Pokemon Predictions: Gotta Model 'Em All!

Member Names: Brandon Timok, Nicholas Casas

Members' role/contributions:

- **Brandon**: Preprocessing (normalization), cross-validation, report writing
- **Nick**: Idea development, dataset selection, optimization, result interpretation
- **Equal Contribution**: Architecture development for each model used

Dataset Used:

- Pokemon Dataset from Kaggle

Libraries Used:

- pandas
- numpy
- sklearn

1. Introduction

Pokémon is a popular franchise with a vast array of creatures, each belonging to one or more types (e.g., grass, fire, water). Understanding a Pokémon's type is crucial for effective gameplay strategies and team compositions. Additionally, predicting the hatch time of Pokémon eggs can provide valuable information to players. The goal of this project is to predict both the type and hatch time of Pokémon based on their characteristics, using machine learning techniques.

2. Computational Challenges

One of the main challenges faced in this project was the computational complexity of processing large datasets of Pokémon with numerous features. Additionally, implementing and tuning multiple machine learning models for both type and hatch time prediction required substantial computational resources and time.

3. Dataset Description

The dataset used for this project contains information about various Pokémon, including their abilities, base stats, typing, and base egg steps (which determine hatch time). The dataset was preprocessed to handle missing values and encode categorical variables for model compatibility.

4. Justification of Machine Learning Techniques

For type prediction, K-Nearest Neighbors, Random Forest, and Logistic Regression were chosen for their effectiveness in handling multi-class classification tasks. For hatch time prediction, Linear Regression was chosen due to its suitability for predicting continuous variables.

5. Approach to Realistic Datasets

The chosen machine learning techniques were applied to realistic datasets of Pokémon, which simulate the variability and complexity of real-world data. By training and evaluating the models on these datasets, we aimed to develop practical and reliable prediction systems for both type and hatch time.

6. Experimental Settings

- Machine Learning Methods for Benchmarking: In addition to the selected models, other methods such as Support Vector Machines and Gradient Boosting were considered but did not perform as well in initial experiments.
- Parameter Tuning: Grid search and cross-validation were used to tune the hyperparameters of each model to optimize performance.
- Evaluation Strategies: K-fold cross-validation with k=5 was employed to evaluate the performance of each model consistently.
- Experiment Repetition: Each experiment was repeated at least 5 times to ensure the stability and reproducibility of the results. The long computation time limited the amount of times we could test overall.

7. Discussion of Experimental Results

The experimental results showed that the Logistic Regression model outperformed the other models in terms of accuracy for type prediction, while the Linear Regression model achieved the lowest error for hatch time prediction. These results indicate the effectiveness of the chosen machine learning techniques for both tasks. It is important to note that these results are based on a random index selection of a Pokémon from the dataset for each prediction task. Nonetheless, these outcomes indicate the effectiveness of the chosen machine learning techniques for both tasks.

Hatch Time Prediction Results:

The output of the hatch time predictions includes three key error metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics assess the accuracy of the models in predicting the base egg steps for Pokémon. A lower value for MAE, MSE, and RMSE indicates that the model's predictions are closer to the actual values. The output includes its predicted base egg steps and the actual base egg steps for comparison.

KNN:

```
--------------------------------------------------
K-Nearest Neighbors Model
--------------------------------------------------

Mean Absolute Error (MAE): 1069.97
Mean Squared Error (MSE): 10458255.54
Root Mean Squared Error (RMSE): 3233.92
Pokémon: Beheeyem
Predicted base egg steps: 5120.0
Actual base egg steps: 5120
```

Random Forest:

```
--------------------------------------------------
Random Forest Model
--------------------------------------------------

Mean Absolute Error (MAE): 723.61
Mean Squared Error (MSE): 2616446.44
Root Mean Squared Error (RMSE): 1617.54
Pokémon: Beheeyem
Predicted base egg steps: 5077.33
Actual base egg steps: 5120
```

Linear Regression:

```
--------------------------------------------------
Linear Regression Model
--------------------------------------------------

Mean Absolute Error (MAE): 1409.56
Mean Squared Error (MSE): 7087184.85
Root Mean Squared Error (RMSE): 2662.18
Pokémon: Beheeyem
Predicted base egg steps: 5961.96
Actual base egg steps: 5120
```

Type Prediction Results:

The output of the type prediction models demonstrates their performance in classifying Pokémon types. The K-Nearest Neighbors model achieved 86.96% accuracy, with strong precision and recall for some types but struggled with others, such as Flying. The Random Forest model improved accuracy to 88.82%, showing strengths in predicting certain types like Dragon and Steel but still facing challenges with others. The Logistic Regression model performed the best, achieving 93.79% accuracy with consistent precision and recall across most types. These results illustrate the models' effectiveness in predicting Pokémon types, highlighting their strengths and limitations for different types.

