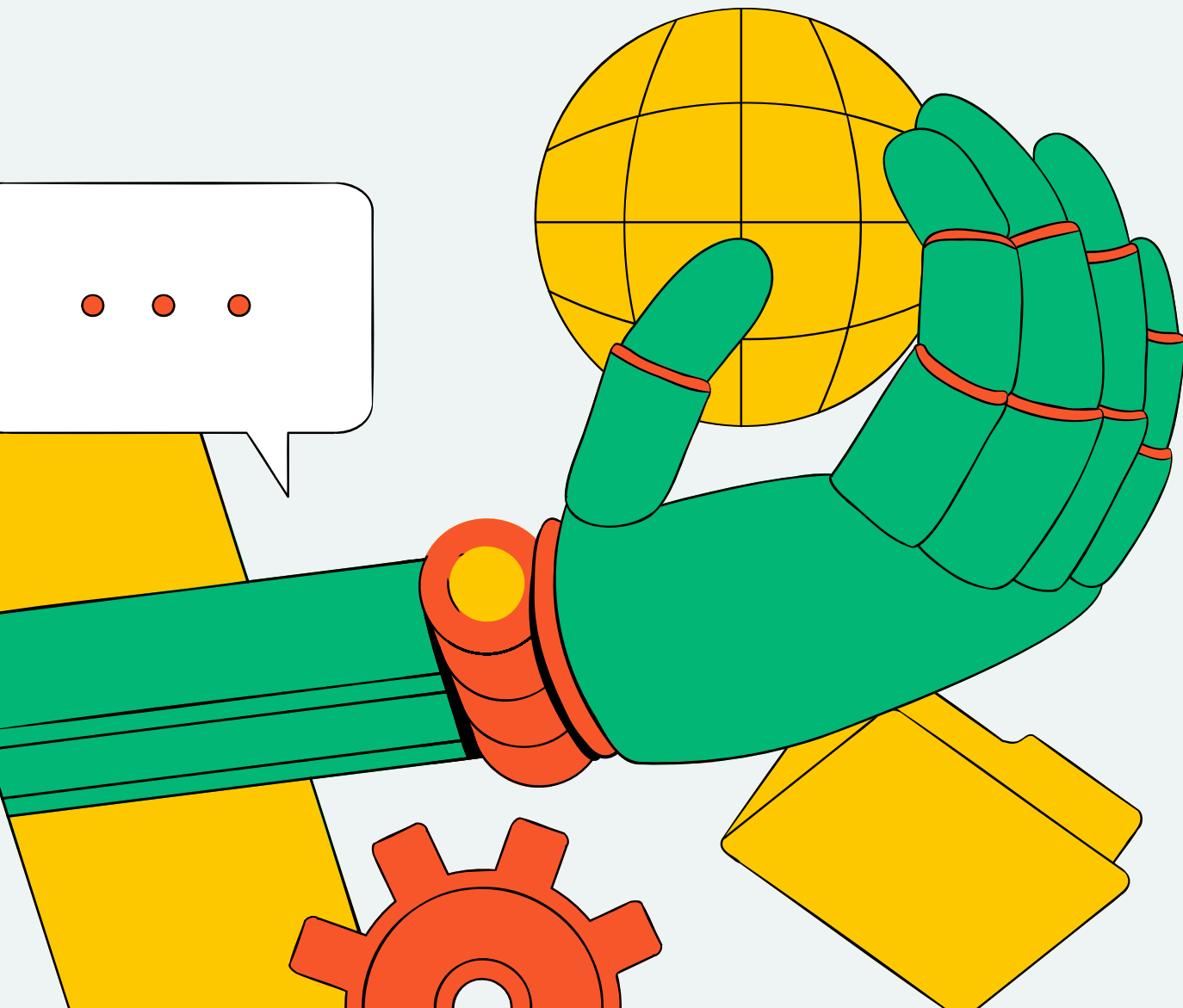
VÉRIFICATION INFLUENCE SUR TARGET

```
cor(data_train[, c("bedrooms", "beds", "bathrooms", "price", "review_scores_rating")])
```



