Pokémon Stat Analysis

Ricky Wong
June 25, 2022

Summary

- Linear regression model of Pokémon game stats to get an edge in competitive Pokémon battles.

- Find the best variables to predict a Pokémons attack

Outline

- Business Understanding
- Data & Methods
- Results
- Assumption checks
- Accuracy
- Conclusion

Business Understanding

- Competitive Pokémon players trying to get an advantage by predicting new generation Pokémons attack and evaluate its viability.

- Use Pokémon game stats data to create linear regression model

Data & Methods

- Data from Kaggle which includes Name and individual stats.
- Check for outliers and mega evolutions then remove them
- Check multicollinearity and normal distribution of data
- Categorical data transformed using one-hot encoding

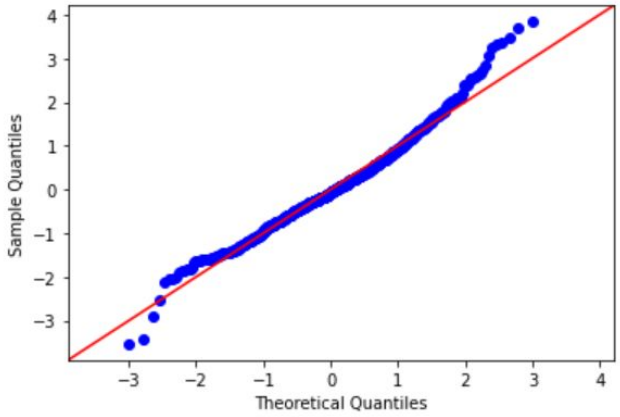
Results - Model 1 & Model 2

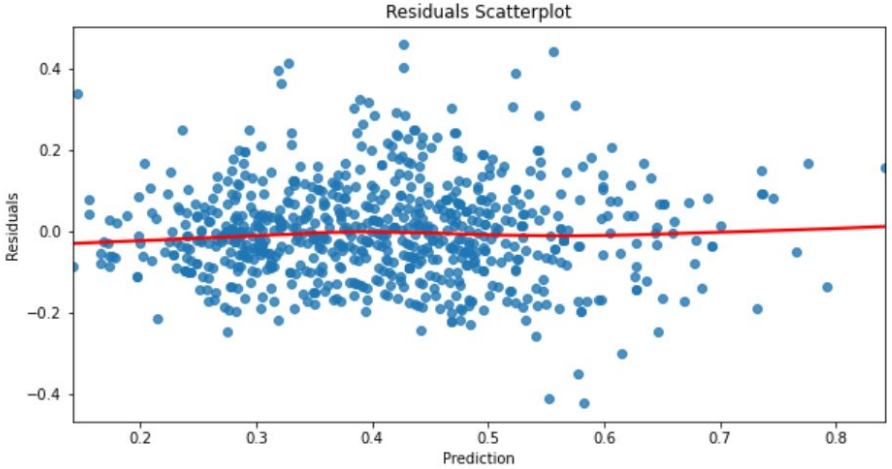
- Used all the available predictors
- R-Squared value of 0.532
- Removed predictors that were not as statistically significant

- Removed p-values over 0.05 lowered R-squared to 0.514
- Still have some predictors over 0.05
- No predictors are highly correlated to another.
- Ran RFE to top 9 predictors

Results - Model 3

- 50.3% of the variance is explained by the model.
- Normality and Homoscedasticity check

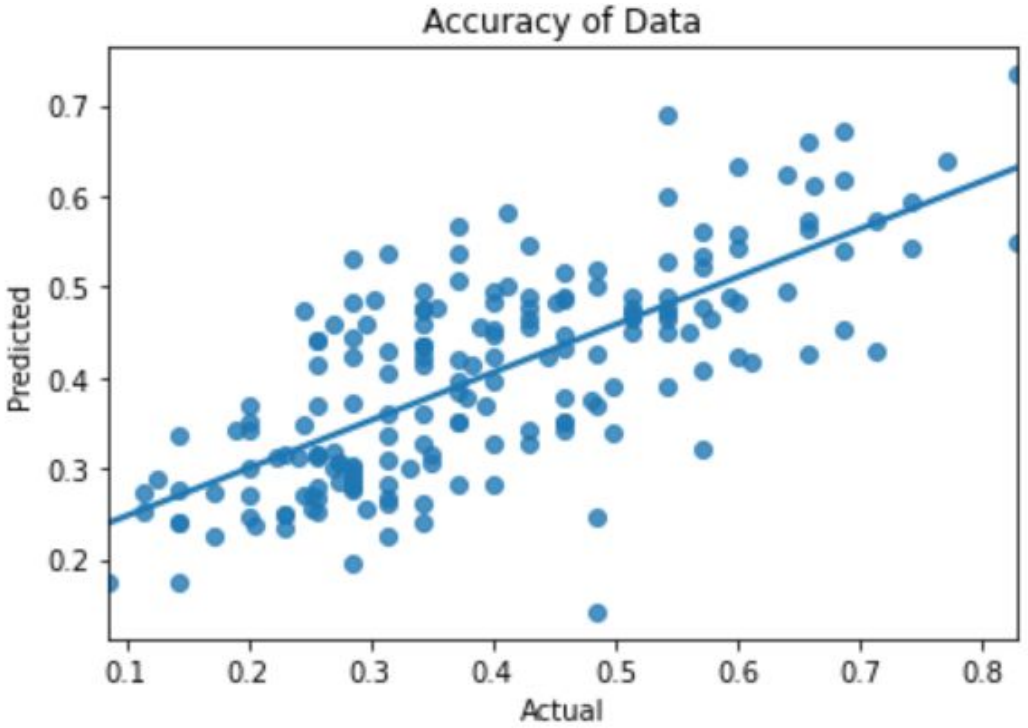




Accuracy

- Train and test MSE very similar

- Accuracy of the model is 51.43%



Conclusion

- HP and defence are best predictors

- Attack stats alone is not enough to determine if a Pokémon is strong or not

- A model that includes Pokémon attacks and type advantage may be more useful

Thank you!

Email: wong_ricky@hotmail.com

GitHub: @wong-ricky