

Prédiction des clients à risque de cesser d'utiliser les services de SyriaTel

Projet de classification binaire

Auteur: Nael Yssa I.A Robert

Objectif du projet

- Prédire quels clients sont à risque de cesser d'utiliser les services de SyriaTel, afin que l'entreprise puisse mettre en place des actions pour réduire les pertes financières.
- Type de problème : Classification binaire (Churn = True/False)
- Public cible : secteur des télécommunications.
- Source de données : Kaggle

Étape 1 : Exploration des données

Jeu de données : 3333 clients, 21 colonnes (informations sur usage des services, plan international, appels au service client, etc.)

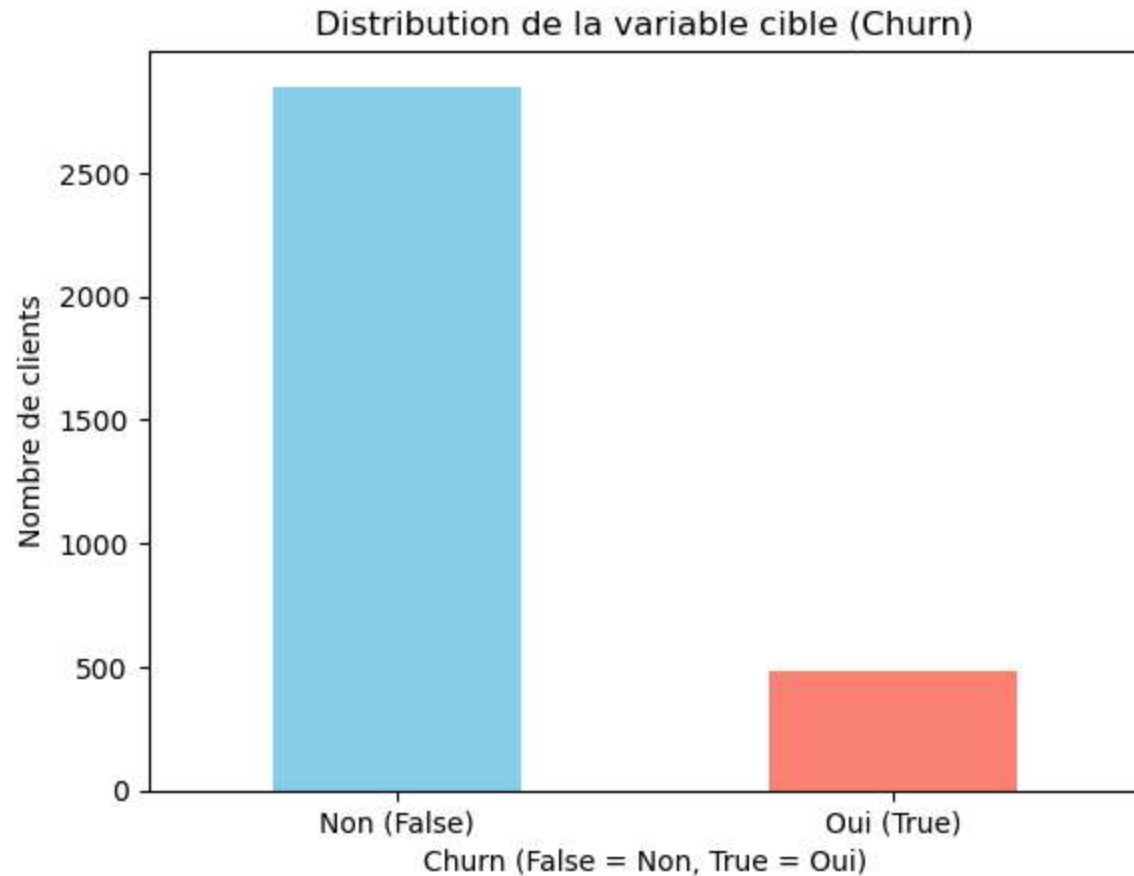
Méthodes utilisées pour l'exploration :

`df.head()`, `df.info()`, `df.isnull().sum()`, `df['churn'].value_counts()`

Analyse initiale :

- Pas de valeurs manquantes
- Classes déséquilibrées : 2850 Non-Churn, 483 Churn (~14% churn)
- Vérification des types de données : numériques et catégorielles

Visualisation de la distribution de la variable cible(churn)



Étape 2 : Préparation des données

Objectif : Mettre les données sous une forme exploitable par les modèles de classification.

