The Dynamics of Basketball Performance

Matteo Saterini- 69602

April 23, 2024

Project Work Description

During the course of the data mining project, I had the opportunity to understand and learn how to conduct in-depth statistical analyses on a real-world dataset. I encountered machine learning algorithms for the first time, such as the fuzzy c-means clustering algorithm and analysis using SVD. Collaborating with my group partner, Mattia, we divided the workload as follows: we both wrote the necessary code (I in R, he in Python) for performing linear regression and PCA/SVD. To better parallelize the work, we then each took on a specific task: I handled the Exploratory Data Analysis, while he focused on the Clustering part. To address this asymmetry and ensure both of us were prepared on the entire project rather than just one aspect, we decided to cross-write the project report; I wrote the clustering section, and he wrote the EDA section. Regarding the linear regression and PCA/SVD portion, we decided to collaborate closely on it.

Overall, it was a rewarding experience collaborating with Mattia. As two Erasmus students passionate about studying these topics, we dedicated a significant amount of time discussing and working on this project together. I believe our shared enthusiasm for the subject made us compatible partners. Despite our differing backgrounds – with Mattia leaning more towards computer science and programming, and myself more inclined towards statistics and interpretation – we complemented each other well. This diversity allowed us to fill in each other's gaps and effectively accomplish our objectives.

Lessons Learned

As for me specifically, I felt a strong connection to this dataset due to its relevance to my personal background. Having played basketball for years and now being a coach, reflecting on the subject and extracting insights proved both beneficial and enjoyable. For an athlete/coach, staying abreast of the changes that a sport undergoes is a fundamental aspect of the improvement process.

Being a student of Mathematical Engineering, specializing in the Statistical Learning major, this project holds pivotal importance for someone aspiring to become a data scientist. It serves as a golden opportunity as it not only teaches the necessary machine learning methods but also prompts you to ask the right questions. For instance: Why am I getting this result? What is it telling me? Are there underlying reasons? What possible ways exist to extend the model? Is the analysis robust? In particular, I really think that our interpretations about the "upsets" phenomenon, and also the meaning of the principal component are really fascinating and may be of great interest for basketball experts.

Closing

In conclusion, this project was a rewarding experience, that allowed me to apply my data analysis studies and develop field-related skills and soft-skills related to collaborating with a teammate, even though our mutual professionalism let us get along with each other really well.

Machine learning algorithms are crucial, but they hold little meaning without proper interpretation, and I believe this course has equipped me with the necessary tools to achieve that.