Article

Unsupervised Learning in NBA Injury Recovery: Advanced Data Mining to Decode Recovery Durations and Economic Impacts

George Papageorgiou [ID], Vangelis Sarlis [ID] and Christos Tjortjis *[ID]

School of Science and Technology, International Hellenic University, 57001 Thessaloniki, Greece; gpapageorgiou2@ihu.edu.gr (G.P.); e.sarlis@ihu.edu.gr (V.S.)
* Correspondence: c.tjortjis@ihu.edu.gr

Abstract: This study utilized advanced data mining and machine learning to examine player injuries in the National Basketball Association (NBA) from 2000–01 to 2022–23. By analyzing a dataset of 2296 players, including sociodemographics, injury records, and financial data, this research investigated the relationships between injury types and player recovery durations, and their socioeconomic impacts. Our methodology involved data collection, engineering, and mining; the application of techniques such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN), isolation forest, and the Z score for anomaly detection; and the application of the Apriori algorithm for association rule mining. Anomaly detection revealed 189 anomalies (1.04% of cases), highlighting unusual recovery durations and factors influencing recovery beyond physical healing. Association rule mining indicated shorter recovery times for lower extremity injuries and a 95% confidence level for quick returns from "Rest" injuries, affirming the NBA's treatment and rest policies. Additionally, economic factors were observed, with players in lower salary brackets experiencing shorter recoveries, pointing to a financial influence on recovery decisions. This study offers critical insights into sports injuries and recovery, providing valuable information for sports professionals and league administrators. This study will impact player health management and team tactics, laying the groundwork for future research on long-term injury effects and technology integration in player health monitoring.

Keywords: anomaly detection; association rules; data mining (DM); sports analytics; text mining

1. Introduction

The applications of machine learning (ML) and data mining (DM) techniques have transformed the understanding of player injuries, performance, and recovery in professional sports, particularly in leagues such as the NBA. These techniques allow valuable insights to be extracted from vast datasets, linking injury patterns with sociodemographic and economic variables [1]. The combination of ML and DM methods creates a data-rich pathway for exploring the physiological and socioeconomic aspects of sports injuries and for forecasting or statistically analyzing athlete performance, providing a comprehensive view of athlete health and performance [2]. For instance, applications such as anomaly detection identify unusual patterns that may be influenced by factors beyond the physical activity, such as a player's sociodemographic information [3].

In addition to the direct implications of sports injuries for player performance, the application of association rule mining, mainly through algorithms such as the Apriori algorithm [4], sheds light on the intricate relationships between injuries, recovery time, and player demographics. By identifying these relationships, sports analysts and team management can create more personalized rehabilitation programs and estimate recovery timelines [5]. Ultimately, ML and DM techniques provide valuable tools for understanding the landscape of sports injuries, allowing stakeholders to make informed decisions that optimize player care and team performance.

A review article has explored the outcomes of sport injuries. The authors conducted a comprehensive review of the literature to analyze the impact of different types of sport injuries on athletes and their subsequent recovery [6]. This study investigated the return to play and performance of NBA players who underwent reconstruction. This was a case series and a literature review. The authors found that there was a high rate of return to play and good performance outcomes following reconstruction in the NBA [7].

Moreover, a study examined the attitudes, priorities, and perceptions of exercise-based activities in youths after a sport-related injury. A sample of youths who sustained a sport-related injury was followed up 12–24 months after the injury. They found that although the youths experienced a decrease in physical activity levels after the injury, they were able to reframe their priorities and attitudes toward exercise-based activities and were motivated to overcome the physical limitations imposed by the injury. This highlights the importance of positive attitudes and priorities toward exercise-based activities in promoting recovery and maintaining physical function after a sport-related injury [8].

An additional study aimed to investigate the potential of using ML methods to predict anterior cruciate ligament (ACL) injuries in female elite handball and soccer players. The study used three-dimensional motion analysis and physical data collected from 791 female athletes and revealed that the best classifier had a mean Area Under the Receiver Operating Characteristic (AUC-ROC) curve of 0.63, which was significantly better than chance, indicating that there is information in the data that may be valuable for understanding injury causation. However, the predictive ability of the NRS-2002 score was low from a clinical assessment perspective, suggesting that the included variables cannot be used for ACL prediction in practice. The study also revealed that class imbalance handling techniques did not improve the results. Therefore, while the study highlights the potential of ML in the field, it also highlights the limitations of current models in making accurate predictions in practice [9].

Based on the results of these assessments, specific interventions can be implemented to reduce the risk of injury and improve overall performance. For example, targeted strength training and proper technique instruction can help to improve knee stability and reduce the risk of injury. Additionally, incorporating stretching and flexibility exercises into an athlete's routine can help to improve overall flexibility and reduce the likelihood of injury. The identification and assessment of risk factors for ACL injuries is an important step in reducing the risk of injury in athletes. By taking a proactive approach and addressing risk factors, athletes can not only reduce their risk of injury but also improve their overall performance [10–16].

A similar study aimed to examine the effect of ACL reconstruction on the athletic performance and career longevity of NBA players. The study involved a sample of NBA players who underwent ACL reconstruction, and their performance data were analyzed before and after surgery. The results showed that although players experienced a decline in performance after surgery, they were able to return to the pitch and had a prolonged career. The study also revealed that the decline in performance was temporary and that many players were able to recover their preinjury level of performance. The findings of the study suggest that ACL reconstruction does not significantly impact the career longevity or athletic performance of NBA players [17].