Méthodes utilisées :

- Encodage des variables catégorielles : `pd.get_dummies()`
- Standardisation des variables numériques : `StandardScaler().fit_transform()`
- Séparation en train/test sets avec stratification : `train_test_split(..., stratify=0.3, random_state=1)`

Actions :

- Encodage des variables catégorielles (One-Hot)
- Standardisation ou normalisation des variables numériques
- Séparation en train/test sets avec stratification pour maintenir la proportion de classes

Étape 3 : Construction des modèles

Objectif : Créer plusieurs modèles de classification pour prédire le churn.

Méthodes utilisées :

- Régression logistique : `LogisticRegression(class_weight='balanced')`
- Arbre de décision : `DecisionTreeClassifier(class_weight='balanced')`
- Évaluation des performances : `accuracy_score()`, `classification_report()`, `confusion_matrix()`

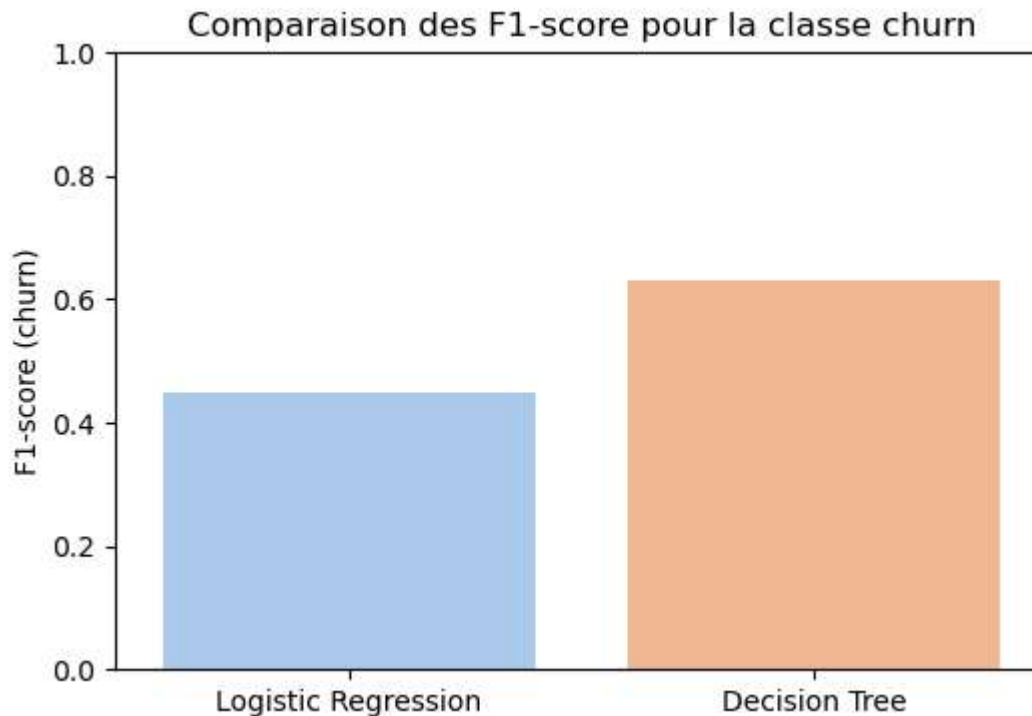
Actions :

- Régression logistique (simple et interprétable)
- Arbre de décision (capte les relations non linéaires)
- Paramètre `class_weight='balanced'` pour gérer le déséquilibre