RANDOMFOREST

Optimisé pour atteindre une erreur moyenne faible ($RMSE < 20$), il fournit des estimations fiables et adaptées aux spécificités locales, offrant ainsi une référence utile pour propriétaires et voyageurs.



```
# Division des données en ensembles d'entraînement et de test
set.seed(seed: 123) # Pour la reproductibilité
train_index <- createDataPartition(data_clean$price, p = 0.8, list = FALSE)

train_data <- data_clean[train_index, ]
test_data <- data_clean[-train_index, ]

# Créer un modèle de régression avec Random Forest
rf_model <- randomForest(price ~ ., data = train_data, importance = TRUE, ntree = 1000)

# Afficher un résumé du modèle
print(rf_model)

# Vérifier l'importance des variables
importance(rf_model)

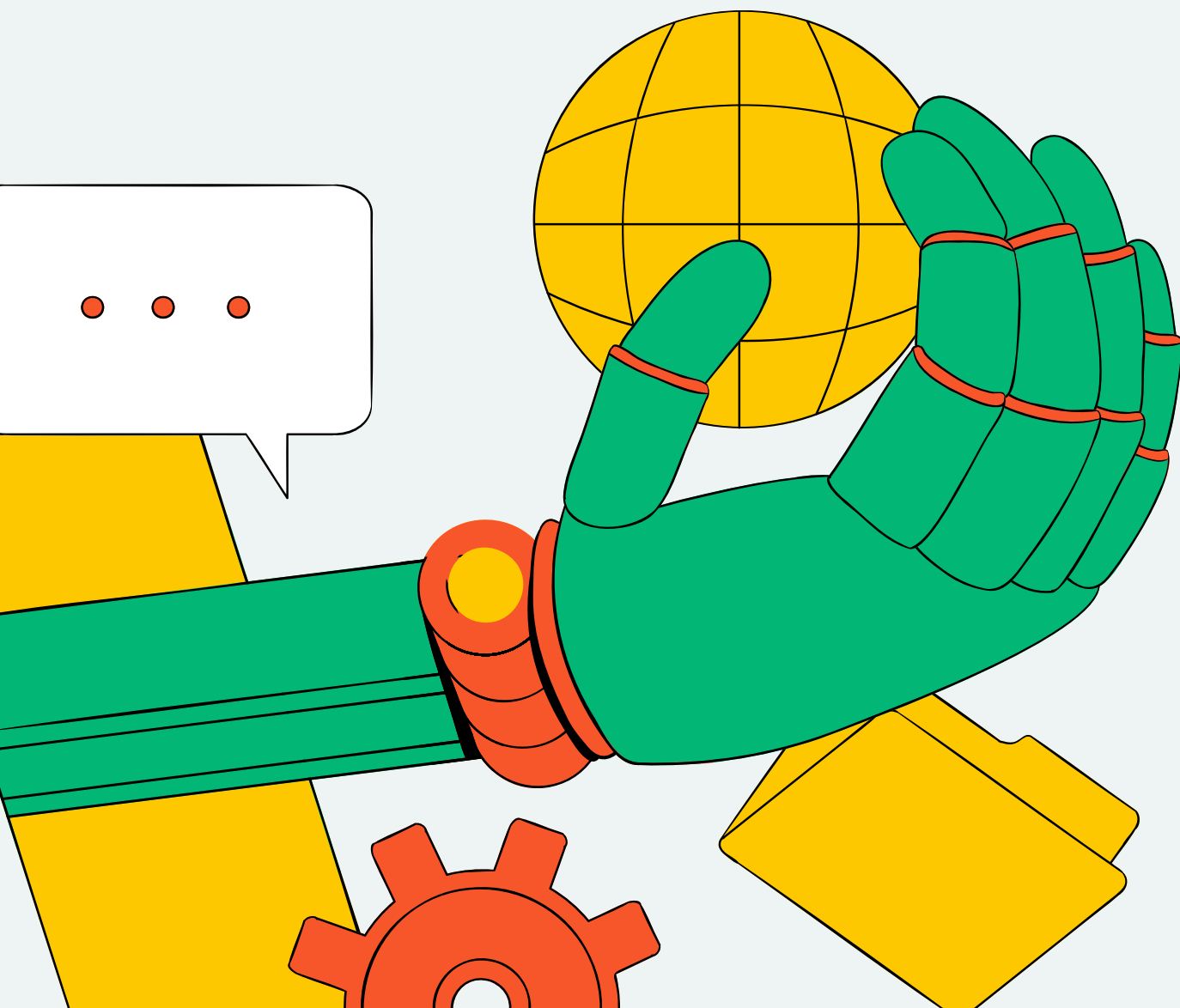
# Prédiction sur les données de test
predictions <- predict(rf_model, test_data)
#
# rf_model_tuned <- randomForest(price ~ ., data = train_data, importance = TRUE, ntree = 1000, mtry = 2)
#
# # Prédiction avec le modèle ajusté
# predictions_tuned <- predict(rf_model_tuned, test_data)
#
# # Calcul de la RMSE après ajustement
rmse_tuned <- sqrt(mean((predictions - test_data$price)^2))
print(paste("RMSE du modèle ajusté : ", rmse_tuned))

summary(rf_model)
```