An additional investigation based on the impact of injuries on postretirement pain and quality of life in professional basketball players was performed. The study was a cross-sectional survey, and the results showed that injuries can have a negative impact on postretirement pain and quality of life for professional basketball players [18]. Researchers have investigated the return-to-sport ratio and performance after reconstruction in NBA players. One study showed that NBA players were able to return to play after reconstruction; however, their performance was impacted [19]. Another research study investigated sex-specific differences in injury types among basketball players. One study revealed differences in the types of injuries sustained by male and female basketball players [20].

Another research paper has analyzed previous studies to determine the best strategies for preventing lower extremity injuries in basketball. The authors conducted a systematic review and meta-analysis of existing related research and found that injury prevention programs that focus on improving strength, balance, and landing technique can be effective at reducing the risk of injury [21]. Another study investigated the types of injuries commonly sustained by basketball players and the strategies that can be used to prevent and manage these injuries [22].

Further research has revealed that players who have sustained a knee injury were more likely to sustain additional knee injuries, such as meniscus tears or cartilage injuries, on the same or opposite knee. Additionally, players who have sustained a knee injury may have a higher risk of developing osteoarthritis, a degenerative joint disease that can lead to pain and disability. It is worth noting that the risk of reinjury after an injury is limited not only to knee injuries but also to other types of injuries, such as ankle sprains, muscle strains, and even concussions. The risk of reinjury after an injury can be due to several factors, such as the quality of the surgical reconstruction, the rehabilitation process, the player's adherence to the rehabilitation protocol, the player's return to sport, and the player's physical readiness. In conclusion, there is some evidence suggesting that players who sustain an injury may be at increased risk of sustaining additional knee injuries and other types of injuries. It is important for players who have sustained an injury to undergo appropriate rehabilitation and to take the necessary precautions to prevent reinjury [23–25].

1.1. Related Work

The integration of ML and DM methods has significantly contributed to the field of sports science in understanding sports injuries, recovery processes, and performance optimization. Several studies have employed these methodologies to explore different aspects of sports injuries.

The research in [26] examined the application of ML for predicting muscle strain injuries in NBA athletes from 1999 to 2019. This study compares traditional logistic regression methods with ML models, such as random forest and extreme gradient boosting (XGBoost), to evaluate injury risk. The purpose of the study was to improve the understanding of risk factors and contribute to better injury prevention strategies in professional basketball. Based on their findings, the XGBoost machine achieved the best performance compared to logistic regression in predicting lower extremity muscle strain.

In another paper, the authors aimed to predict the risk of injury in elite athletes using ML algorithms. The authors used data from an extensive screening test battery of 880 female athletes to develop an ML model that could predict the risk of injury. The results showed that the model was able to predict injury risk with high accuracy. The authors concluded that their findings may have important implications for the prevention of injuries in female athletes [27].

A different research of injury forecasting in NBA basketball used a deep learning approach. The authors analyzed data from players in the NBA and used ML algorithms to make predictions about the likelihood of injury. One study revealed that deep learning can be an effective method for forecasting injuries in professional basketball. The authors developed a deep learning model, called the Multiple bidirectional Encoder Transformers for Injury Classification (METIC) model, to predict future injuries in NBA players. Longitudinal data on NBA player injuries were collected using publicly available data sources, and a model that performed significantly better than other traditional ML approaches was developed. The METIC model uses feature learning to create interactive features that become meaningful when combined with each other. The study suggested that the model can be used by practitioners and staff to improve athlete management and reduce injury incidence, potentially saving millions of sports teams' revenue due to reduced athlete injuries. The study addresses the issue of small sample sizes and imbalanced data, which have been a challenge in sports medicine when predicting athlete injury risk. The study's results

indicate that this model can be a valuable tool for sports teams to predict and prevent injuries in their athletes [16].

However, the above-mentioned studies have yet to investigate the connections between recovery times, sociodemographic, injuries, and financial information spanning over two decades in the NBA directly. This novel approach utilizes advanced machine learning and data mining techniques, such as anomaly detection and association rule mining, to analyze two decades of data. This new approach offers insights into the relationships between recovery times, sociodemographic factors, and financial aspects of sports injuries, marking significant justifications in the field.

1.2. Research Overview

This research paper tackles an extensive examination of player injuries in the NBA using advanced DM techniques and ML methodologies. This study aimed to unravel the complex patterns of injury frequencies and recovery patterns and their wider implications for team dynamics and league operations. The research covers the period ranging from 2000–2001 to the 2022–2023 NBA seasons. The dataset included 2296 NBA players and included their performance metrics, injury records, and financial information. The primary focus of our study is to address two key research questions. First, we aim to utilize unsupervised learning techniques such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect unusual recovery times for different kinds of injuries. Second, we aim to investigate the relationships between variables such as injury type, recovery period, sociodemographic factors, and financial implications for both teams and the League.

Our methodology follows a three-dimensional approach consisting of data collection, engineering, and mining. We initiated the process by conducting a thorough data collection and preprocessing phase, which involved the use of advanced techniques, such as web scraping, data normalization, and feature selection [28]. Subsequently, we utilized sophisticated DM methods, including anomaly detection through DBSCAN [29], isolation forests [30], Z score techniques [31], and association rule mining (ARM), using the Apriori algorithm [32]. This study introduces the innovative use of advanced DM techniques employing a synergistic blend of anomaly detection methods—DBSCAN, isolation forest, and Z scores—complemented by ARM through the Apriori algorithm, enabling the identification of significant patterns, anomalies, and associations within the NBA injury data.

In the realm of anomaly detection, we employed a novel approach by integrating multiple algorithms to accurately identify unusual recovery time patterns in professional basketball. This enabled us to gain a better understanding of the various factors that influence recovery times. We used the ARM [33] to examine the relationships between injury type and recovery duration and between injury type and sociodemographic and financial implications.

