

Final Research Project Report

Project Title: From data to theory: analyzing dynamics in team sports

Duration: October 2018 - August 2023

Project Team (after the end of the project):

- Principal Investigator: Laszlo Toka
- Researcher: Gergely Biczok
- PhD candidate: Pegah Rahimian
- MSc and BSc students: Balazs Mihalyi, Kristof Koflanovics, Benedek Zentai, Csaba Czuth, Fanni Berkovics, Hanna Biro, Patrik Szmidá, Patrick Bitá, Daniel Kovacs, Donat Hajszer

Project Overview:

The sports analytics research project, spanning five years from October 2018 to August 2023, aimed to advance the field of sports analytics through data-driven research, model development, and extensive publication efforts. The project was conducted in collaboration with various academic institutions and sports organizations, with a primary focus on soccer analytics. The project evolved each year, adapting to new challenges and opportunities. The topic gained an extremely large traction among computer science students at the Budapest University of Technology and Economics (BME).

Goals and contributions summary:

The project aimed to develop mathematical models and analytical results in the field of sports analytics. Additionally, the project aimed to create a freely available analytics framework for sports analytics. A patent was planned to be filed for this framework. The main goal was to advance the quality of sports analytics research in Hungary and enhance the reputation of domestic engineering research.

The project involved data acquisition, data processing, and result publication over the four plus one years. In the first year, the focus was on obtaining sports data, designing a database layout, and conducting preliminary data analysis. In the second year, mathematical models and data mining methods were developed, and player- and team-wide analysis was performed. The third year saw the development of more complex models, real-time analytics, and a patent filing. The fourth year aimed to validate models and the analytics framework on data from Europe, with plans to publish in open-access journals. The fifth additional year completed the publication phase.

The project also aimed to engage students, researchers, and collaborators, and regular open meetings and seminars were organized to promote science. The project targeted several conferences and journals for publication, and it was expected to yield a minimum of 1-2 journal articles and 2 other conference publications yearly on average. Additionally, the project planned to involve PhD, MSc, and BSc students in research tasks and thesis preparation. In total, 8 journal articles and 12 conference papers were published, while dozens of BSc and MSc theses were written in the topic.

Next, a yearly overview of planned tasks and key achievements is provided.

Year 1 (Oct 2018 - Sep 2019):

Research Focus:

- Data acquisition (play-by-play, extended play-by-play, and video tracking data)
- Database design and creation
- Statistical calculations and analytical model development
- Software development for data analysis
- Taxonomy-like article summarizing the state of sports analytics
- Conference paper and journal article publications

Key Achievements:

- Expanded the project team to include students from BME, specializing in data mining, deep learning, and analytics framework development.
- Submitted and presented two papers at BME's Students' Scientific Conference 2019.
- Established a strong collaboration with the Hungarian Football Federation (MLSZ) for data sharing.
- Collected football raw data from multiple sources besides MLSZ.
- Created multiple databases for versatile analytics.
- Published research results at international and national conferences.
- Developed the first version of the analytics framework, including data preprocessing and web components.
- Initiated work on a survey article.
- Submitted a journal article titled "Different creatine kinase responses to training and match load in elite youth soccer players."
- Presented papers at the European twin conference of KDD (ECML-PKDD).

Year 2 (Oct 2019 - Sep 2020):

Research Focus:

- Characterization of data based on existing models
- Research of novel data descriptive models
- Inspection of database scaling properties
- Porting the sports analytics framework to big data technologies
- Conference paper and journal article publications

Key Achievements:

- Coordinated and expanded the student team at BME.
- Supervised BSc and MSc theses.
- Continued collaboration with MLSZ.

- Conducted player- and team-wide analysis using deep learning and statistical methods.
- Developed mathematical models and data mining processes.
- Enhanced the analytics framework, including player detection and tracking components.
- Published papers at ECML-PKDD and other conferences.
- Submitted articles to journals, including Sports Health.

Year 3 (Oct 2020 - Sep 2021):

Research Focus:

- Advanced analytics models
- Real-time data evaluation using big data technologies
- Intensive publication activities
- Patent application
- Conference paper and journal article publications

Key Achievements:

- Managed the student team at BME.
- Supervised BSc and MSc theses.
- Initiated collaborations with new partners.
- Conducted complex player, team, and league-wide analyses.
- Developed and validated mathematical models.
- Expanded analytics into other sports.
- Submitted papers to MIT and KDD.
- Published articles in various conferences and journals, including Sports Health.

Year 4 (Oct 2021 - Aug 2022) :

Research Focus:

- Validation and fine-tuning of models and procedures
- Testing the analytics framework in other sports
- High-volume publishing
- Conference paper and journal article publications

Key Achievements:

- Continued coordination of the student team at BME.
- Supervised BSc and MSc theses.
- Continued collaborations with existing and new partners.
- Conducted comprehensive analyses across various sports.

- Developed and refined mathematical models.
- Expanded analytics into new domains.
- Achieved recognition with papers reaching the second round of evaluation at MIT Sloan Sports Analytics Conference 2022.
- Published papers in ECML-PKDD MLSA and other forums.

Year 5 (Sep 2022 - Aug 2023):

Research Focus:

- Continued validation and application of models and analytics framework
- Final round of publishing

Key Achievements (Up to August 2023):

- Coordinated the student team at BME.
- Supervised BSc and MSc theses.
- Maintained collaborations with partners.
- Continued research and analysis across multiple sports.
- Accepted papers at ECML-PKDD MLSA 2023 and Opta Forum 2023.
- Secured a patent for an online gaming apparatus.
- Submitted additional journal papers and conference papers.

Publications:

- Numerous student conference papers, international workshop papers, international conference papers, and scientific journal articles.
- Notable awards and recognitions received for some publications.
- Continuing efforts to publish the remaining articles, with several already accepted.

Budget:

- The project budget has been expended, except for a part of Open Access fees.

Conclusion:

The Sports Analytics Research Project has made significant contributions to the field of sports analytics over five years. The project achieved its goals by conducting extensive research, developing advanced machine learning models, and publishing research findings in various top-tier international forums. The team's dedication and collaborative efforts have led to valuable insights and innovations in sports analytics. This report serves as the culmination of the project's achievements and reflects its ongoing commitment to advancing the field.