

Análise de Táticas em Tênis Profissional com Descoberta de Subgrupos

Orientador: Renato Vimieiro

Pedro Cimini Mattos de Freitas
Universidade Federal de Minas Gerais, Belo Horizonte, Brasil
ciminipedro@gmail.com

Abstract—Nesse projeto, buscamos realizar a replicação do artigo [19]. Para isso, selecionaremos os dados manualmente anotados de jogos de Roger Federer e aplicaremos Descoberta de Subgrupos. Temos a disposição 7722 pontos. Desses, 3686 são pontos de serviço. Focaremos em encontrar padrões em pontos de serviço bem sucedidos a partir dos atributos desses dados. Vamos mostrar que é possível aplicar a metodologia do artigo original em outros dados e retornar resultados estatisticamente significativos, demonstrando assim a eficácia da metodologia original mesmo quando retirada de seu contexto original.

I. INTRODUÇÃO

A utilização de dados em esportes está se tornando cada vez mais comum, como podemos ver no **basquete**[1], no **futebol**[3] [13] e no **beisebol**[20]. **Tênis**, também, é um esporte que, cada vez mais, busca a utilização de dados para o desenvolvimento do próprio esporte e seus praticantes[26]. O uso de dados sobre diferentes aspectos do jogo, como saques, ralis e golpes, já é utilizado pelos melhores atletas [24] para avaliar e melhorar seu desempenho. Na primeira parte do projeto, fizemos uma análise estatística extensa em cima de dados de milhares de jogos para 3 jogadores específicos, na expectativa de encontrar características específicas a cada jogador, presentes nos dados, que nos permitiriam diferenciar entre o estilo de jogo de cada um deles. Os resultados mostraram que é possível discernir essas diferentes características a partir dos dados. Além disso, através da utilização de mineração de sequências, conseguimos encontrar padrões frequentes nas ações dos jogadores. A partir desses resultados, portanto, queremos dar o próximo passo nessa pesquisa. Tendo conhecimento que os dados são capazes de nos retornar informações relevantes quanto às características dos jogadores, queremos extrair informações que se repetem durante diversos jogos e pontos. Ou, de forma mais direta, queremos extrair táticas.

Entendemos que existem padrões frequentes (táticas) que se repetem nos dados e que, através deles, podemos encontrar informações sobre os jogadores[5]. Essas informações seriam muito valiosas para treinadores e jogadores, pois permitiriam uma análise mais aprofundada do estilo de jogador e a identificação de áreas de força e de fraqueza, que poderiam ser exploradas ou aprimoradas, por meio de treinos específicos ou mudanças nas táticas utilizadas, dessa forma sendo muito

útil para um aumento de desempenho dos jogadores. Para encontrar esses padrões, utilizaremos a técnica de **Descoberta de Subgrupos**[17]. Neste projeto, estaremos tentando replicar os resultados encontrados em [19], onde os autores foram bem-sucedidos em encontrar padrões em pontos de serviço. No artigo, os pesquisadores coletaram dados de 31 jogos de tênis, fazendo uma anotação ponto-a-ponto das jogadas em cada um dos jogos. Essa anotação contém desde informações mais gerais, como o número de rebatidas de um ponto e como ele foi finalizado, até informações mais específicas, como a direção e o tipo de cada rebatida. Nesses dados, é feito ainda um processo de engenharia de *features* para enriquecer ainda mais as informações sobre o ponto. Para a experimentação, foi utilizado Descoberta de Subgrupos[17], encontrando subconjuntos de interesse para caracterizar pontos de saque bem sucedidos. A ideia é comparar os padrões de um jogador "X" de interesse contra os padrões de seus oponentes e, dessa forma, encontrar padrões que são específicos ao jogador de interesse. O conjunto de dados usados será o mesmo que foi utilizado na primeira parte do projeto, por ter características similares ao encontrado no artigo que estamos tentando replicar, e por conter um número elevado de pontos para serem analisados.