This study provides valuable insights into sports injuries, rehabilitation strategies, and economic factors related to player health in the context of professional basketball. The research findings can guide sports scientists, medical professionals, team managers, and league policymakers in making informed decisions and shaping future policies [34]. Furthermore, the study's results pave the way for future research into the long-term impact of injuries on players' careers and the incorporation of innovative technologies in managing player health.

2. Data and Methods

In this study, we aimed to understand the complex dynamics of player injuries in the NBA and their various impacts. We used advanced DM techniques and a comprehensive analytical approach to unravel these complexities. Our extensive dataset, covering player sociodemographic, injury, and financial data from the 2000–2001 to the 2022–2023 NBA seasons, forms the basis of our research.

2.1. Research Questions/Hypothesis

1. Can we identify anomalous recovery times for various types of body injuries in the NBA using unsupervised learning methods? (DBSCAN)
2. What patterns and associations exist between types of injuries, recovery durations, sociodemographics, and their impact on team financial outcomes in the NBA?

2.2. Methodology

Our research utilized a diverse range of data sources [35,36] to conduct a robust and comprehensive analysis. The process of retrieving and preprocessing the data presented several challenges, which we addressed by consolidating these varied data into a supervised data model, with a strong focus on data quality. Our research methodology comprises three main dimensions: Data Collection, Data Engineering, and DM methodologies. The Data Collection phase involved meticulous gathering [37] and web scraping, including a comprehensive set of NBA player sociodemographic, injury, and salary information spanning the seasons from 2000–2001 to 2022–2023. Data engineering involved data cleansing, structuring, and enhancing the datasets, including data normalization and feature engineering and selection, aimed at refining and standardizing the data for analysis [38]. In the DM phase, we utilized sophisticated methods, such as anomaly detection via DBSCAN [29], isolation forests [30], Z scores [31], and ARM via the Apriori algorithm [32]. These techniques were instrumental in identifying significant patterns, anomalies, and associations within the data, ultimately leading to insightful conclusions about player injuries in the NBA.

2.2.1. Data Collection

The structure of the collected data was unorganized and heterogeneous, which necessitated the implementation of a thorough methodology for comprehending the information and obtaining valuable insights. This methodology included several steps, such as data collection, preprocessing, analysis, and result evaluation. The data were collected through Python scripts [37] and underwent preprocessing procedures to identify and eliminate missing values, outliers, and irrelevant data. Furthermore, the data were transformed through scaling and normalization and structured into a usable format, and only the most pertinent features were selected. To enhance the quality of the data, an extract, transform, and load (ETL) process was employed to standardize and homogenize the data.

This research paper presents a comprehensive analysis of player sociodemographic, injury, and salary data for all NBA players from 2000–2001 to 2022–2023. The data were collected from various sources, and this subsection provides details about the data acquisition process, the type of data, and the corresponding dataset shapes.

To collect player sociodemographic and injury data, we utilized the nba_api [35], a robust API that pulls data directly from the NBA's official website and database. The data were collected over a specific time span of 22 seasons, starting from 2000–2001 and continuing through 2022–2023. Our comprehensive dataset included 2296 players, representing the entire pool of players who participated in the NBA during the selected seasons.

Two separate scrape runs were performed on nba_api to extract data on NBA players. The first scrape targeted the regular season per-game records, while the second scrape focused on playoff records. These datasets contained comprehensive statistics on all players who participated in the NBA within the chosen time frame. The datasets were merged using identical player_id and game_date fields, creating a complete dataset that was stored locally in a PostgreSQL database. The datasets also included comment-based metadata, which provided reasons for player absence, including coaching decisions and injuries.

In addition, another dataset was acquired through web scraping from the ESPN [36] open website, which centralized information on players' signed contracts. This contract dataset also spans from the 2000–2001 season to the 2022–2023 season, ensuring consistency across the different datasets. Overall, these datasets provide a comprehensive and detailed perspective on NBA player sociodemographics and absences, which can be used to gain

insights into player behavior and team performance. A summary of the schemas for these datasets is provided in Table 1.

Table 1. Research datasets on demographics, injuries, and contrasts in the seasons from 2000 to 2023.

Name (Type)	Shape (Rows, Columns)
Player sociodemographic information	(781,406, 26)
Injuries data (on and off game)	(58,151, 4)
Contracts data (signed over seasons)	(7257, 6)

2.2.2. Data Engineering

After collecting the raw data, the next critical phase was data engineering, involving data cleansing, structuring, and enhancing the datasets. This phase was crucial in preparing the injury and contract datasets for proceeding with anomaly detection and ARM. This process was important—especially for heavy text datasets, such as injury reports and contract details—to ensure accuracy. Structuring the data made it easier to organize and access the information for text-mining purposes. Additionally, transforming contract details into a comprehensive salary dataset allowed us to apply advanced DM techniques later.

Text Mining and Categorization in Injury Data

For the injury dataset, text-mining techniques [39,40] were applied to extract relevant injury information from textual descriptions of players' injuries. A customized dictionary was developed to classify injuries into predefined categories. For instance, an injury description such as "placed on IL with bone spur in left heel" was mapped to the "Heel" → "Foot" category. Moreover, the dataset contained multiple duplicates due to the absence of players missing more than one game due to injury. To address this, duplicate cases were identified, and only the first instance along with the date of the initial injury was retained under specific conditions. After the textual descriptions were mapped [41,42], records were flagged as "duplicate" (TRUE) if they referred to the same injury type for the same player within a 15-day window from the last reported occurrence of that particular injury type. This approach helped to reduce the dataset's size while retaining the relevant information. The study results were obtained using text mining and categorization techniques, which are essential for injury data analysis in sports.

