

## Rapport Final

### Introduction

Le rapport présente une analyse détaillée d'un modèle de détection de fraude basé sur les données de carte de crédit. L'objectif principal était de développer un modèle prédictif capable d'identifier les transactions frauduleuses dans un ensemble de données très déséquilibré.

### Processus Suivi

1. **Exploration des Données** : Les données ont été chargées et une analyse exploratoire a été réalisée pour comprendre la distribution des classes, la répartition des caractéristiques, les valeurs manquantes, et la corrélation entre les variables.
2. **Prétraitement des Données** : Les valeurs manquantes ont été vérifiées et aucune n'a été trouvée. Les doublons ont été supprimés pour assurer la qualité des données. De plus, une mise à l'échelle de la colonne 'Amount' a été effectuée.
3. **Gestion des Déséquilibres** : Étant donné le déséquilibre entre les classes, la méthode SMOTE a été utilisée pour suréchantillonner la classe minoritaire.
4. **Entraînement du Modèle** : Un modèle de régression logistique a été instancié, entraîné sur les données suréchantillonnées et évalué avec des mesures telles que la précision, le rappel et le F1-score.
5. **Validation Croisée** : Une validation croisée à 5 plis a été effectuée pour évaluer la généralisation du modèle.
6. **Analyse des Caractéristiques Importantes** : Les caractéristiques les plus importantes identifiées par le modèle ont été analysées pour comprendre leur impact sur les prédictions.

### Résultats Obtenu

- **Performances du Modèle** :
  - Accuracy : 0.9708
  - Precision : 0.9807
  - Recall : 0.9607
  - F1-score : 0.9706
- **Scores de Validation Croisée** : [0.9413, 0.9700, 0.9714, 0.9714, 0.9683]
- **10 Caractéristiques les plus Importantes** :
  - V14, V17, V3, V4, V12, V10, V16, V9, V8, V13

### Conclusions

Le modèle de régression logistique a montré des performances encourageantes dans la détection de fraudes sur des données de carte de crédit. Les scores élevés en précision, rappel et F1-score indiquent

une capacité à identifier avec précision les transactions frauduleuses, minimisant ainsi les faux positifs et faux négatifs.

#### Recommandations

- Explorer des modèles plus complexes pour améliorer encore les performances.
- Collecter des données supplémentaires pour enrichir l'ensemble de données existant et renforcer la capacité du modèle à généraliser.
- Continuer à surveiller et à réévaluer régulièrement le modèle pour l'adapter aux nouvelles tendances de fraude.

#### Perspectives Futures

Le développement de ce modèle peut être étendu à d'autres techniques d'apprentissage automatique, telles que les arbres de décision, les méthodes ensemblistes ou les réseaux neuronaux, pour rechercher des améliorations continues des performances de détection de fraude.

#### Conclusion

Ce rapport offre un aperçu détaillé du processus suivi pour développer un modèle de détection de fraude. Bien que le modèle actuel montre de bonnes performances, des améliorations continues et une surveillance constante sont essentielles pour rester efficace dans la détection des fraudes dans les transactions par carte de crédit.