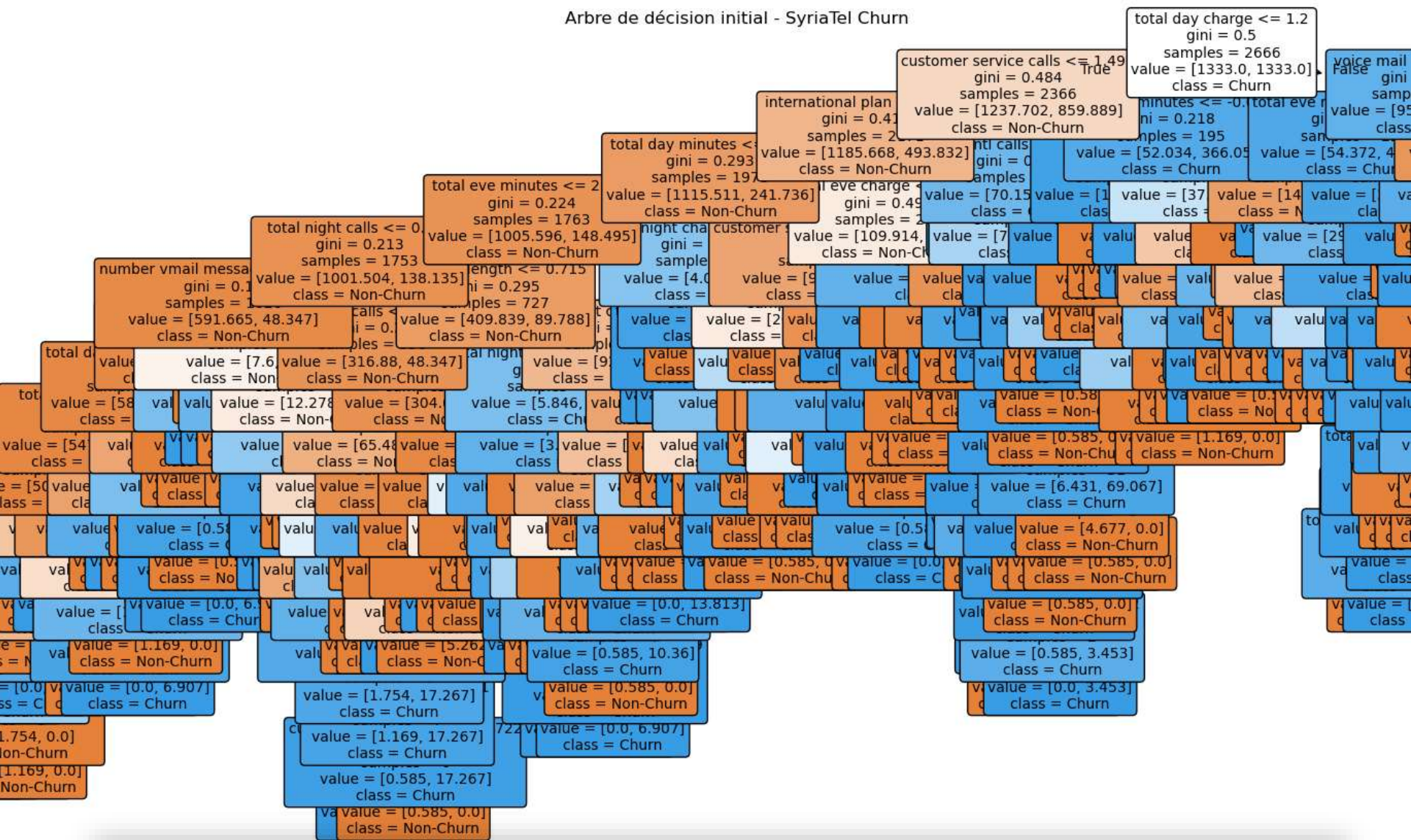
Résultats:

- **Régression logistique**
 - Accuracy : 0.747
 - Classe Churn : Precision 0.33, Recall 0.70, F1-score 0.45
 - Limitation : faible précision pour détecter les clients à risque (beaucoup de faux positifs).
- **Arbre de décision initial**
 - Accuracy : 0.892
 - Classe Churn : Precision 0.63, Recall 0.63, F1-score 0.63
 - Avantage : meilleure détection des clients à risque, capture les relations non linéaires.

