Understand Elite Powerlifters' future performances

Yunzhe Yu

Brief Introduction About Powerlifting

Kyota Ushiyama

Joy Sistina

Jamal Browner

Composed of three movements: Squat, Bench and Deadlift AKA (SBD).

Lifting the heaviest gets the highest rank (Same weight class)

DOTs, or Wilks Score to measure from different weight classes.

Problem Statement

Object: My project aims to predict the future career outcomes of elite powerlifters using historical competition data. Seeking to understand factors that most significantly influence their career longevity and success.

Methods

Data Collections: Data collected from openpowerlifting, focusing on elite athletes.

Data Preparation: Converting different rating tools into DOTs, which used by USAPL

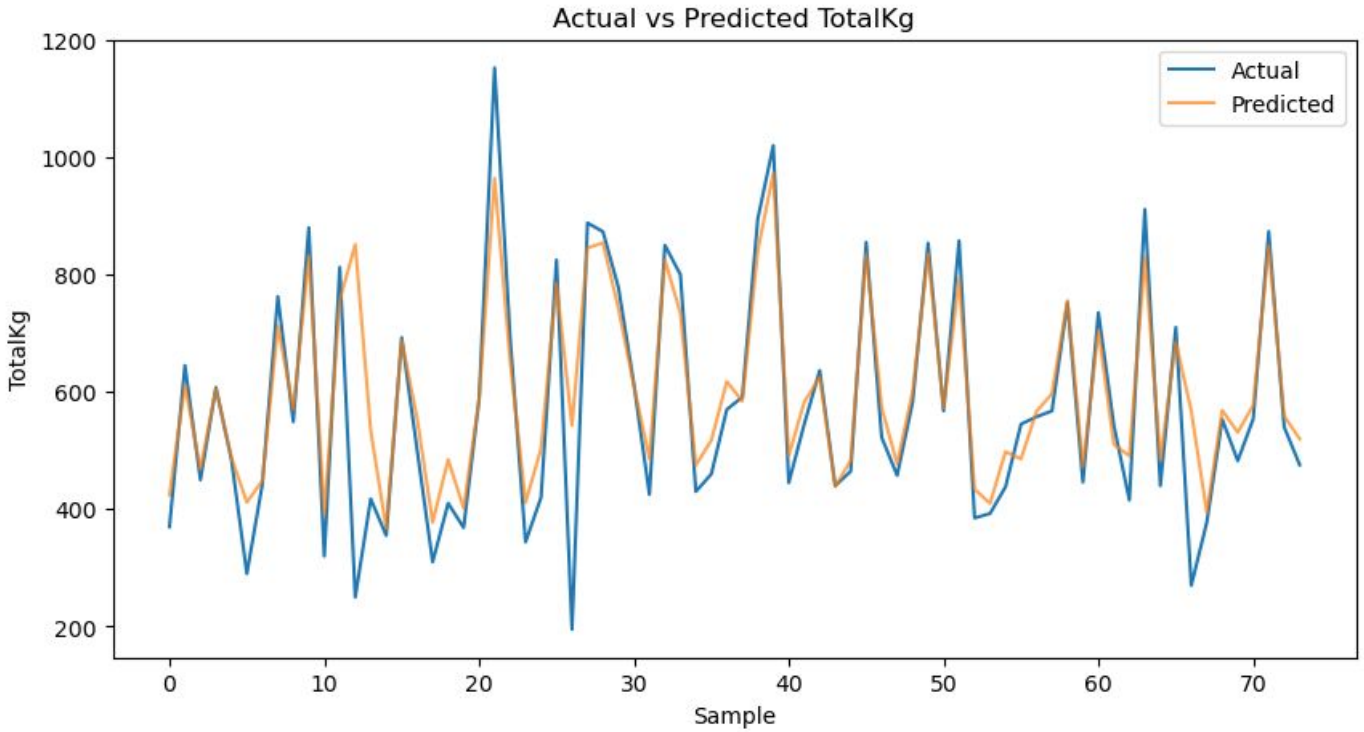
Feature Engineering: Developed features such as competition frequency, year-over-year performance improvements, and athlete's age

Model Development: Chose a Long Short-Term Memory model due to its ability to learn from sequences and remember long-term patterns, which is ideal for tracking athletes' performance over time.

Challenge

1. Age
   - Some Athletes' age information are not available for on the sheet.
   - Using existing meeting date and age to predict the age
2. Which coefficient to choose from
   - Different federations have different coefficient measurement.
   - Choosing only one coefficient which is DOTs.
3. LSTM Model Complexity
   - Complex LSTM Model does not perform well, causing overfitting and underfitting.
   - Using a simpler Model

Res


Actual vs Predicted TotalKg
Actual
Predicted
TotalKg
Sample
1200
1000
800
600
400
200
0
10
20
30
40
50
60
70

Conclusion

Feature engineering is important. By identifying and selecting pertinent attributes such as competition frequency and performance improvements, I learned how to transform raw data into meaningful predictors that significantly enhance a model's predictive capability.

Then, I learned how to adjust hyperparameters and evaluate their effects on model performance. This process also highlighted the importance of balancing model complexity with the computational limitations of the hardware at hand.

Data preprocessing presented its own set of challenges, especially when faced with missing values. The creative and logical strategies I developed to impute these gaps not only preserved the integrity of the dataset but also reinforced the meticulous nature required in data science.

Moreover, this project emphasized the importance of interdisciplinary knowledge. Understanding the nuances of powerlifting and the variables affecting an athlete's performance was crucial in interpreting the data accurately and making informed decisions throughout the modeling process.