Binary Classification Model for Pokémon Battle Prediction

Machine Learning project of team "*Totally Spies*"

Eleonora Liani 2078147
Andrea De Marco 1999447
Marianna Esposito 2022242

Abstract

This paper discusses the design and building of an effective classification model that predicts the winner of a Pokémon battle from the perspective of "Player 1".
The model was trained on a dataset composed of thousands of battles from real human-vs-human 1st Generation battles on the Pokémon Showdown simulator.

1. EDA and Data inspection

A brief inspection of the first battle allowed us to determine the dataset's structure, particularly regarding team details (pokemon types, pokemon names, statuses and their distributions). We also checked for number of turns per battle and class balance before proceeding with our actual code.

2. Model Choice and Tuning

In evaluating different modeling approaches, logistic regression offered a simple and interpretable baseline. For this reason, it was the initial model implemented. However, the complexity and nonlinearity inherent in the features suggested the need for a more expressive model. Therefore Extreme Gradient Boosting (XGBoost) was chosen as the final model due to its parallel tree boosting capabilities.
To tune the best hyperparameters for the model, we randomly sampled them from a predefined range deemed optimal.

3. Features

3.1. Most effective

In choosing the best features to implement in the model, we mainly focused on: the number of K.O.s, of status and of seen pokémon in each team. Given the model choice, these dynamic features (with multiple checkpoints at different turns) proved to be the most suitable and impactful ones during the training phase. Indeed, these features exploited the fact that the development of the whole 30-turns battle was given as part of the data, not only the composition of the team members.

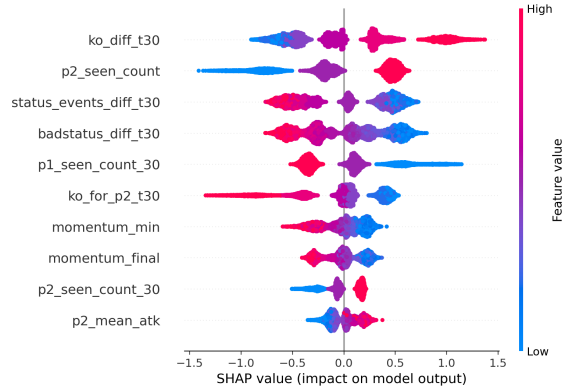

Figure 1. Top 10 features according to SHAP values.

Furthermore, all of these features take into consideration the difference between the two groups, in order to highlight the discrepancy among them.

3.2. Least effective

Some other important but secondary features were: momentum indicators, base stats of the teams and pokémon win-rate. As previously said, marginal features rely only on the knowledge of the team composition. We also implemented the pokémon official type chart to take into account attacks effectiveness, however this feature didn't affect the final accuracy therefore we removed it.

4. Model Training and Cross validation

We evaluated our model using a custom cross-validation procedure that recomputes win-rate features inside each fold (fold-safe), assembles all feature groups for train and validation, trains the selected model, reports accuracy/F1-score/AUC/precision/recall for each fold. This approach ensures that no information from the validation split leaks into the training process, making the evaluation reliable even when using derived statistics such as win rates.

5. Results

This model managed to reach an accuracy of 0.84.