To ensure the integrity and quality of our injury data, every stage of the data engineering process was monitored and reviewed. Firstly, we supervised the results of the removal of the common and unnecessary words from the dataset, which made it more organized and formatted for categorization. Secondly, we carefully checked the accuracy of the categorization process to ensure that no data were missing, and all the information was properly classified. Finally, we analyzed the word count in each category and checked each record for duplicated injuries, to confirm that only relevant data were included and to increase the precision of our analysis.

Transformation of Contract Data into Salary Data

The objective was to transform contract data into a more analytical salary dataset. To achieve this goal, we applied text-mining techniques to extract contract lengths and amounts from the textual descriptions contained in the scraped data. For instance, a contract description such as "signed as a free agent (from Lakers) to a 2-year, $22 M contract" was divided into a length of 2 years and an amount of USD 22 million.

Moreover, inflation rate [43] data for the relevant years were used to standardize salary figures, enabling more meaningful comparisons of player salaries while considering the year the contract was signed and the economic context of the U.S., where the NBA operates. Formula (1) was used to perform these calculations.

To confirm the quality of the contract data, we initially focused on the cleansing process, removing stop words and common irrelevant text. This step was crucial in refining

the contract data, ensuring it was free from irrelevant or duplicative information that could impact the quality and reliability of our analysis. Furthermore, after categorizing the contract years and amounts, we cross-checked to ensure that every player who had participated in the league had an assigned contract amount for each of their years. Finally, we checked for significant deviations in each player's salary amount over the years to identify potential errors and ensure the data quality for proceeding further.

2.2.3. Anomaly Detection Methodology

This study utilized a combination of algorithms—namely, DBSCAN, an isolation forest, and the Z score—to detect unusual patterns in the recovery time data of NBA players. By leveraging each algorithm's unique capabilities, we aimed to improve the accuracy and reliability of the detection process.

DBSCAN Algorithm Application

The present study employed the DBSCAN algorithm [29,44,45], a density-based clustering approach, to identify the clusters and noise present within the data. DBSCAN was selected because of its capability to accurately detect clusters within complex datasets, which is essential for comprehending injury duration patterns due to the data's characteristics. It is an excellent option for identifying anomalies in large datasets, allowing unusual recovery periods to be distinguished from typical ones.

To enhance the sensitivity of the algorithm to the density of data points, the radius of the neighborhood (ϵ) was reduced by a factor of 0.5 to allow for more granular detection of clusters. This approach enabled the algorithm to distinguish between more compact clusters and potential outliers. Additionally, the algorithm's selectivity in recognizing core points was improved by doubling the minimum number of points needed to form a dense region (min_samples). The optimal (ϵ) was calculated mathematically through the nearest neighbors technique, which involved constructing a k-distance graph and identifying the point of maximum curvature (knee) as the indicator of the optimal (ϵ) value; thus:

$$\epsilon_{\text{optimal}} = \text{k-distance(knee point)} \tag{1}$$

where the knee point was found by the maximum gradient change in the sorted k-distance plot.

Isolation Forest Algorithm

The isolation forest technique [30,46] is a powerful tool for detecting outliers in datasets with high dimensions. Isolation Forest was selected for its efficiency in detecting anomalies in datasets with high deviation, which is particularly useful for identifying atypical recovery durations in the extensive NBA data. Its effectiveness in handling complex injury records makes it a valuable option for this analysis.

In this study, we set the contamination rate to 0.01, which is believed to represent the percentage of anomalies in the overall population, providing a balance between sensitivity to outliers and retaining the integrity of the dataset's overall pattern. Using an unsupervised learning approach, the algorithm isolates observations by randomly choosing a feature and then selecting a split value between the minimum and maximum values of that feature. As a result, anomalies are expected to have shorter path lengths on the tree and are thus isolated closer to the root.

Detection of Statistical Anomalies via the Z Score

The present study incorporates a statistical approach through Z score analysis to detect outliers within the dataset [31,47]. The Z score technique is utilized for its straightforward approach to identifying outliers by estimating the deviation from the mean, providing a clear metric and a basis for anomaly detection. Its simplicity and efficiency are vital in quantifying the extent of deviation in recovery times within the large dataset. A criterion for identifying anomalies was employed, set at a Z score greater than 3. In the context of a

normal distribution, this criterion accounts for approximately 99.7% of the data, identifying those data points in the upper 0.3% tail as outliers, minimizing the risk of false positives, and focusing on the most significant anomalies.

Ensemble Anomaly Detection Strategy

A consensus strategy was developed where an observation is classified as an anomaly if it is identified by at least two out of three methods. This ensemble approach is represented as follows:

$$\text{Anomaly Indicator} = 1 \text{ if } \sum_{j=1}^{3} M_j(i) \geq 2, \text{ otherwise } 0 \qquad (2)$$

In Equation (2), (Anomaly Indicator) is the indicator for the (ith) observation, and $(M_j(i))$ is the detection result from the (jth) method for the same observation. This approach minimizes the likelihood of false positives, ensuring a more robust identification of true anomalies.

The algorithms used in this approach were carefully tuned to the dataset's intrinsic characteristics and were applied iteratively to distinct groupings based on "Detailed Body Part". By combining the strengths of each algorithm—DBSCAN for local density, isolation forest for global anomaly separation, and Z score for statistical deviation—the ensemble framework yielded a refined dataset that was rigorously screened for outliers. This approach ensured the integrity of the subsequent data analysis stages and the association rules that were derived from the data.

2.2.4. Methodology for Association Rules

ARM is a key technique in DM that aims to discover interesting relationships, frequent patterns, associations, or causal structures among sets of items in transaction databases or other data repositories [48]. The primary objective is to identify rules that occur frequently in a dataset. The Apriori algorithm was chosen for our study because it efficiently uncovers frequent patterns and associations within large datasets. It is particularly relevant for our research, as it enables the detailed analysis of convoluted relationships between various aspects of sports injuries, such as types of injuries, recovery times, and sociodemographic and financial variables in the NBA dataset.

