Link to the code: https://github.com/EzioA666/DS340/tree/main/Project

Author: Yunzhe Yu

Introduction:

As someone who has competed in powerlifting, I'm always looking for new ways to boost my performance. In powerlifting, we focus on three main lifts: the squat, the bench press, and the deadlift. Success in this sport is measured by the total weight you can lift across these three events—the heavier, the better. I'm interested in finding out how data science can be integrated into my daily training to help me lift more and climb higher in the rankings.

Methodology:

In this project, I gathered data from the OpenPowerlifting website, a reputable source for global powerlifting rankings. My focus was on elite powerlifters, specifically those who have competed in world championships, due to their high frequency of participation in competitions. This ensures a robust dataset for analysis. I selected twenty-four athletes from the recent SBD Sheffield World Championship for my study.

The first step in my analysis involved cleaning the data. I removed several columns that were not pertinent to my objectives, including 'AgeClass', 'BirthYearClass', 'WeightClassKg', 'Squat4Kg', 'Bench4Kg', 'Deadlift4Kg', 'Wilks', 'Glossbrenner', 'Goodlift', 'Country', 'State', 'MeetCountry', 'MeetState', 'MeetTown', 'Federation', 'ParentFederation', and 'MeetName'. These columns were excluded as they do not contribute to the performance analysis.

Next, I addressed the issue of missing age data. Some athletes did not have recorded ages. To estimate these missing values, I used the last known competition date and age, applying a forward-fill method within each athlete's data group. This approach helped in maintaining a continuous set of age data for better analysis.

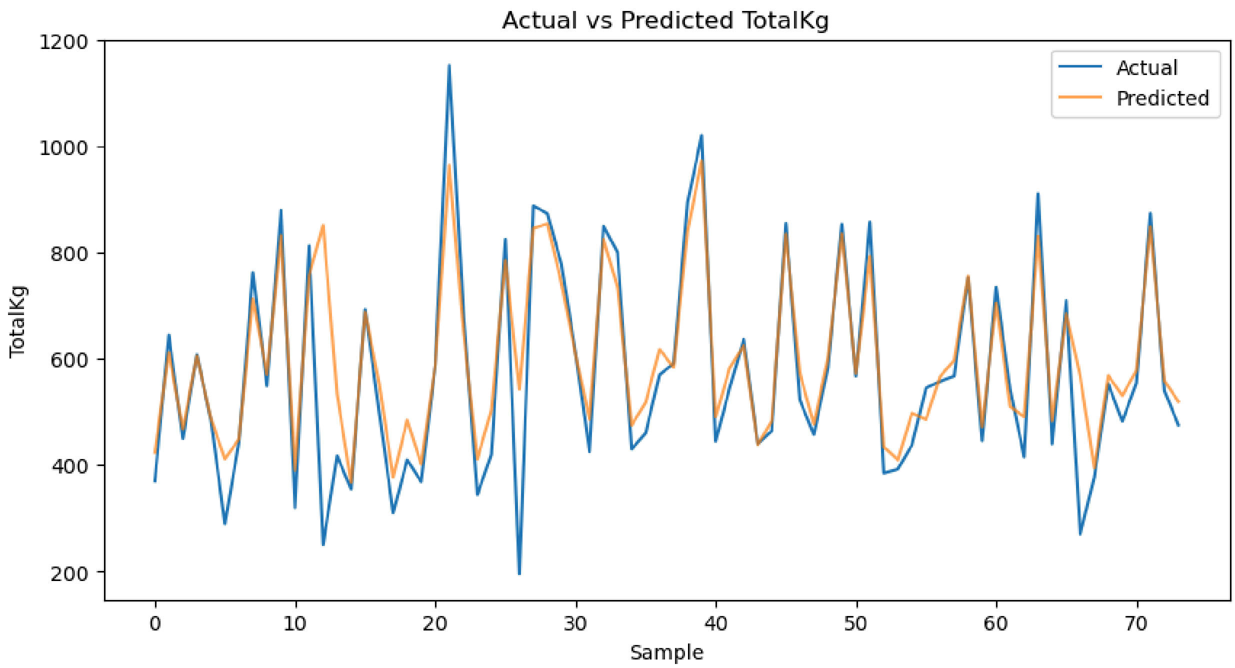
Following data preparation, I moved on to feature engineering. I transformed the 'Date' column into datetime format and added a 'Year' column to denote the year of the competition. I also computed each athlete's competition frequency and analyzed year-over-year performance improvements by comparing their maximum total lifts across years.