II. REFERENCIAL TEÓRICO

A técnica de Descoberta de Subgrupos [17] é uma técnica supervisionada de mineração de dados que busca encontrar subconjuntos de dados onde a distribuição de um variável específica é diferente da distribuição dessa mesma variável quando levando em conta o conjunto de dados completo. Essa técnica é usada para diversas aplicações: em [2], ela é usada para encontrar padrões interessantes em visualizações de mapas, o que permite encontrar informações demográficas relevantes. Similarmente, em [15], o uso de Descoberta de Subgrupos é capaz de descobrir padrões no fluxo de migrações de distritos, encontrando uma relação entre alto fluxo de migração e desemprego. Podemos ainda citar o uso de Descoberta de Subgrupos para analisar o tráfego do Reino Unido[14], para descobrir relações entre os métodos de aprendizado, as atividades realizadas e a nota final de um aluno num contexto de aprendizado remoto [4], além de seu uso em contextos médicos, como para diagnósticos de câncer [10]. Esses são

alguns exemplos de aplicações e eficácia dessa técnica para encontrar padrões/relações em diferentes contextos. Porém, estamos interessados em outro contexto: o esportivo.

Técnicas de mineração de dados já são amplamente utilizados[25]. Nesse contexto, também, podemos citar o uso de Descoberta de Subgrupos com resultados positivos. Em [16], os pesquisadores utilizaram Descoberta de Subgrupos, com modelagem linear, para encontrar modelos interessantes a partir dos dados, visando encontrar padrões interpretáveis, que poderiam servir de guia para aprimoramento no regimento de treinos de um patinador. No vôlei, podemos citar a utilização de Descoberta de Subgrupos tanto para análises de desempenho, como em [6], onde a técnica é utilizada para encontrar relações entre o tipo de treinamento realizado e características físicas do jogador, com o seu desempenho ofensivo, quanto em [7], onde ela é utilizada para monitorar a situação física de jogadores de elite e, dessa forma, monitorar possíveis lesões. Em [22], Descoberta de Subgrupos é utilizado para encontrar táticas bem sucedidas de times de futebol, através da análise de dados de *tracking*, o que permitiu encontrar padrões nas relações espaciais dos jogadores, informação que pode ser útil para comissões técnicas de times de futebol. Mais próximo do contexto desse projeto temos o artigo [9], que busca encontrar e visualizar táticas no tênis de mesa utilizando-se da técnica de Descoberta de Subgrupos, encontrando padrões nos pontos bem sucedidos de um determinado jogador de interesse, mapeando as ações e reações dos jogadores durante um jogo. Além disso, temos o artigo [19], que tentaremos replicar nesse projeto, onde, através da utilização de Descoberta de Subgrupos sobre um conjunto de pontos e suas *features*, conseguiu encontrar padrões em pontos de serviço bem sucedidos.

III. METODOLOGIA

Iremos replicar o artigo [19], sendo assim a metodologia empregada nesse projeto será a mesma que a encontrada no artigo original: utilização de dados ponto-a-ponto, que serão especificados mais abaixo. Além disso, faremos também um processo de engenharia de *features* sobre esses pontos, com a intenção de criar um maior nível de detalhamento. Feito isso, podemos passar aos experimentos, onde tentaremos, através do uso da técnica de Descoberta de Subgrupos, encontrar subconjuntos de dados interessantes que caracterizem um jogador de interesse "X" sendo possível, dessa forma, fazer uma análise sobre as táticas utilizadas em pontos de saque bem sucedidos. Além disso, como já foi dito anteriormente, tentaremos também realizar essa mesma análise sobre pontos de recepção, como uma forma de levar adiante o que foi proposto no artigo original.

Os dados que serão utilizados provém do *Tennis Abstract Match Charting Project (TAMCP)*, um projeto coletivo que coleta dados detalhados de partidas de jogos profissionais de tênis. Ele possui dados de 14,416 jogos, 2,250,373 pontos e 8,617,097 rebatidas. Esse número considerável de dados, possibilita uma análise extensa de diversos pontos e, esperamos, com um nível de detalhes suficiente para demonstrar as diferentes táticas empregadas. Não trabalharemos com

todos os dados, optando por, assim como no artigo original, escolher um jogador de interesse e analisar uma parcela de jogos da carreira desse jogador.

Para entender as análises feitas abaixo, é interessante ter um conhecimento dos conceitos básicos do tênis. Vamos deixar explicitado alguns mais importantes para os nossos interesses. Porém, [12] contem uma visão geral sobre as regras e conceitos do tênis.

