1.1: The History and Tools of Machine Learning

By Daniel Mueller

Chosen Case Study *"Simulating World Cup Matches."*

❖How is machine learning used in the following environment? What does it that humans can not?

In the realm of sports analytics, machine learning offers capabilities that far surpass human limitations, especially in predicting complex tournament outcomes like the FIFA World Cup. By analyzing vast datasets encompassing team statistics, historical performances, and player metrics, machine learning models can identify intricate patterns and relationships that are not immediately apparent to human analysts.

The case study was classified as a **classification problem**, as it aimed to predict **discrete values**—specifically, whether **Team 1 or Team 2 would win** during the **knockout stage** of the **2018 FIFA World Cup**. The algorithms used included **logistic regression**, a **random forest classifier**, and a **linear support vector classifier (SVC)**.

These algorithms were trained on **binary classification data**, where:

- **1** = Team 1 wins
- **-1** = Team 2 wins

These labels were determined by subtracting the number of goals scored by Team 2 from Team 1 in historical match data.

After running the models, the **SVC algorithm achieved 70% accuracy**, which was sufficient to build a **simulation model**. This model used a **Monte Carlo simulation**, a technique for estimating possible outcomes in uncertain situations—such as knockout matches—by applying **random sampling** to generate probable results.

One significant advantage of machine learning in this environment is its ability to handle vast datasets and perform complex computations rapidly. For instance, simulating an entire tournament involves calculating the probabilities of each possible match outcome, which can number in the millions.

The machine learning model in this case study processed a **huge amount of historical match data**, identifying patterns that might be **too time-consuming** or **difficult for humans to detect manually**. This allowed for more accurate predictions of match outcomes.