

Exploring Pokémon Data with Unsupervised Learning

Lucien BAUER

Master 2 Data Science
Université de Strasbourg

6 Février 2026

Vue d'ensemble

Objectif

Découvrir des patterns cachés dans les données Pokémon en utilisant l'unsupervised learning

Datasets

- **Pokémon** : ~1000 Pokémon avec stats (HP, Attack, Defense, etc.)
- **Moves** : ~900 moves avec descriptions textuelles
- **Learnset** : Qui apprend quoi ? (~50 moves/Pokémon en moyenne)

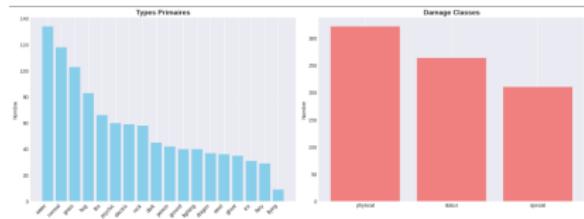
Question principale

Les Pokémon se regroupent-ils naturellement selon leurs caractéristiques ?

Part 1 : Understanding the Data

Statistiques clés

- 1000+ Pokémons
- 900+ moves
- ~50 moves/Pokémon
- 18 types différents



Valeurs manquantes

- type_2 : optionnel
- power : 0 pour status
- Approche sémantique

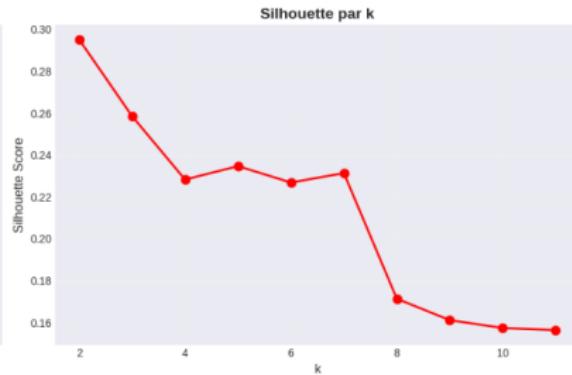
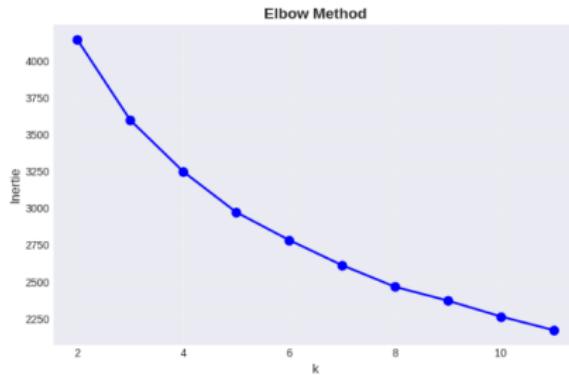
Observation

- Données déséquilibrées
- Water, Normal, Grass surreprésentés

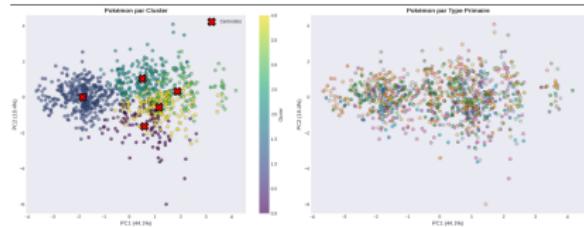
Part 2 : Clustering - Méthodologie

Choix techniques

- **Normalisation** : StandardScaler (préserve structure)
- **Algorithme** : K-Means (archétypes bien définis)
- **k=5 clusters** : Elbow + Silhouette ($\sim 0.35-0.40$)
- **Visualisation** : PCA (60% variance expliquée)



