FDS: Pokemon Battles prediction 2025 - Team dugTRIO

Line 60: emiliano.grandis5@gmail.com, eyahouij02@gmail.com, filippovitozzi@gmail.com

Filippo Vitozzi 2265118, Emiliano Grandi 2258965, Eya Houij 2238231

1. Feature Engineering

The dataset provides both initial team information and full turn-by-turn battle logs. From these sources we derive a compact set of matchup-level and temporal descriptors designed to capture strategic asymmetries and momentum shifts throughout the battle.

Our features include: (i) **stat-based differences** (speed, bulk, offenses, total base stats, lead advantage), (ii) **type- and coverage-related signals** (STAB differential, coverage differential, weak/strong attack potential, net attack advantage), (iii) **status and reliability indicators** (effective status differential, last-turn status effects), and (iv) **dynamic battle-flow metrics** such as HP momentum, alive-difference evolution, switch efficiency, first-strike potential, and round-count ratio.

These engineered features summarize complex battle trajectories into stable numerical representations, enabling downstream models to learn both structural matchup properties and temporal dynamics without processing raw sequences.

2. Models Evaluated

We experiment with three classes of supervised models of increasing expressive power: a linear baseline, a lightweight ensemble, and a gradient-boosted decision tree architecture. The goal is to progressively assess how much non-linearity is required to capture the strategic dependencies encoded in the engineered features.

2.1. Logistic Regression (Baseline).

Our first approach is a regularized Logistic Regression model trained within a standardized pipeline. We perform an extensive grid search over penalties (L1/L2), inverse-regularization strengths C, and optional class weighting. Despite careful tuning and 5-fold stratified validation, the model saturates around a training accuracy of 0.796, failing to exceed the 0.80 threshold.

2.2. Soft Voting Ensemble.

To introduce limited non-linearity while retaining interpretability, we train a soft-voting ensemble combining: (i) a scaled Logistic Regression classifier, (ii) an SVM with probabilistic outputs, and (iii) a shallow Decision Tree (max depth 4). This ensemble benefits from complementary inductive biases and successfully surpasses the 0.80 accuracy barrier on the training set. However, its performance remains unstable across validation splits, suggesting that deeper nonlinear modeling is required to fully exploit the engineered feature space.

2.3. XGBoost (Final Model).

Our best-performing model is an XGBoost classifier trained with histogram-based tree construction. We perform a 40-iteration RandomizedSearchCV over tree depth, learning rate, number of estimators, child weight, subsampling ratios, column sampling, and L1/L2 regularization. The best configuration achieves a cross-validated AUC of 0.8777. We then retrain the model on the training split with an 80-round early-stopping patience on the validation set, selecting the optimal number of boosting rounds and preventing overfitting. The final classifier is trained on the full dataset and delivers strong generalization, achieving a validation AUC of 0.8844 and a public-leaderboard score of 0.8213. This demonstrates that gradient-boosted trees are well suited to the complex, nonlinear structure of Pokémon battle dynamics.