A. Definições

A estrutura de um jogo de tênis segue o formato: **Game - Set - Match**. Quem ganha 6 **Games**, ganha 1 **Set**. Ganhar 2 ou 3 **Sets** (dependendo do nível da competição), ganha o **Match** (jogo). Além disso, aqui estão algumas definições de conceitos base:

- **Serviço** - O golpe que inicia o ponto.
- **Forehand** - Golpe executado com a mão dominante voltada para a direção do golpe.
- **Backhand** - Golpe executado com as costas da mão dominante voltadas para a direção do golpe.
- **Erro Não Forçado** - Pontuação na qual um jogador, sem nenhuma influência de seu adversário, comete um erro no momento de tentar atravessar a bola para o lado do adversário.

Para os nossos propósitos, essas definições devem ser o suficiente

B. Seleção dos Dados

Para realizar as análises que intencionamos, precisamos selecionar os dados. Como em [19] não é mencionado o jogador de interesse, escolhemos utilizar Roger Federer como nosso jogador alvo. Utilizaremos dados de todos os jogos de Federer no ano de 2006, que foi, estatisticamente, o mais vitorioso de sua carreira. Assim, selecionamos 7722 pontos de 44 jogos oficiais, sendo 3686 desses pontos, pontos de serviço de Federer. Para esses pontos, temos os seguintes atributos iniciais:

- O placar do jogo, jogadores e quadra em que o jogo está sendo disputado
- O game e set que está sendo disputado
- Quem está no serviço
- Informações sobre se é o primeiro ou segundo serviço
- Codificação do Rally

Os 4 primeiros são informações gerais, porém o *rally* nos permite descobrir diversas outras informações a partir dele, como:

- Golpes presentes no ponto
- Como o ponto foi finalizado (*Winner*, erro forçado, *Ace*, etc.)
- Características sobre cada um dos golpes individualmente, como direção e tipo
- Número de golpes em um *rally*
- Informação sobre a direção do serviço

Essa lista não é exaustiva, mas deixa o entendimento do tipo de informação que pode ser retirado a partir desses

dados. Tendo todos esses atributos em mãos, podemos então passar para a próxima etapa. Para utilizar a tarefa que nos propusemos, esses atributos não são o suficiente, mas a partir deles criamos diferentes *features*, que serão utilizadas para realizar a descoberta de subgrupos.

C. Criação de Features

Dividimos as *features* criadas em diferentes grupos para facilitar o entendimento, utilizando os mesmos grupos descritos em [19]. A maioria das *features* são as mesmas do artigo original, porém como existem mudanças, é importante explicitar exatamente quais *features* foram utilizadas e o que elas denotam.

1) *Features do Ponto (10 features)*: Características gerais do ponto. O número total de golpes e a forma como o ponto foi finalizado. Além disso criamos 3 *features* binárias que separam os pontos entre pequeno (3 golpes), médio (4 a 7 golpes) e grande (7 golpes). Quando uma dessas *features* é 1, naturalmente as outras tem valor 0. Como no artigo original, queremos também saber como o ponto finaliza. Dessa forma fizemos 5 *features* com todas as situações possíveis: **Erro Forçado, Erro não forçado, Winner, Ace e Dupla Falta.**

2) *Features do Jogo (13 features)*: Nessa parte, consideramos a situação do jogo. Fizemos uma *feature* para determinar qual o número do game ou set está sendo disputado, denotados por um inteiro $i \in [1,5]$ para *Sets*, e $i \in [1,13]$ para os *Games*. Além disso, em relação aos *Sets*, é interessante saber se ele é (ou não) um *tiebreak*, portanto criamos uma *feature* binária para esse caso. Além disso, temos uma *feature* numérica que denota o tamanho da sequência de saque e outra para o retorno.

Além dessas características mais gerais, gostaríamos também de levar informações temporais. Dessa forma, criamos 3 *features* que determinam se o jogador está a frente no *Game*, no *Set* e no *Match*. Cada uma delas pode assumir os valores 1,0 e 1, que, respectivamente, sinalizam que o jogador está atrás, empatado ou à frente em cada um dos quesitos. Criamos, também, 5 *features* diferentes que nos dizem quantas pontos o jogador de interesse fez nos últimos 5, 4, 3, 2 e 1 pontos anteriores. Por exemplo, se nos últimos 3 pontos, o jogador venceu 2 deles, a *feature* dos últimos 3 pontos teria o valor inteiro "2". Essas últimas *features* descritas nos dão informação sobre o desempenho do jogador em um momento específico do jogo.