KNN:

```
--------------------------------------------------
K-Nearest Neighbors Model
--------------------------------------------------

Accuracy: 86.96%
              precision    recall  f1-score   support

         bug       0.79      1.00      0.88        11
        dark       1.00      0.67      0.80         6
      dragon       0.67      0.80      0.73         5
    electric       0.83      0.83      0.83         6
       fairy       0.80      1.00      0.89         4
    fighting       1.00      0.91      0.95        11
        fire       1.00      0.91      0.95        11
      flying       0.00      0.00      0.00         0
       ghost       0.80      0.67      0.73         6
       grass       1.00      0.79      0.88        14
      ground       0.80      0.67      0.73         6
         ice       0.75      1.00      0.86         3
      normal       0.93      0.93      0.93        15
      poison       0.91      1.00      0.95        10
     psychic       0.82      1.00      0.90        14
        rock       0.82      0.90      0.86        10
       steel       0.75      1.00      0.86         3
       water       0.95      0.77      0.85        26
```

```
    accuracy                           0.87       161
   macro avg       0.81      0.82      0.81       161
weighted avg       0.89      0.87      0.87       161

Pokémon: Wartortle
Predicted type: water
Actual type: water
```

Random Forest:

```
--------------------------------------------------
Random Forest Model
--------------------------------------------------

Accuracy: 88.82%
              precision    recall  f1-score   support

         bug       1.00      0.64      0.78        11
        dark       1.00      1.00      1.00         6
      dragon       0.71      1.00      0.83         5
    electric       1.00      1.00      1.00         6
       fairy       0.80      1.00      0.89         4
    fighting       1.00      0.91      0.95        11
        fire       0.83      0.91      0.87        11
       ghost       1.00      0.67      0.80         6
       grass       0.92      0.79      0.85        14
      ground       0.71      0.83      0.77         6
         ice       0.75      1.00      0.86         3
      normal       1.00      1.00      1.00        15
      poison       1.00      0.70      0.82        10
     psychic       1.00      0.93      0.96        14
        rock       0.71      1.00      0.83        10
       steel       0.38      1.00      0.55         3
       water       1.00      0.92      0.96        26
```

```
    accuracy                           0.89       161
   macro avg       0.87      0.90      0.87       161
weighted avg       0.92      0.89      0.89       161

Pokémon: Wartortle
Predicted type: water
Actual type: water
```

Logistic Regression:

```
--------------------------------------------------
Logistic Regression Model
--------------------------------------------------

Accuracy: 93.79%
              precision    recall  f1-score   support

         bug       0.85      1.00      0.92        11
        dark       1.00      1.00      1.00         6
      dragon       0.80      0.80      0.80         5
    electric       1.00      0.83      0.91         6
       fairy       0.80      1.00      0.89         4
    fighting       1.00      1.00      1.00        11
        fire       1.00      1.00      1.00        11
      flying       0.00      0.00      0.00         0
       ghost       1.00      0.83      0.91         6
       grass       0.92      0.79      0.85        14
      ground       1.00      1.00      1.00         6
         ice       1.00      1.00      1.00         3
      normal       1.00      1.00      1.00        15
      poison       0.91      1.00      0.95        10
     psychic       1.00      0.93      0.96        14
        rock       1.00      0.90      0.95        10
       steel       0.75      1.00      0.86         3
       water       0.96      0.92      0.94        26
```

```
    accuracy                           0.94       161
   macro avg       0.89      0.89      0.89       161
weighted avg       0.95      0.94      0.94       161

Pokémon: Wartortle
Predicted type: water
Actual type: water
```

8. Conclusion

In conclusion, this project successfully demonstrated the feasibility of predicting both the type and hatch time of Pokémon using machine learning techniques. The results highlight the importance of choosing appropriate models and features for each prediction task. Future research could focus on further improving the accuracy and efficiency of the prediction models, as well as exploring additional features or datasets to enhance the predictions.

9. Miscellaneous

- Obstacles Overcome: One of the main obstacles was the preprocessing of the datasets to handle missing values and categorical variables. This was addressed by carefully designing preprocessing steps and ensuring data integrity throughout the process.
- Issues Encountered: An issue encountered was the imbalance in the distribution of Pokémon types, which could affect the performance of the models. This was mitigated by using appropriate evaluation metrics and techniques for imbalanced datasets.
- Future Plans: In future research, we plan to explore advanced machine learning techniques and ensemble methods to further improve the prediction accuracy for both type and hatch time prediction. Additionally, we aim to incorporate additional features, such as Pokémon abilities and evolutionary stages, to enhance the predictive power of the models.