Part 2 : Les 5 Archétypes



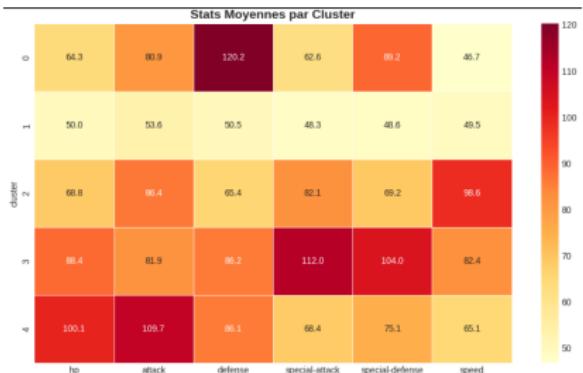
Clusters identifiés

- ① **Fast Sweepers**
Att. élevée + vitesse
- ② **Defensive Walls**
Def. + Def. Spé élevées
- ③ **Special Attackers**
Att. Spé élevée
- ④ **Bulky Pokémons**
HP élevé
- ⑤ **Balanced**
Stats équilibrées

Part 2 : Insight Principal

Découverte clé

Les clusters NE correspondent PAS aux types officiels !



Interprétation

- **Types = Résistances élémentaires**
- **Clusters = Rôles de combat**

Exemple

- Un Pokémon Water peut être dans *n'importe quel* cluster selon ses stats
- Le type définit "contre quoi il est fort"
- Le cluster définit "comment il joue"

Part 3 : Text Analysis

TF-IDF sur descriptions des moves

Révèle des sous-catégories fines au-delà de physical/special/status

Physical	Special	Status
damage	user	target
power	special	stage
attack	stat	effect
contact	turn	lowers
physical	target	raises

Patterns découverts

- Healing moves, Status effects, High-damage, Defensive, Stat modification
- Le texte capture la **mécanique**, pas juste la catégorie

Part 4 : Stats vs Moves

Question

Les Pokémon similaires en stats sont-ils similaires en moves ?

Approche

- Représentation move-based : counts + stats moyennes + types
- Comparaison de similarités

Découverte

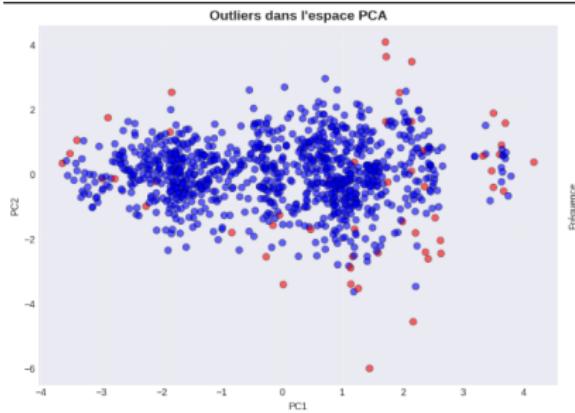
- Corrélation modérée
- Stats = POTENTIEL
- Moves = OPTIONS
- Les deux sont complémentaires

[Graphique : Corrélation
Stats vs Moves]

Part 5 : Anomaly Detection

Méthode

Isolation Forest (contamination = 5%)



Résultats

- ~5% outliers (50-60 Pokémons)
- Stats extrêmes ou distributions inhabituelles

Causes

- ➊ Une stat extrême
- ➋ BST exceptionnel
- ➌ Combinaison rare
- ➍ Spécialisation extrême

Part 5 : Légendaires vs Normaux

Découverte majeure

Les légendaires sont 5x plus susceptibles d'être des outliers !

Légendaires (BST > 580)		Normaux
Total	~100	~900
Outliers	15-25%	3-5%

Interprétation

- Design intentionnel : légendaires *doivent* être exceptionnels
- Les choix des développeurs sont **mathématiquement visibles**
- Pas soumis aux mêmes contraintes d'équilibrage

Conclusion

Découvertes principales

- ① 5 archétypes émergent des stats (Fast Sweepers, Walls, etc.)
- ② Les clusters révèlent des **rôles de combat**, pas des types
- ③ Le **text mining** capture des mécaniques au-delà des catégories
- ④ Stats et moves sont **complémentaires** (potentiel vs options)
- ⑤ Les légendaires sont **mathématiquement différents** par design

Insight général

L'unsupervised learning révèle la **structure cachée du game design**

Limites et Extensions

Limites

- K-Means force chaque Pokémon dans un cluster (pas de flou)
- PCA perd 40% de variance
- Text analysis limité par qualité des descriptions

Extensions possibles

- Analyse temporelle : évolution des designs à travers les générations
- Network analysis : graphes de similarité
- Supervised learning : prédire le tier compétitif
- Recommender system : suggérer des équipes équilibrées

Questions ?

Merci pour votre attention !