Theoretical Basis

In our research, we utilized the Apriori algorithm, a widely recognized algorithm in DM, to extract frequent itemsets and association rules. Introduced by Agrawal (1994) [49], this algorithm is a fundamental approach in DM for discovering frequent itemsets and generating association rules. The Apriori [32] property is a key feature of this algorithm, which states that all nonempty subsets of a frequent itemset must also be frequent.

The Apriori algorithm [50] starts by examining the dataset to establish support for each item and identifying frequent single-item sets (1-itemsets). In the candidate generation phase, these itemsets are extended to (k)-itemsets by combining frequent $(k-1)$-itemsets. During pruning, it discards (k)-itemsets containing any infrequent $(k-1)$-subset, in line with the Apriori principle that all subsets of a frequent itemset must also be frequent. The algorithm then estimates the support of each candidate (k)-itemset, retaining those above a minimum support threshold. This repetitive process continues until no new frequent itemsets are found, at which point the algorithm generates association rules from these identified frequent itemsets.

The theoretical foundation of association rules lies in the concepts of support, confidence, and lift. These metrics evaluate the strength and significance of the discovered rules:

- Support [51] indicates the frequency or prevalence of an item set in the dataset.

○ The support of an itemset (A) is defined as the proportion of transactions in the dataset that contain the itemset. Mathematically, this process is expressed as follows:

$$\text{Support}(A) = \frac{\text{Number of transactions containing } A}{\text{Total number of transactions}}$$

- Confidence [52] measures the likelihood of occurrence of the consequent in a transaction given the presence of the antecedent.

○ The confidence of a rule $(A \Rightarrow B)$ (where (A) and (B) are disjoint itemsets) measures the likelihood of (B) being present in transactions that contain (A). It is calculated as follows:

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

- Lift [53] assesses the strength of a rule over the random occurrence of the antecedent and consequent, indicating the rule's effectiveness in predicting the consequent.

○ Lift evaluates the performance of a rule compared to the expected performance if (A) and (B) are independent. The lift of a rule $(A \Rightarrow B)$ is given by:

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Confidence}(A \Rightarrow B)}{\text{Support}(B)}$$

Implementation Process

The methodology employed in our study involved a series of important steps for the application of association rules. First, we prepared the NBA player data by converting the continuous variables into discrete groups based on predetermined criteria. The "Salary Per Game" was divided into five categories: "0–150 k", "150–300 k", "300–450 k", "450–600 k", and "600 k+", to reflect different financial tiers in NBA player contracts, capturing the diversity in player earnings. For "Team Losses", we created categories such as "0–25 M", "25–50 M", "50–75 M", "75–100 M", and "100 M+", segmenting the data based on the financial impact of player injuries on teams. Player age was grouped into six ranges: "0–20", "20–25", "25–30", "30–35", "35–40", and "40+", allowing an in-depth analysis of the influence of age on injury occurrence and recovery. Height in centimeters and weight in kilograms were similarly categorized: "<175", "175–200", "200–225", "225–250", "250+" and "<70", "70–100", "100–130", "30–160", "60+", respectively—to investigate correlations with injury patterns and recovery times. Lastly, "Recovery Time" in days was segmented into "0–10", "10–30", "30–90", "90–180", and "180+", with an additional category "career ending" specifically for career-ending injuries, aiding in understanding the duration and severity of injuries. These groupings were crucial in dissecting the intricate relationships between various player attributes and their professional experiences in the NBA.

Second, we encoded the data using one-hot encoding to transform the dataset into a binary matrix representation [54]. Following the categorization of our mentioned sociodemographic and finance features, the encoding process was applied to the injury types, most common position, salary per game, team losses, age group, height, weight, and recovery duration. One-hot encoding effectively converted these categorical variables into a binary format, where each unique category within a variable is represented by its column, marked with a "1"/True or "0"/False to indicate the presence or absence of that category for each player.

As a result of this encoding, the dimensionality of our dataset expanded considerably from the initial eight columns. Post encoding, our dataset comprised sixty-six columns, each representing the unique categories derived from our grouped variables and the detailed body parts identified from the injury data. This expanded dataset was thus properly structured, enabling a deeper exploration of patterns within the NBA player data.

The Apriori algorithm was subsequently applied to identify frequent itemsets within the dataset. From these frequent itemsets, we generated association rules by calculating the confidence and lift for each rule. Finally, we analyzed the extracted rules to draw meaningful insights, focusing on the implications in the context of NBA player injuries, recovery times, and their financial impacts on teams.

Significance in Research

Our research study utilized association rules to uncover hidden patterns and relationships in complex datasets of NBA player sociodemographic information [55], injury information [56], and financial information [57]. This methodological approach offered a quantitative means to address our research questions, specifically about the financial effects of player injuries and recovery durations on NBA teams. Using a comprehensive and systematic application of association rules, our research contributes to the broader understanding of the interplay between player health, sociodemographic characteristics, and team economics in professional sports, focusing on the NBA.

3. Results

In the Results section, we present a detailed analysis of NBA player injuries and their recovery times. We examine and identify some unusual patterns in recovery duration, as well as the influence of various factors, such as player position, injury type, and economic impact. We first highlight the differences in recovery times among players and their positions. Then, we explored the associations between injury types and recovery times and team financial outcomes, providing a complete understanding of the injury associations in professional basketball.

