Leonardo Sani 2059860, Mohammad Sorooshmehr 1947890, Simone Mantero 2014057

Abstract

Modeling began with a baseline **Logistic Regression (L2)** to establish a simple linear reference. To address feature redundancy and improve stability, the model was extended with **PCA**. Potential nonlinear interactions were explored using **Polynomial Logistic Regression**, which underperformed. To capture non-linear patterns more effectively, several tree-based methods were employed. While a **soft voting classifier** was explored to leverage complementary strengths, the single **XGBoost** model ultimately achieved the highest predictive accuracy (0.8532).

1 Feature Engineering

Our feature set captures several complementary aspects of competitive battles, including type effectiveness, base and boosted stats, Pokémon identity, temporal game flow, status conditions and damage-related quantities [Bulbapedia contributors(2024)].

1.1 Top Features

Average Final HP. $\frac{1}{6}\sum_{i \in P} h_i$, where h_i is the last observed HP percentage over 30 turns; unseen Pokémon are assigned value 1.

Voluntary switches Number of times that a pokemon with strictly positive Hp is changed.

Pokémon Encoding. Given the fixed species pool and uniform level, Pokémon identity is encoded directly : one-hot encoding for linear models and label encoding for tree-based models [Paul(2023)].

1.2 Feature Variants

To capture temporal patterns, each feature was split into early, mid and late intervals (0–10, 10–20, 20–30 turns). Inter-player differences can also be used to reduce dimensionality while preserving relative comparisons.

1.3 Correlation and Final Selection

Most features exhibited low correlation (Figure 1), with only a few expected overlaps. Manual feature pruning was found to reduce performance, so we instead applied automated techniques. For linear models, principal component analysis (PCA) was used to compress correlated features. For tree-based models, hyperparameter-based regularization, including colsample_bytree to randomly subsample features per tree, mitigated feature redundancy and improved generalization. This strategy maintained model robustness while simplifying inputs without sacrificing accuracy.

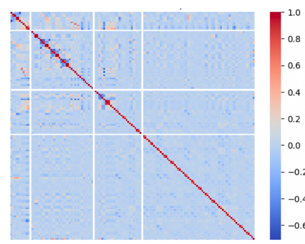

Figure 1: Correlation heatmap for the logistic-regression feature set

To further streamline experimentation, a dedicated feature-aggregation tool enabled rapid activation or removal of feature groups, producing tailored subsets for each model type.

2 Cross-Validation Performances

Table 1: Model accuracy (mean ± std, 10-fold CV)

Model	**Accuracy**
Logistic Regression (L2)	0.8504 ± 0.0065
Logistic Regression (PCA + L2)	$\mathbf{0.8516 \pm 0.0094}$
Logistic Regression (Poly + L2)	0.805 ± 0.0097
Random Forest	0.8376 ± 0.0118
XGBoost	$\mathbf{0.8532 \pm 0.0134}$
soft Voting Classifier (XGB + RF + LR(PCA))	$\mathbf{0.8517 \pm 0.011}$

References

[Bulbapedia contributors(2024)] Bulbapedia contributors. Damage — bulbapedia, the community-driven pokémon encyclopedia. https://bulbapedia.bulbagarden.net/wiki/Damage, 2024. Accessed: 2025-11-04. 1

[Paul(2023)] Rohan Paul. Ml interview q series: Explain one-hot vs label encoding and select the right encoding for a model. https://www.rohan-paul.com/p/ml-interview-q-series-explain-one, 2023. Accessed: 2025-11-04. 1