

## Rapport d'Analyse de Données

**Objectif** Ce projet vise à prédire les joueurs de StarCraft 2 en utilisant un modèle de machine Learning à partir de données de jeu. Plus spécifiquement, notre objectif est de déterminer quel joueur a réalisé certaines actions au cours d'une partie.

**Base de Données** : Nous avons utilisé un dataset contenant des informations sur les parties de StarCraft 2. Ce dataset inclut diverses caractéristiques telles que le nombre d'actions par minute, les ressources collectées, le temps de jeu, et d'autres métriques de performance des joueurs.

**Traitement des Données** : Il était nécessaire de procéder à un prétraitement des données car elles présentaient des valeurs manquantes. Pour résoudre ce problème, nous avons employé la technique de suppression des valeurs manquantes. De plus, les données ont été standardisées afin d'assurer que toutes les caractéristiques soient homogènes, ce qui améliore les performances des modèles d'apprentissage automatique.

**Choix et Entraînement des Modèles** : Nous avons opté pour l'utilisation d'un RandomForestClassifier en raison de sa robustesse et de son efficacité pour les données complexes. Effectivement, cela facilite la gestion des interactions non linéaires entre les caractéristiques. Voici les étapes que nous avons suivi :

- **Chargement des données** : Importer les données depuis un fichier CSV.
- **Prétraitement** : Suppression des valeurs manquantes et normalisation des données.
- **Séparation des données** : Division des données en ensembles d'entraînement et de test.
- **Entraînement** : Entraîner le modèle RandomForest sur les données d'entraînement.

**Évaluation de l'Entraînement** : L'entraînement s'est déroulé sans problème, avec une précision de 85% sur les données de test. Effectivement, le modèle a correctement identifié les motifs des données d'entraînement.

**Comparaison des Modèles** : Nous avons comparé le RandomForest avec un LogisticRegression. Le RandomForest a montré de meilleurs résultats en termes de précision et de recall, justifiant notre choix initial.

**Optimisation** : Afin d'optimiser les résultats, nous avons employé une méthode d'optimisation des hyperparamètres connue sous le nom de "recherche en grille". Cette méthode implique de tester régulièrement différentes combinaisons d'hyperparamètres afin de déterminer la configuration optimale pour le modèle. Nous avons pu améliorer les performances du modèle grâce à cette optimisation, même si les améliorations sont limitées.

**Conclusion et Perspectives** : En résumé, le modèle de RandomForest a démontré son efficacité dans la prédiction des joueurs de StarCraft 2. Toutefois, il serait pertinent d'expérimenter d'autres modèles plus avancés et de collecter davantage de données afin d'améliorer encore davantage les prédictions.