3.1. Anomaly Detection during Recovery

We conducted a comprehensive analysis of NBA player injuries and the time required for them to recover. Our study examined a dataset of 18,184 entries, and we identified 189 anomalies, which represented approximately 1.04% of the cases. These anomalies were noted for their unusual recovery durations, which were significantly different from the established patterns observed in the majority of the dataset. It is important to highlight that these anomalies primarily involved cases of players who left the NBA due to injuries and returned to the league after a prolonged absence, resulting in unusually long recovery times. Our investigation included various player variables, such as their position, age, body parts affected, and other relevant metrics.

Our investigation into the abnormal recovery times of players with injuries revealed that there is a wide distribution across different player positions. Notably, no specific position has a higher concentration of unusual recovery times, indicating that this phenomenon can affect players in any position. An interesting case was observed in the "Center" position, where a player with an "Ankle" injury had an extraordinarily long recovery time of 711 days, which was accompanied by a high salary per game and significant team losses. This finding suggests that there is a complex relationship between player value for the team and the impact of injury. To provide a visual representation of these extended recovery times across various player positions, we have included Figure 1, which highlights the diverse nature of these anomalies. This violin plot visualizes the range and density of recovery times for each player position among the anomalies detected. This highlights the variability in recovery periods postinjury, irrespective of the player's position.
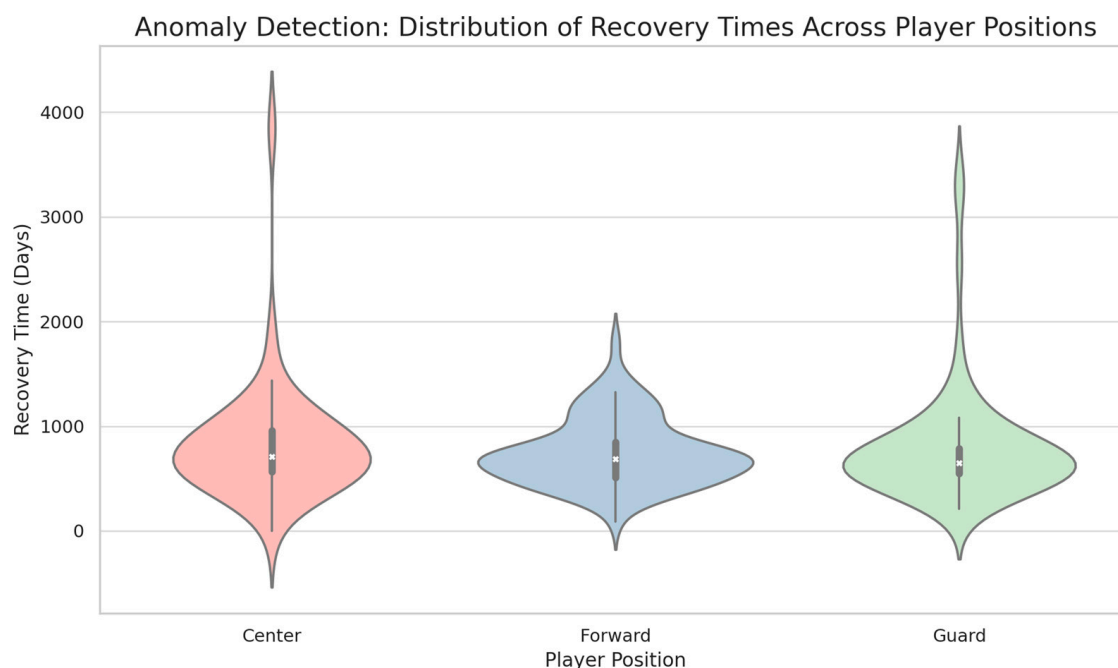
Figure 1. Distribution of recovery times among anomalies detected by player position.

Anomaly analysis revealed that atypical recovery times were diversely distributed across different player positions. No particular position showed a disproportionate concentration of extended recovery periods, indicating that any player position can be affected by this phenomenon. One instance of prolonged recovery was noted in the "Center" position, where a player with an "Ankle" injury took 711 days to recover due to a brief departure from and return to the NBA. This case also featured significant team losses and a high salary per game, suggesting a complex relationship between a player's value to the team and the impact of their injury. The violin plot in Figure 1 portrays a visual representation of these prolonged recovery times. Colors serve to differentiate between the distributions of recovery times for three player positions: Center (red/pink), Forward (blue), and Guard (green), highlighting the variation and density of recovery times for each position.

3.2. Association Rules in Recovery Times

The detailed development of the association rules from NBA data offers valuable insights into how injuries, recovery periods, and various factors, such as salary and physical attributes, correlate within the professional basketball landscape. Table 2 displays the correlation between various types of injuries in NBA players and their short-term recovery (0–10 days), along with the financial impacts on teams (team losses of USD 0–25 million). The metrics included support, confidence, and lift, highlighting the recovery efficiency for different injuries.

The core of our investigation, as presented in Table 2 (in descending order on the "Lift" variable), focused on the correlation between specific bodily injuries and the predicted duration of recovery. Our analysis reveals that players who are absent due to "Rest"—a term encompassing a range of health measures—exhibit a remarkable 94.75% level of confidence in returning to play within a 0–10-day window. This figure is significantly higher than the standard, as indicated by a lift of 1.71. Our empirical analysis provides quantified evidence of this relationship within a large-scale, real-world dataset, confirming a common assumption with robust data. The significance of this finding goes beyond what might be presumed; it establishes a benchmark for expected recovery times associated with rest and facilitates comparison with other, more severe injuries, offering a more defined picture of their recovery revolution. Furthermore, it has important implications

for player health management policies, as it underscores the effectiveness of rest periods in recovery protocols in NBA. While the correlation between "Rest" and recovery time might be anticipated, its documentation within our study is crucial for an evidence-based approach to sports medicine and team management. These findings suggest that the preventive rest policies of the NBA are widely adopted and highly effective at maintaining player health.