The core of my project was developing a predictive model using LSTM networks. I normalized the 'TotalKg' feature to prepare for training and created sequences that the LSTM model could use to learn patterns in performance over time. After training the model, it achieved a mean squared error of 0.299, indicating a strong predictive performance.

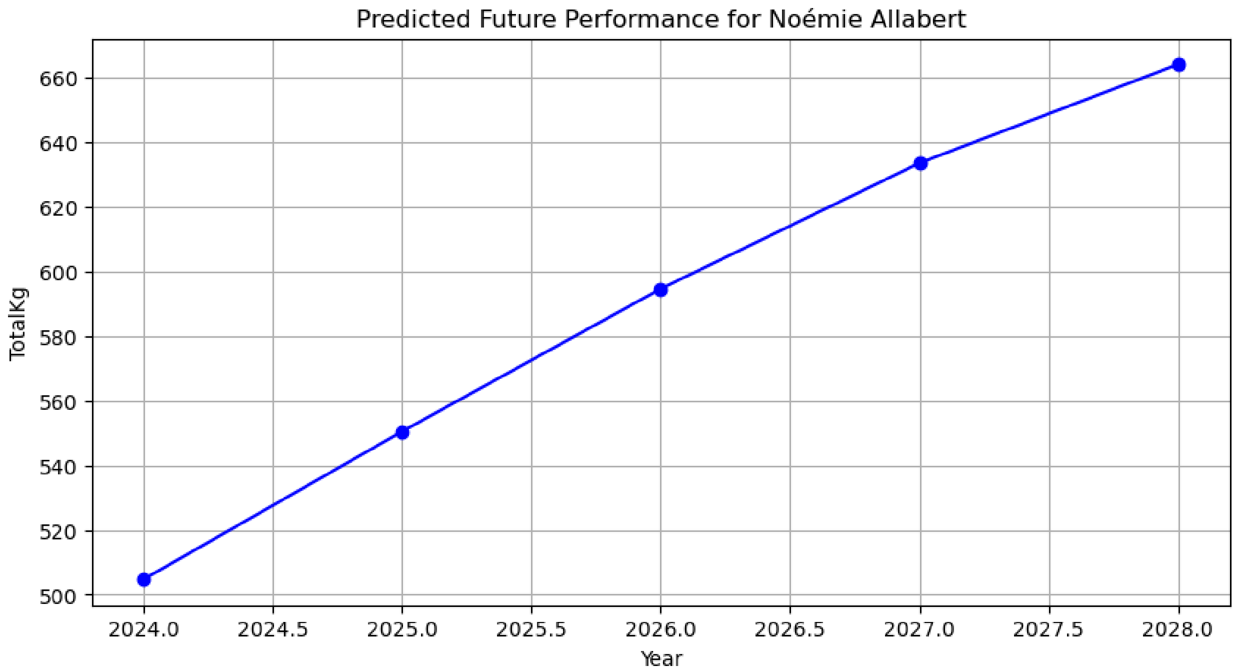
Finally, I used this model to gain insights into the potential performance trajectories of elite powerlifters, guiding future training and competition strategies.

Results:

