

README IA & DATA : SUUUUPER TOURNOI

Équipe : Amadou Mbodj , Mehita Fanny , Aymeric Thevenet

Spécialité : Intelligence Artificielle et Data Science

Livrable Clé : Moteur de Matchmaking ELO

I. Démarche et Objectif IA/Data

L'objectif de la partie IA/Data était double :

1. **Fiabiliser** le jeu de données historique des matchs.
2. **Diagnostiquer** l'expérience utilisateur pour identifier le levier d'amélioration principal.
3. **Fournir un modèle** capable de créer des équipes équilibrées (Matchmaking) pour augmenter la satisfaction.

1. Actions Menées (Référence : [nettoyage_hackathon.ipynb](#))

La première phase a consisté à transformer le jeu de données brut et incohérent en une base exploitable (95 340 lignes nettoyées).

- **Nettoyage Intensif** : Standardisation des scores, des couleurs d'équipe, des rôles et des localisations.
- **Feature Engineering Temporel** : Conversion des durées de match de formats hétérogènes (HH:MM:SS, "10min", décimal) en secondes.
- **Standardisation des Ratings** : Conversion des notes de satisfaction (étoiles, émojis, texte) en une échelle numérique standardisée de 1 à 5.

II. Choix Techniques et Modélisation

1. Choix de l'Algorithme ELO

- **Choix Technique** : Mise en œuvre de l'algorithme de notation **ELO** (K=40) itéré sur les 23 835 matchs du dataset.
- **Justification** : L'ELO est l'algorithme standard pour évaluer la compétence relative des joueurs dans les jeux compétitifs (échecs, e-sport). Il fournit un score unique permettant au système de **quantifier le niveau réel** de chaque joueur, et ainsi de prédire le résultat probable d'un match.

2. Livrable et Intégration

- **Livrable IA** : Un classement dynamique (score ELO) pour les 800 joueurs uniques de la base.
- **Requis Fonctionnel pour le DEV** : L'application doit désormais constituer les équipes (RED et BLUE) en s'assurant que la **somme des scores ELO** des deux équipes est la plus proche possible.

III. Difficultés Rencontrées

La principale difficulté a été le **nettoyage des données brutes** (phase 1). Le dataset initial présentait une très forte hétérogénéité des formats de saisie :

- **Gestion du Temps** : L'interprétation des durées a nécessité la création de fonctions de parsing robustes.
- **Conversion Texte/Nombre** : Les statistiques (buts, assists) étant parfois saisies en texte ("two goals", "one save"), l'utilisation de bibliothèques comme `word2number` a été indispensable pour la conversion.

Ces difficultés ont pris 60% du temps initial de l'équipe, mais ont garanti la fiabilité du modèle ELO final.

IV. Collaboration avec les Autres Spécialités

La collaboration a été essentielle et s'est concentrée sur la fourniture de la donnée critique :

- **Collaboration avec DEV/Infra** : L'équipe IA/Data a livré le **Classement ELO final** (`ELO_Classement_Final.csv`). Ce classement est la source de vérité que l'équipe DEV/Infra doit intégrer à l'API de Matchmaking. Notre travail a permis de passer d'une simple collecte de données à un **système de recommandation actif**.
- **Collaboration avec Design/Marketing** : L'EDA a permis d'identifier des opportunités de Gamification (ex: **Maria Bianchi** comme Top Scoreuse) et des insights pour améliorer l'expérience utilisateur (le besoin d'un meilleur arbitrage).

V. Conclusion en tant qu'Expert IA/Data

Le projet a démontré que l'expertise Data est le fondement de la solution *next-gen*. En diagnostiquant l'insatisfaction (faible plaisir de jeu) et en identifiant la cause (déséquilibre des équipes), nous avons pu implémenter une solution mathématique éprouvée (ELO).

Le résultat final est un système intelligent qui ne se contente pas d'enregistrer les matchs, mais les optimise activement pour maximiser le plaisir des 1000 étudiants d'Ynov.