Table 2. Correlations between injury types and short-term recovery and team financial outcomes.

Antecedents: Injury Type Consequents: Recovery (0–10 Days) and Team Losses (USD 0–25 million)	Support	Confidence	Lift
Rest	2%	95%	1.71
Respiratory	1%	79%	1.43
General illness	6%	77%	1.39
Neck	1%	70%	1.27
Abdominal	1%	66%	1.19
Hip	2%	64%	1.15
Back	5%	63%	1.14
Cranial	1%	63%	1.14
Unclassified	1%	62%	1.12
Ankle	8%	58%	1.05
Facial subareas	1%	57%	1.03
Heel	1%	54%	0.98
Arm	1%	54%	0.97
Groin	1%	53%	0.95
Thigh	4%	52%	0.94
Knee	7%	52%	0.93
Foot	3%	51%	0.92

Our findings also indicate that individuals with respiratory-related absences and general illnesses tend to recover rapidly, with confidence levels of 79.45% and 77.15%, respectively, and corresponding lifts of 1.43 and 1.39, respectively. These conditions are managed efficiently, enabling players to return to play quickly and minimize downtime.

Delving into the realm of injuries more directly associated with the physical rigor of the game, we see a similar trend. The findings revealed that players with lower extremity injuries, particularly those affecting the thigh, foot, or groin, exhibited shorter-than-expected recovery times than did those with a confidence level greater than 50%. These findings suggest that customized advanced treatment and rehabilitation protocols for these injuries contribute to faster recovery and reintegration.

Notably, the data revealed an association between knee injuries and both short and moderate recovery times. This finding indicates that while knee injuries are often serious, there is a subset of these injuries where players are able to recover and return to play in a shorter timeframe than traditionally anticipated.

Furthermore, injuries to the upper extremities and the head—such as those to the arm and cranial regions—show a broad spectrum of recovery times. This diversity in recovery reflects the varied impact of these injuries and underscores the personalized nature of treatment and recovery plans.

In addition to the physical impact of injuries on basketball players, the analysis revealed that recovery times are influenced by a range of factors. Players within the lower

salary bracket, notably those earning USD 0–150k, tend to have shorter recovery periods. This could be due to a combination of economic factors and the contractual obligations of these players, which may encourage a faster return to the pitch. Table 3 illustrates (in descending order on the "Lift" variable, per recovery in days group), the associations between NBA players' recovery periods and their physical attributes, such as height and weight, players' salaries, and the economic impacts on teams. It covers various recovery timeframes and shows how these aspects are interrelated, with support, confidence, and lift values for each category.

Table 3. Associations between recovery periods, player attributes, and team economic impacts.

Antecedents: Recovery in Days	Consequents: Height (H) in cm Weight (W) in kg Team Losses (TL) in USD M Salary per Game (S) in USD k	Support	Confidence	Lift
0–10	W: 100–130 kg and TL: 0–25 M	28%	50%	1.01
	H: 200–225 cm and TL: 0–25 M	32%	58%	1
	W: 100–130 kg	28%	50%	1
	TL: 0–25 M	55%	100%	1
	H: 200–225 cm	32%	58%	0.99
	S: 0–150 k	38%	68%	0.96
	S: 0–150 k and TL: 0–25 M	38%	68%	0.96
10–30	S: 0–150 k and TL: 0–25 M	14%	75%	1.06
	S: 0–150 k	14%	75%	1.05
	H: 200–225 cm	11%	60%	1.02
	H: 200–225 cm and TL: 0–25 M	11%	60%	1.02
	TL: 0–25 M	18%	100%	1
30–90	S: 0–150 k and TL: 0–25 M	7%	73%	1.03
	S: 0–150 k	7%	73%	1.03
	W: 100–130 kg	5%	51%	1.02
	W: 100–130 kg and TL: 0–25 M	5%	51%	1.02
	H: 200–225 cm and TL: 0–25 M	6%	59%	1.02
	H: 200–225 cm	6%	59%	1.01
	TL: 0–25 M	10%	100%	1
90–180	H: 200–225 cm	2%	61%	1.03
	H: 200–225 cm and TL: 0–25 M	2%	60%	1.03
	TL: 0–25 M	3%	100%	1
	S: 0–150 k	2%	69%	0.97
	S: 0–150 k and TL: 0–25 M	2%	69%	0.97
180+	S: 0–150 k	7%	73%	1.03
	W: 70–100 kg	5%	50%	1.03
	S: 0–150 k and TL: 0–25 M	7%	73%	1.02
	H: 200–225 cm	6%	58%	0.99
	TL: 0–25 M	10%	98%	0.98
	H: 200–225 cm and TL: 0–25 M	6%	56%	0.97
Career ending	S: 0–150 k and H: 200–225 cm and TL: 0–25 M	2%	51%	1.24
	S: 0–150 k	3%	87%	1.23
	S: 0–150 k and H: 200–225 cm	2%	51%	1.23
	S: 0–150 k and TL: 0–25 M	3%	87%	1.23
	W: 70–100 kg	2%	50%	1.03
	W: 70–100 kg and TL: 0–25 M	2%	50%	1.03
	TL: 0–25 M	4%	100%	1
	H: 200–225 cm	2%	57%	0.97
	H: 200–225 cm and TL: 0–25 M	2%	57%	0.97

The physical characteristics of athletes, including their height and weight, have a significant impact on their recovery duration. In particular, players who are tall, specifically those with a height range of 200–225 cm, and those who fall within specific weight categories exhibit distinct recovery patterns. These findings underscore the significance of a player's physical profile in the development of injury treatment and recovery protocols.