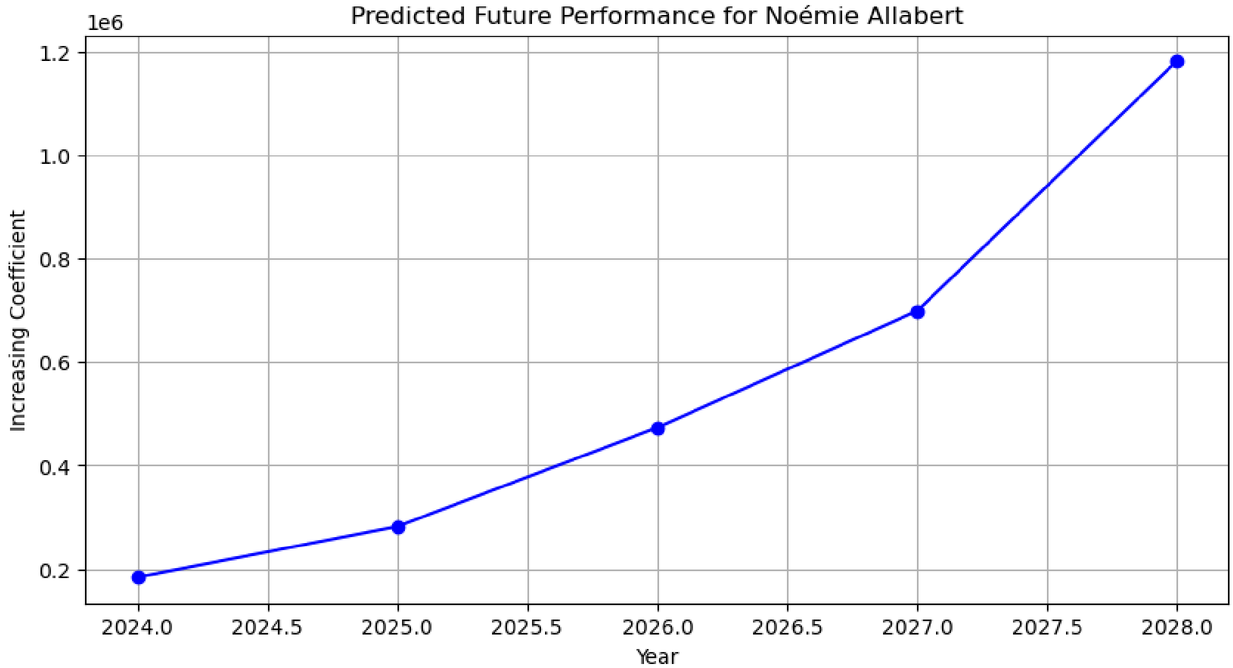
.



Initially, I conducted a predictive analysis of athletes' future performances using historical data. The resulting graph clearly illustrated that the predicted outcomes closely aligned with the actual performances, confirming the reliability of the model as a solid foundation for further exploration. Building on this success, I also specifically applied the model to forecast the future performance of Noemie Allabert, a distinguished athlete in the dataset. This targeted prediction helps in understanding her potential trajectory in upcoming competitions and provides valuable insights into how she might optimize her training and strategy based on the model's outputs.



This the application of the model to forecast Noemie Allabert's performance over the next five years. The predictive graph generated from this analysis shows a consistently linear improvement in her performance, with gains recorded annually. This trend suggests that Noemie is likely to enhance her abilities steadily, year over year. Such insights are invaluable not only for Noemie in strategizing her training regimen and competition schedule but also for her coaches to tailor specific aspects of her training to ensure continued progress. This visualization of linear growth highlights the model's potential to serve as a strategic tool in long-term athlete development.



Following the initial analysis, I enhanced the model by integrating additional features such as competition frequency, bodyweight, and age, aiming to provide a more nuanced understanding of performance dynamics. I then calculated a performance coefficient for Noemie Allabert, which factors in these variables to assess her potential for improvement. The updated graph reveals a promising increase in Noemie's performance coefficient over time, indicating that her capabilities are not only improving but accelerating. This refined model offers a more comprehensive view, suggesting that her training adaptations and physical development are effectively contributing to her progressive performance enhancements. Such detailed insights can be pivotal for fine-tuning training strategies and optimizing performance in competitive settings.

Conclusion:

an analysis of the general performance graph for the athlete reveals distinct peaks and troughs in performance over time. This visual representation allows us to predict potential peak and downturn phases in the athlete's performance trajectory. With these insights, we can strategically adjust the training schedule to capitalize on predicted peak periods and mitigate downturns. Specifically, by aligning training intensity and recovery periods with these forecasts, we can optimize the athlete's preparation and ensure they are peaking at the right moments. This proactive approach not only enhances performance but also helps in managing the athlete's physical conditioning more effectively, aiming to extend their peak performance years and reduce the risk of injury during expected lows.