MODÈLE DE REGRESSION

Le modèle utilise ces facteurs pour identifier les relations avec le prix. Une phase de validation sur un échantillon testé permet d'évaluer sa précision, offrant ainsi des prévisions fiables et adaptées aux caractéristiques du marché local.



```
# data split
split = 0.7
trainIndex<-createDataPartition(data$price, p=split, list = FALSE)

data_train<-data[trainIndex, ]
data_test<-data[-trainIndex,]

cor(data_train[, c("bedrooms", "beds", "bathrooms", "price", "review_scores_rating")])

# On peut voir que les variables "room_type", "neighbourhood" et "number_of_reviews" ne sont
# On va donc les retirer du modèle
modele <- lm(price ~ bedrooms + beds + bathrooms + review_scores_rating, data = data_train)

x_test<-data_test[, c(-2)]
y_test<-data_test[,2]

# prediction
predictions<-predict(modele, data_test)

summary(modele)
```

PRÉSENTATION SHINY



L'utilisateur peut explorer l'impact de différentes caractéristiques, telles que le nombre de chambres, de lits et de salles de bains, sur le prix des logements.

Cette interface permet de comparer les prix réels et prévus, offrant ainsi une compréhension intuitive du modèle et de ses résultats en fonction des spécifications des logements.

A screenshot of a Shiny web application interface. The browser address bar shows 'http://127.0.0.1:6431'. The page title is 'Prédiction de prix avec un modèle de Régression linéaire'. On the left, there is a form with five input fields: 'quartier:' with value '9', 'évaluation cliente :' with value '4,21', 'chambres:' with value '1', 'lits:' with value '4', and 'salles de bains:' with value '1'. Below these fields is a button labeled 'Prédire le prix'. On the right, under the heading 'Résultat de la prédiction :', there is a box displaying 'Prix prédit: 507 USD'. The background of the application is a solid red color.

Prédiction de prix avec un modèle de Régression linéaire

quartier:
9

évaluation cliente :
4,21

chambres:
1

lits:
4

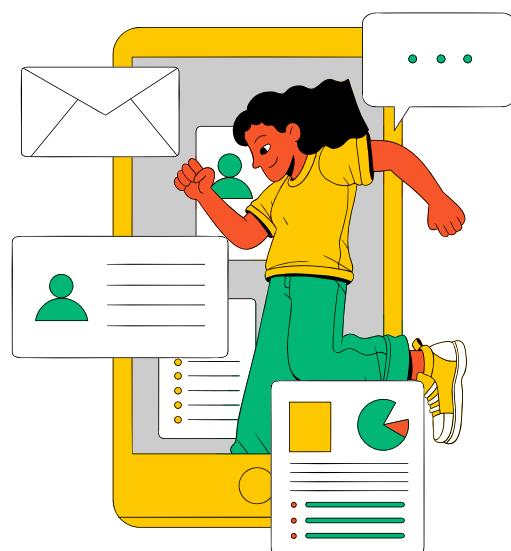
salles de bains:
1

Prédire le prix

Résultat de la prédiction :
Prix prédit: 507 USD

LOGEMENT DE TEST AIRBNB

PREMIER LIEN



Cliquez-moi

DEUXIÈME LIEN



Cliquez-moi

TROISIÈME LIEN



Cliquez-moi



POUR CONCLURE

Notre modèle prédictif offre une estimation précise des prix des logements Airbnb en Île-de-France, en prenant en compte des facteurs clés comme le type de logement et le nombre de chambres.

Grâce à une approche de régression robuste et à l'utilisation d'une transformation logarithmique du prix, nous avons optimisé les prédictions. L'application Shiny permet une exploration interactive des résultats, offrant ainsi un outil utile pour les propriétaires et les voyageurs.