Visualisation de la comparaison des f1_score de chacun des modèles



Arbre de décision initial - SyriaTel Churn



Visualisation de l'arbre de décision

Étape 4 : Optimisation de l'arbre de décision

- **Meilleurs paramètres** trouvés :

```
{'criterion': 'entropy', 'max_depth': 5,  
'min_samples_leaf': 1,  
'min_samples_split': 10}
```

- **Résultats sur le test set :**

- Accuracy : 0.891

- Classe Churn (clients à risque) : Precision 0.61, Recall 0.68, F1-score 0.64

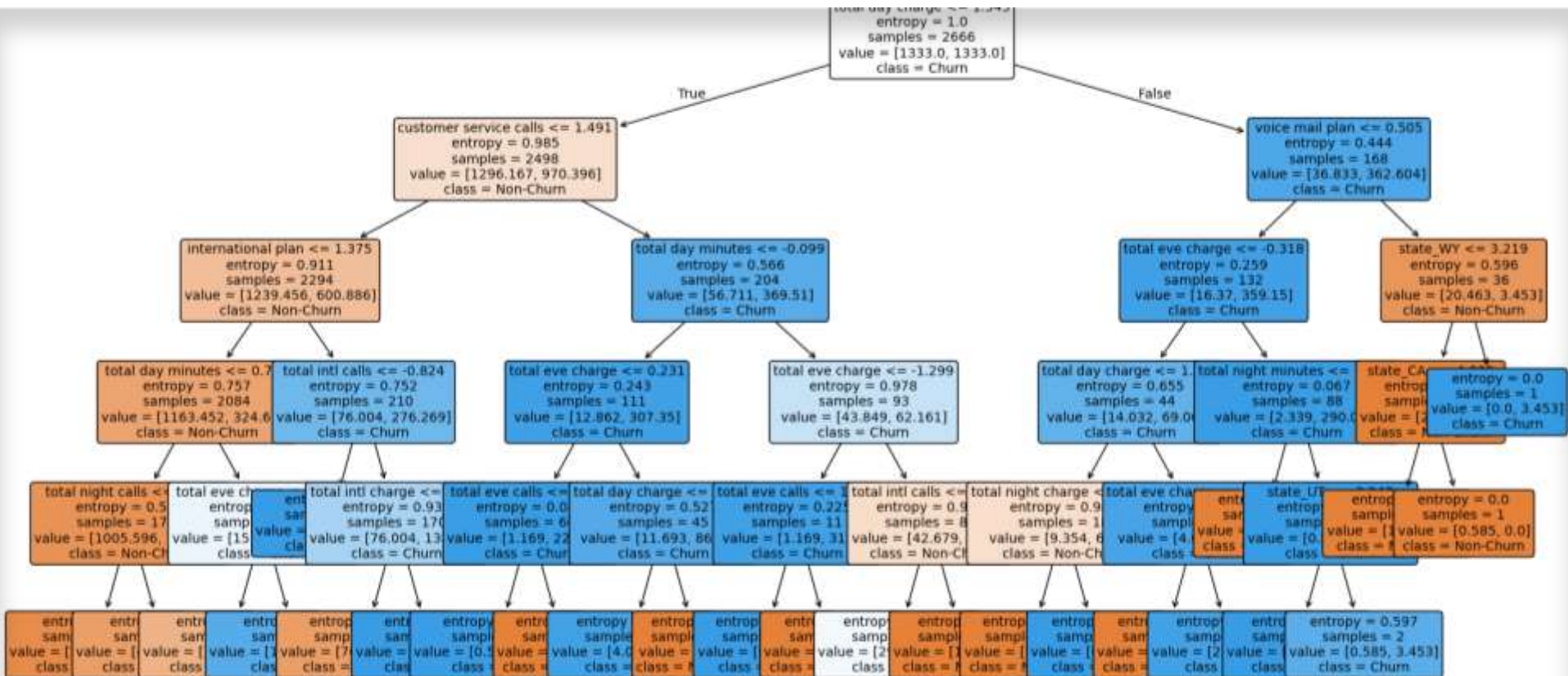
- Confusion Matrix :

```
[[528 42] [ 31 66]]
```