3) *Features dos Golpes (8 features)*: Temos 2 *features* que caracterizam o saque: uma nominal relacionada a direção (Aberto, No Corpo, ou na linha do T) e outra binária que identifica se a situação de "Saque e voleio" aconteceu. Fazemos, também, *features* relacionadas ao retorno, especificando o tipo de golpe e a direção do mesmo. Fazemos o mesmo para o último golpe realizado por cada jogador.

Aqui, vale mencionar o artigo [19]. Os pesquisadores utilizaram de características como "intencionalidade", que são impossíveis de replicar através da nossa base de dados, pois elas dependem de uma análise subjetiva visual da pessoa realizando a anotação. Portanto, temos um número bem menor

de *features* relacionadas diretamente com a característica dos golpes.

4) *Features do Rally (44 features)*: O rally é denotado pela sequência de golpes desde o saque até a finalização dos pontos. Portanto, as *features* do rally são relacionadas a toda essa sequência e seus componentes individuais. Assim, inicialmente criamos *features* binárias relacionadas a presença (ou não) de um determinado golpe. Fazemos também a distinção entre se esse golpe foi realizado pelo sacador ou pelo recebedor. Da mesma forma, fazemos essa mesma distinção para se uma direção está (ou não) presente durante um ponto. Como um exemplo, caso em um ponto disputado tenha uma sequência como:

[Saque aberto, *Forehand* Cruzado, *Forehand* Cruzado, *Backhand* Meio, Winner]

as *features* relacionadas a "*Forehand*" e a direção "Cruzado" estariam assinaladas com 1, enquanto para o recebedor as *features* de "*Forehand*" e "*Backhand*", assim como as direções "Cruzado" e "Meio" estariam assinaladas. Ao final, temos 44 *features* binárias diferentes.

Aqui, novamente, precisamos falar das decisões que foram tomadas em relação as *features* apresentadas no artigo original. Não é explicitado em [19] todos os diferentes tipos de golpes/direção contemplados. Sendo assim, foi necessário nos ater aos golpes que estão presentes na anotação manual realizada em nossa base de dados, dessa forma chegando ao número de *features* apresentado.

IV. EXPERIMENTOS

Buscamos encontrar padrões que representem padrões para pontos de serviço bem sucedidos de Roger Federer. Além disso, queremos encontrar outros padrões como esses, porém em referência aos adversários de Federer durante esses mesmos jogos. Também, vamos comparar os resultados encontrados aqui com os reportados em [19].

A. Método

Vamos utilizar a técnica de Descoberta de Subgrupos[17]. Como estamos tentando utilizar essa técnica para caracterizar os pontos de um determinado jogador, nos encontramos num problema de classificação: 1 para pontos vencidos, e 0 para pontos perdidos. Sendo assim, utilizaremos, assim como em [19] a ferramenta Cortana [18], e a medida de qualidade introduzida em [16]. A profundidade de busca de cada experimento segue os mesmos parâmetros delimitados no artigo original[19].

Para saber se os resultados são estatisticamente significantes, vamos recorrer ao artigo [8], também citado no artigo original. Nele fica demonstrado que o *threshold* para essa específica métrica de qualidade, com um $\alpha = 0.05$ é próximo de 0.06. Como veremos, os subgrupos listados superam esse *threshold* com certa folga e, portanto, são, também, estatisticamente significativos.

TABLE I
TABELA DE SERVIÇO DE FEDERER. REFERENTE A 3686 PONTOS, DOS
QUAIS 2544 (69%) FORAM VENCIDOS POR FEDERER

#	Coverage	Winning percentage	Quality	Condition
1	1517	91.8%	0.440	Tamanho do ponto é no máximo 2 golpes
2	2597	77.6%	0.284	Federer não comete um erro não forçado
3	3131	75.2%	0.248	Federer ganha o Game no qual o ponto foi disputado
4	2780	75.3%	0.224	Federer não utiliza Backhand
5	2306	76.4%	0.216	Federer disputa o ponto no 1o Serviço

B. Results

A Tabela I mostra os melhores subgrupos com profundidade de busca 1 para os pontos bem sucedidos de Roger Federer.

