

Rapport de Projet Big data

Introduction

Notre projet vise à développer une solution d'analyse et de prédiction des interactions en combat entre les Pokémon en utilisant des techniques de Big Data.

L'objectif principal est de comprendre comment les types et leurs avantages ou désavantages influencent les probabilités de victoire.

Les données proviennent de deux sources principales :

- Statistiques des Pokémon : Détails sur les attributs/statistiques de base de chaque Pokémon.
- Matchups de types : Relations entre les types de Pokémon (avantages/désavantages).

Ces informations ont été combinées et nettoyées pour créer un pipeline de données complet, prêt à l'analyse et à la modélisation.

Objectifs

- Identifier les types de Pokémon ayant un avantage stratégique en combat.
- Analyser les interactions entre les types pour proposer des recommandations stratégiques.

Sources des données

- [Pokémon Database](#) : Données sur les relations entre types (avantages et désavantages).
- [Pokémon Complete](#) : Informations détaillées sur les statistiques de base de chaque Pokémon.

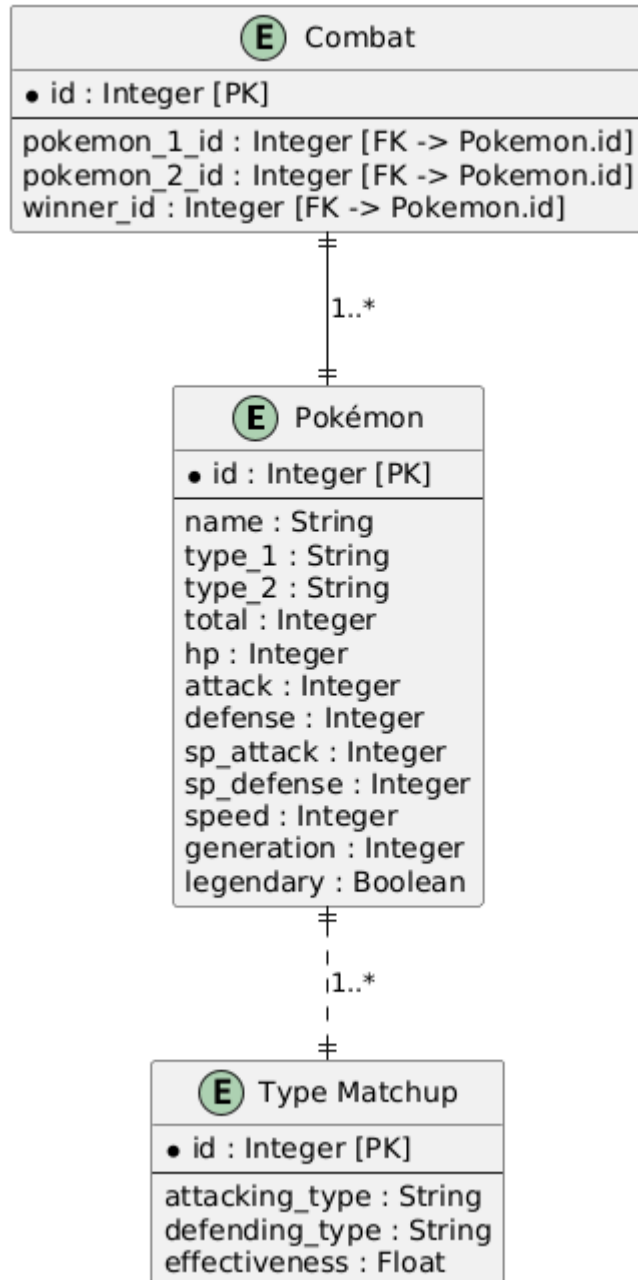
Dataset des Pokémon : Contient les informations sur les statistiques de base de chaque Pokémon.

Dataset des Matchups : Fournit les relations entre les types de Pokémon, indiquant les avantages et les désavantages.

Ces données ont été pré-traitées et intégrées dans notre pipeline.

Modèle Conceptuel de Données

Un modèle en étoile a été utilisé pour structurer les données dans la zone Gold :



Entité **Pokemon** :

- Contient les caractéristiques principales des Pokémon (statistiques, types, génération, etc.).
- Les champs **type_1** et **type_2** permettent de définir les types primaires et secondaires.

Entité **TypeMatchup** :

- Définit les relations entre les types avec le champ **effectiveness** (coefficient indiquant l'avantage ou le désavantage).

Entité **Combat** :

- Enregistre les combats entre deux Pokémon et le vainqueur.
- Les clés étrangères **pokemon_1_id**, **pokemon_2_id**, et **winner_id** pointent vers l'entité **Pokemon**.

Relations :

- Une relation entre **Pokemon** et **TypeMatchup** (un Pokémon peut être associé à plusieurs types).
- Une relation entre **Combat** et **Pokemon** (chaque combat concerne deux Pokémon, avec un vainqueur).

Architecture et Traitements

Les données ont été organisées dans une architecture de type Lakehouse :

Zone Bronze : Contient les données brutes sous forme de fichiers CSV.

Zone Silver : Données nettoyées, avec suppression des doublons et des entrées manquantes.

Zone Gold : Données transformées et préparées pour l'analyse, stockées dans un modèle optimisé.

Nettoyage et Transformation

Les étapes suivantes ont été suivies :

1. Suppression des doublons et des valeurs manquantes.
2. Normalisation des colonnes :
 - Uniformisation des noms.
 - Association des colonnes **Number** (Pokémon Complete) et **pokedex_number** (Pokémon Database).
3. Standardisation des formats de données.

Analyse et Résultats

Exploration des données

- **Fréquences des types** : Les types les plus fréquents sont analysés.
- **Avantages stratégiques** : Analyse des relations entre types pour identifier les plus forts/faibles selon les situations.

Prédictions

En utilisant des techniques de machine learning, nous avons prédit les probabilités de victoire en fonction des types et des statistiques des Pokémon.

- A MODIFIER ! Pas sûr pour cette partie et on a pas encore de graphiques.

Visualisations

Les résultats sont présentés sous forme de graphiques interactifs :

1. **Carte thermique** : Avantages/désavantages entre types.
2. **Histogrammes** : Répartition des types et leurs performances en combat.
3. **Courbes** : Taux de victoire moyen par type.

Conclusion et Perspectives

Ce projet met en évidence les interactions complexes entre les types Pokémon et leurs impacts sur les résultats des combats. Les travaux futurs pourraient inclure :

- L'intégration de données supplémentaires, comme les capacités spéciales.
- L'analyse des combinaisons multi-types pour des prédictions plus fines.