

# League of Legends Meta Analysis - Machine Learning

## Problématique générale

**Comment maximiser ses chances de victoire dans une partie de League of Legends en s'appuyant sur l'analyse des données de jeu et des modèles de machine learning, à travers les choix stratégiques (champions, items, draft) et les performances individuelles en partie ?**

Cette problématique vise à **comprendre, quantifier** et **modéliser** les éléments qui influencent la victoire, afin de transformer les données brutes en **recommandations stratégiques exploitables**.

## Phase 0 — Préparation et Qualité des Données

Pourquoi cette phase est cruciale :

- La qualité des données impacte directement la fiabilité des modèles
- Un nettoyage rigoureux évite les biais et erreurs de prédiction

Objectifs de la phase :

1. Rassembler et contrôler : Charger tous les fichiers CSV du dataset
2. Nettoyer : Gérer valeurs manquantes, données corrompues etc.
3. Échantillonner : Conserver uniquement les modes CLASSIC et ARAM (exclure les autres)

Tâches typiques :

- Identification et traitement des valeurs manquantes
- Validation des clés étrangères (ChampionId, ItemId, MatchId)
- Détection d'anomalies (runes à 0, sorts d'invocateur à 0, items invalides)
- Création d'une version "clean" des tables pour analyses ultérieures
- Segmentation et création de splits train/test (ex : 80/20 par patch ou split stratifié)

## Phase I — Compréhension de la méta actuelle (ce qui gagne)

Objectif :

- Identifier les champions et items qui contribuent le plus aux victoires, par contexte (lane, rank, durée, patch)

### I.1 — Efficacité des champions

Question de recherche

Quels champions présentent les meilleurs taux de victoire en fonction du rank, de la lane et de la durée de la partie ?

Actions proposées :

- Calcul des taux de victoire par champion
- Analyses croisées : Champion × Lane, Champion × Champion adverse
- Visualisation et comparaison des distributions
- Identification de tendances fortes et de champions surperformants

Objectif :

Mettre en évidence les champions dominants de la métaprogression

I.2 — Détection sur/sous-performance par lane et par rank

- Détection des sur- et sous-performers en comparant winrates par champion vs moyenne de la lane et par rank
- Filtrage par nombre minimal de matchs pour garantir la robustesse

---

## Phase II — Optimisation du champion (builds & items)

Objectif :

- Jouer un champion de manière optimale selon les données

Identification des builds optimaux

Question de recherche

Quels items sont le plus fréquemment associés aux victoires ?

Approches :

- Identification des items « core » par champion (ex : item présent dans  $\geq 50\%$  des victoires)
- Étude des items situationnels associés à un taux de victoire élevé
- Construction de "build\_key" / "build\_str" et agrégation sur les parties gagnées

Objectif :

Apprendre automatiquement les configurations d'items les plus favorables à la victoire.

---

## Phase II.b — Modélisation de la victoire (ML explicatif et prédictif)

Objectif :

- Formaliser la relation entre plusieurs variables et l'issue d'une partie

Question de recherche

Peut-on prédire l'issue d'une partie à partir des choix stratégiques et des statistiques de jeu ?

Approches Data Scientist :

- Préparation et nettoyage des données
- Séparation des jeux d'entraînement et de test
- Modèles : Régression logistique (interprétable), Random Forest / autres arbres (non linéaires)
- Évaluation : accuracy, AUC, classification report

Approche Expert IA / Machine Learning :

- Régression logistique pour interprétation des coefficients (features standardisées)
- Random Forest (ou Gradient Boosting) pour robustesse et importances des variables
- Analyse de l'importance des variables (feature importance) et comparaison inter-modèles

Limites :

- Variables post-game (gold, dégâts, CS) expliquent beaucoup mais induisent un biais d'interprétation (modèles explicatifs post-hoc)
- Variables pré-game permettent des estimations probabilistes mais pas de prédiction parfaite

Objectif final :

Obtenir un classement quantifié des facteurs de victoire et des modèles probabilistes d'aide à la décision.

---

## Phase III — Draft et contre-stratégies (analyse pré-game)

---

Objectif :

- Analyser l'impact des choix pré-game (champion, lane, matchup, rank, patch) sur la victoire, sans utiliser les statistiques post-game.

Question de recherche

Quels champions réduisent significativement le taux de victoire d'un champion donné ?

Actions :

- Analyse des matchups fréquents (Champion vs EnemyChampion par lane)
- Calcul des winrates champion contre champion (filtrer par volume minimal)
- Analyse relative (delta à la moyenne de la lane) pour identifier vrais counters
- Construction de tableaux de counters et identification des matchups défavorables

Objectif :

Construire des tableaux de counters et identifier les matchups défavorables.