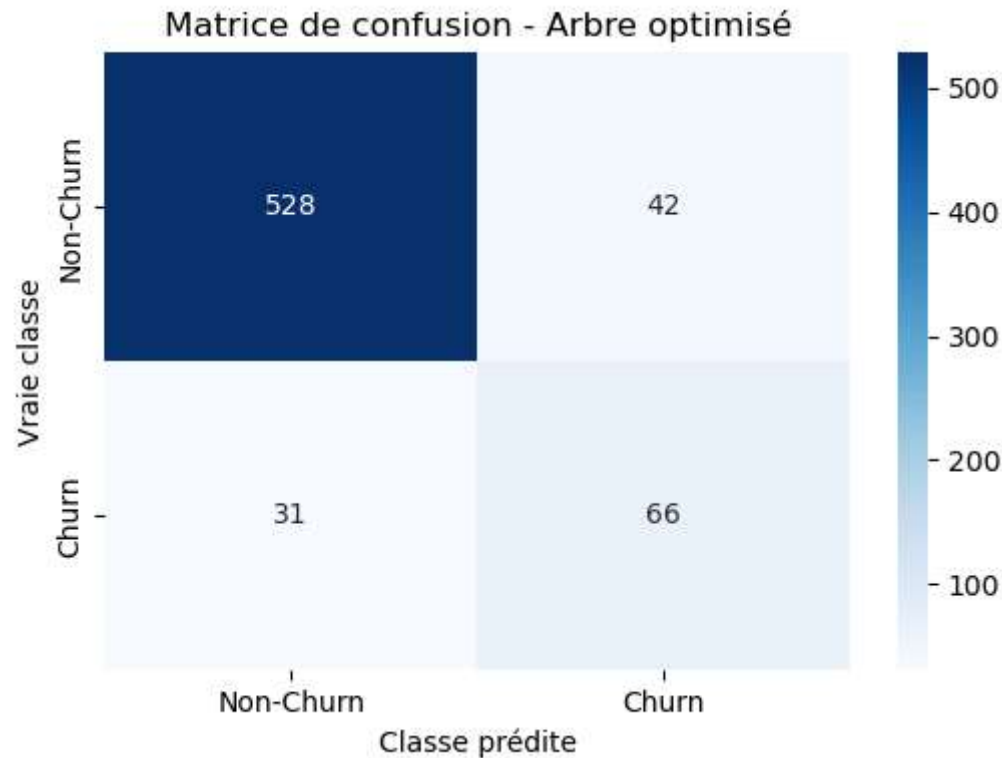
- **Nombre total de clients identifiés comme à risque : 108**
(66 vrais churn + 42 faux positifs)

- Optimisation : meilleur compromis entre précision et rappel pour détecter les clients à risque.

Visualisation de l'arbre optimisé



Visualisation de la matrice de confusion de l'arbre optimisé



Recommandations

L'arbre de décision optimisé est **le modèle le plus efficace** pour prédire les clients à risque de churn.

Actions pour SyriaTel :

Contacter les **108 clients identifiés comme à risque** avec des offres ou un suivi personnalisé.

Réévaluer régulièrement le modèle pour maintenir la précision.

Améliorations possibles :

Ajouter plus de données comportementales ou démographiques pour mieux détecter les clients à risque.

Conclusion

Le modèle optimisé permet à SyriaTel d'identifier les clients à risque de cesser d'utiliser ses services (churn) et de mettre en place des actions ciblées pour réduire les pertes financières.

MERCI !!!