Como podemos ver, a característica que melhor garante sucesso é que o ponto seja curto, com no máximo dois golpes. De forma interessante, temos uma característica que se refere a cada uma das 4 divisões de *features* que realizamos, mostrando que o impacto encontrado nos pontos bem sucedidos de serviço não são específicos de nenhum domínio, mas sim de uma junção de vários aspectos do jogo. A imagem descrita por essas características, é uma em que Federer é capaz de dominar seus oponentes, sacando bem, acabando com os pontos de forma rápida, sempre usando sua mão dominante e não cometendo erros. Por outro lado, podemos olhar agora para os pontos de serviço bem sucedidos para os adversários de Federer durante esses mesmos jogos. Os resultados estão na Tabela II.

TABLE II
TABELA DE SERVIÇO DOS ADVERSÁRIOS DE FEDERER. REFERENTE A
4036 PONTOS, DOS QUAIS 2333 (69%) FORAM VENCIDOS PELOS
ADVERSÁRIOS (57%)

#	Coverage	Winning percentage	Quality	Condition
1	1265	84.6%	0.345	Tamanho do ponto é no máximo 2 golpes
2	2597	70.8%	0.345	Adversário ganha o Game no qual o ponto foi disputado
3	2572	65.6%	0.205	Federer não utiliza o Backhand
4	1369	71.2%	0.186	Adversário força o erro de Federer
5	20476	65.5%	0.160	Federer não utiliza o Forehand

Como podemos ver, o tamanho máximo do golpe e a característica de "ganhar o game disputado" se repetem, demonstrando a importância tanto de acabar um ponto de forma rápida, mas também demonstrando como os games de

serviço são geralmente vencidos pelos sacadores. No entanto, as outras 3 características são bem diferentes. Na verdade, o que elas mostram é que a possibilidade para o adversário vencer um ponto contra Federer naquele momento de sua carreira passava pela necessidade de encerrar o ponto de forma rápida, forçando um erro e impedindo que ele retornasse o serviço, incapacitando Federer de usar ou seu *Forehand* ou seu *Backhand*. Se expandirmos nossa busca para a profundidade 2, encontraremos para ambos, Federer e seus adversários, que os subgrupos com a maior valor de qualidade para a métrica são aqueles pontos que tem 2 ou menos golpes e Dupla Falta = 0, com valores de 0.491 e 0.411 respectivamente.

Sabendo dessas características, podemos também fazer análises em cima de situações mais específicas como, por exemplo, para pontos de serviço de tamanho menor que 3. Quando fazemos isso, encontramos que o subgrupo de maior qualidade, na profundidade 1, é quando Federer não comete um erro não forçado, com o valor da métrica sendo 0.373. Para os adversários, a condição é que eles forcem um erro de Federer, com valor da métrica 0.400. Se procurarmos na profundidade 2, tanto para Federer quanto para seus adversários, o subgrupo com maior valor na métrica encontrado é aquele onde o ponto é disputado em primeiro serviço e onde o jogador não comete um erro não forçado, com qualidade de 0.59 e 0.54 respectivamente.

Todas essas análises apontam as táticas utilizadas pelos jogadores. E, como vimos quando limitamos os pontos àqueles de tamanho menor que 3, ao mudarmos a situação de jogo, podemos encontrar táticas diferentes, o que mostra que esse método se adapta a condição do jogo e é capaz de extrair táticas específicas para diferentes cenários e para diferentes jogadores.

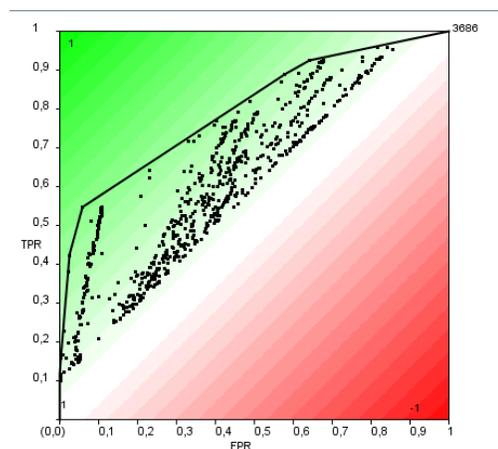


Fig. 1. ROC que ilustra a qualidade de classificação para pontos de serviço bem sucedidos de Federer. Cada ponto é um subgrupo obtido através de Descoberta de Subgrupos com busca de profundidade 2. A AUC é 0.8

Como dissemos anteriormente, estamos num cenário de classificação. Portanto, para checar a qualidade do nosso modelo, iremos utilizar a *Receiving Operation Characteristics (ROC) curve*. Em [21], citado também pelo artigo original,

fica demonstrado que a *Area Under the Curve* (AUC) de uma ROC quantifica o desempenho da classificação. Podemos ver a ROC na **Figura 1**. Com profundidade de busca 2, ao calcularmos a AUC, chegamos a um valor de 0.8 para os pontos de Federer e 0.76 para os pontos dos adversários. Portanto, a classificação é razoavelmente boa.

C. Comparações

Quando comparamos nossos resultados com os encontrados em [19], podemos perceber algumas informações interessantes. Primeiramente, para ambas as análises, a condição que tem maior valor de qualidade são pontos com no máximo 2 golpes. Isso mostra como essa característica é fundamental, independentemente de qual jogador estamos analisando. Finalizar pontos de forma rápida é a maneira que melhor garante o sucesso do ponto de serviço. No entanto, as outras características são bem diferentes. Para nosso jogador de interesse, as condições são variadas, com características relacionadas ao ponto, a situação de jogo, a características dos golpes e a características de *rally* se apresentando como importantes para decidir se o ponto será bem sucedido. No artigo, no entanto, encontramos duas características de "intencionalidade" nos golpes, que não pudemos reproduzir. No que se refere aos pontos do adversário, nossa análise mostrou a necessidade fundamental de encerrar os pontos de forma rápida e dominante, se utilizando do primeiro serviço. No caso do artigo original, no entanto, não houve uma grande diferença encontrada entre as condições para pontos de serviço bem sucedidos do jogador de interesse e de seus adversários.

Em relação ao desempenho da classificação, o valor de AUC que encontramos, 0.8 e 0.76, é parecido com os valores do artigo original 0.79 e 0.74. Nossos valores são ligeiramente superiores, mas o importante é perceber que ambos denotam uma qualidade de classificação razoavelmente boa.

V. CONCLUSÃO

Nesse projeto, nós replicamos os experimentos realizados em [19]. Através da Descoberta de Subgrupos, encontramos, para Roger Federer e para seus adversários, *features* que caracterizam pontos de serviço bem sucedidos. Encontramos que os pontos bem sucedidos de Federer são caracterizados por pontos curtos, onde utiliza sua mão dominante e no qual ele não comete erros não forçados. Igualmente, para os adversários, encontramos que o sucesso de um ponto de serviço passa pela realização de um primeiro saque bom e dominante, que termina o ponto de forma rápida, sem deixar que Federer responda. Fizemos isso através da replicação da metodologia do artigo original, obtendo resultados que são estatisticamente significativos e comparáveis aos resultados originais. A utilização dessa mesma metodologia em um grupo de dados diferente demonstra que ela parece ser válida para contextos diferentes do que os analisados na pesquisa original. Como mostramos através da AUC, encontramos uma qualidade de classificação razoavelmente boa, ligeiramente melhor que as encontradas em [19].

VI. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Sports analytics — evaluation of basketball players and team performance. *Information Systems*, 93:101562, 2020.
- [2] Natalia Andrienko, Gennady Andrienko, Alexandr Savinov, Hans Voss, and D. Wettschereck. Exploratory analysis of spatial data using interactive maps and data mining. *Cartography and Geographic Information Science*, 28:151–165, 07 2001.
- [3] Mark Carey. How data departments have evolved and spread across english football clubs. *New York Times - The Athletic*, September 3 2024.
- [4] Cristóbal J. Carmona, Pedro González, María José Del Jesus, Cristóbal Romero, and Sebastian Ventura. Evolutionary algorithms for subgroup discovery applied to e-learning data. pages 983 – 990, 05 2010.
- [5] Miguel Crespo and Machar Reid. Introduction to tennis tactics. *ITF Coaching and Sport Science Review*, 01 2002.
- [6] Arie-Willem de Leeuw, Rick van Baar, Arno Knobbe, and Stephan van der Zwaard. Modeling match performance in elite volleyball players: Importance of jump load and strength training characteristics. *Sensors*, 22(20), 2022.
- [7] Arie-Willem de Leeuw, Stephan van der Zwaard, Rick van Baar, and Arno Knobbe and. Personalized machine learning approach to injury monitoring in elite volleyball players. *European Journal of Sport Science*, 22(4):511–520, 2022. PMID: 33568023.
- [8] Wouter Duivesteyn and Arno Knobbe. Exploiting false discoveries – statistical validation of patterns and quality measures in subgroup discovery. pages 151–160, 12 2011.
- [9] Pierre Duluard, Xinqing Li, Marc Plantevit, Céline Robardet, and Romain Vuillemot. Discovering and visualizing tactics in a table tennis game based on subgroup discovery. In Ulf Brefeld, Jesse Davis, Jan Van Haaren, and Albrecht Zimmermann, editors, *Machine Learning and Data Mining for Sports Analytics*, pages 101–112, Cham, 2023. Springer Nature Switzerland.
- [10] Dragan Gamberger, Nada Lavrač, Filip Železný, and Jakub Tolar. Induction of comprehensible models for gene expression datasets by subgroup discovery methodology. *Journal of Biomedical Informatics*, 37(4):269–284, 2004. Biomedical Machine Learning.
- [11] Floris R. Goes, Michel S. Brink, Marije T. Elferink-Gemser, Matthias Kempe, and Koen A.P.M. Lemmink and. The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *Journal of Sports Sciences*, 39(5):523–532, 2021. PMID: 33106106.
- [12] International Tennis Federation (ITF). Itf rules of tennis, 2013. United States Tennis Association Website (USTA).
- [13] Claire Justin, Tom Griggs, and Richard Topping. The

- business of football: the big data arms race. *Financial Times*, September 27 2024.
- [14] Branko Kavšek and Nada Lavrac. Using subgroup discovery to analyze the uk traffic data. *Metodoloski Zv.*, 1:249–264, 01 2004.
- [15] Willi Klösigen and Michael May. Spatial subgroup mining integrated in an object-relational spatial database. In *European Conference on Principles of Data Mining and Knowledge Discovery*, 2002.
- [16] Arno Knobbe, Jac Orié, Nico Hofman, Benjamin van der Burgh, and Ricardo Cachucho. Sports analytics for professional speed skating. *Data Mining and Knowledge Discovery*, 31:1–31, 11 2017.
- [17] Petra Kralj Novak, Nada Lavrac, and Geoffrey Webb. Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, 10:377–403, 01 2009.
- [18] Arie-Willem Leeuw, Mathieu Heijboer, Mathijs Hofmijster, Stephan Van der Zwaard, and Arno Knobbe. Time series regression in professional road cycling. 10 2020.
- [19] Arie-Willem Leeuw, Aldo Hoekstra, Rens Meerhoff, and Arno Knobbe. Tactical analyses in professional tennis. 09 2019.
- [20] M. Lewis. *Moneyball: The Art Of Winning An Unfair Game*. Business book summary. WW Norton, 2003.
- [21] Simon Mason and N.E. Graham. Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, 128:2145 – 2166, 07 2002.
- [22] L. A. Meerhoff, F. R. Goes, A.-W. De Leeuw, and A. Knobbe. Exploring successful team tactics in soccer tracking data. In Peggy Cellier and Kurt Driessens, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 235–246, Cham, 2020. Springer International Publishing.
- [23] Rodolfo Metulini. Players movements and team shooting performance: a data mining approach for basketball, 2018.
- [24] Jornal Nacional. Informações precisas coletadas com base em tecnologia de ponta podem mudar os rumos dos jogos de tênis, 31/03/2025.
- [25] Robson Parmezan Bonidía, Jacques Duilio Brancher, and Rosangela Marques Busto. Data mining in sports: A systematic review. *IEEE Latin America Transactions*, 16(1):232–239, 2018.
- [26] Tatiana Sampaio, João P Oliveira, Daniel A Marinho, Henrique P Neiva, and Jorge E Morais. Applications of machine learning to optimize tennis performance: a systematic review. *Applied Sciences*, 14(13):5517, 2024.