The study reveals a significant correlation between team financial losses and player recovery times. According to the findings, teams that suffer fewer financial losses (up to USD 25 million) tend to have shorter recovery periods for their players. This observation implies that financial considerations might play a role in how teams deal with injuries.

4. Discussion

In this comprehensive study, we delved into the complex domain of player injuries in the NBA, utilizing a multifaceted DM approach to dissect two decades of large amounts of quality data. Two pivotal questions propelled our research: first, the feasibility of identifying anomalous recovery times for various types of body injuries in the NBA using unsupervised learning methods and, second, unravelling the intricate patterns and associations between injury types, recovery durations, and sociodemographic and financial repercussions. The depth and breadth of our dataset, encompassing player sociodemographics, injury specificity, and financial data, set the stage for a nuanced exploration of these dynamics.

Our research not only provides insights into injury management within the NBA but also emphasizes several data mining challenges and the broader implications for the field and big-data studies. The integration of heterogeneous data sources required specific steps, as described in the previous methodology section, and a significant degree of result supervision, reflecting the effort-intensive nature of high-quality data curation in sports analytics.

The data engineering techniques could vary; however, the initial conceptualization of the final dataset required for analysis was crucial. Given the nature of the scraped data, cleansing, engineering, and the application of DM and ML techniques were crucial to enhancing data quality. An example of this is the implementation of anomaly detection in this research. As big-data problems often include abnormal records, the use of anomaly detection for each category was essential to achieving optimal data quality.

The application of anomaly detection algorithms sets unique challenges due to the diversity of the data. The iterative tuning of these algorithms highlights the importance of domain knowledge in sports data mining, as various factors influencing player recovery cannot be directly inferred from the data without the supervision of the results. Based on the aforementioned, the decision was made to select a voting ensembling technique for anomaly detection, targeting to optimize the efficiency of the data quality objective. This emphasizes the need for interdisciplinary expertise and supervision when conducting similar data mining research.

4.1. Unconventional Recovery Durations

The ensemble of DBSCAN, isolation forest, and Z score techniques utilized in our anomaly detection methodology played a pivotal role in uncovering unusual recovery durations in NBA players. Our findings suggest that only 1.04% of the total cases were considered to be anomalies, a testament to the specificity and effectiveness of our approach. These anomalies were primarily characterized by instances where players had unusually prolonged recovery periods, often due to leaving and then returning to the NBA. This aspect of our research is particularly noteworthy, as it extends beyond the conventional understanding of recovery timelines and introduces a new dimension to the understanding of player injuries and recoveries in professional sports.

The effectiveness of our approach for detecting anomalies in NBA recovery timelines was due to the successful integration of three distinct methods. We leveraged the clustering and noise identification capabilities of DBSCAN, the ability of the isolation forest to identify anomalies in high-dimensional spaces, and the statistical severity provided by Z score

analysis. This comprehensive approach not only improved the accuracy of our findings but also provided a more in-depth understanding of the recovery timelines. By carefully tuning these algorithms to the unique characteristics of our dataset, we were able to effectively filter out data points that deviated from the norm. This approach allowed us to conduct subsequent analyses with greater accuracy and insight, proceeding with association rules. Our results demonstrate that the integration of these methods is a promising approach for identifying anomalies in complex datasets.

4.2. Interplay between Player Attributes, Injury Types, and Recovery

Our research utilized ARM to uncover complex associations among NBA player data, focusing on injury types, recovery times, and their financial implications. We observed a notable correlation between specific injuries, such as thigh, foot, and groin injuries, and recovery duration, indicating the effectiveness of advanced treatment protocols in the NBA. Furthermore, players who opted for "Rest" exhibited a high level of confidence in returning within 0–10 days, highlighting the efficiency of preventive rest policies in the league.

Differences in recovery times for various injury types highlight the importance of customized treatment and rehabilitation strategies. Knee injuries, which are typically viewed as severe, occurred in a subset of players with unexpectedly short recovery times, indicating improvements in treatment techniques. On the other hand, injuries to the upper extremities and cranial areas exhibited a wide range of recovery times, indicating the varied effects of these injuries and the customized nature of treatment plans.

4.3. Financial Aspects and Recovery Dynamics

Our analysis revealed a captivating aspect of the financial consequences of player injuries. It was observed that players with lower salaries had relatively shorter recovery periods. This trend may be influenced by economic factors and contractual commitments, which implies that there could be a certain amount of pressure on these players to return to play earlier than they should. This finding points to a confluence between economics and player well-being that demands further study, particularly regarding the decision-making processes concerning injury management in professional sports.

In addition, the association between the economic losses of a team and the duration of a player's recovery time provides a new framework for examining the economic implications of injuries in professional sports. It was observed that teams with less financial loss had shorter player recovery times, indicating that financial factors might play a role in how teams manage player injuries. This correlation implies a complicated interplay between economic factors and sports medicine practices in the NBA, which could be indicative of more extensive trends in professional sports.

4.4. Physical Attributes and Recovery Patterns

The study also revealed that the physical characteristics of players significantly impact their recovery process. Our findings indicate that taller players, specifically those within the 200–225 cm height range and certain weight categories, exhibit unique patterns of recovery. These observations highlight the need for a comprehensive approach that considers an athlete's physical attributes while developing injury treatment and recovery protocols [58]. They also underscore the importance of gaining a deeper understanding of the interplay between physical traits, medical interventions, and training practices in sports.

4.5. Threats to Validity

In this section, we acknowledge and address potential limitations and biases that could have influenced the outcomes of our study. The primary threats to validity include the data collection and preprocessing processes in real-world big-data studies. Although our data collection was extensive, encompassing multiple seasons and a broad range of player information, the ordering, joining, and validation of the data were essential. Furthermore, the manual aspects of our data engineering and supervision processes were critical to this

research and could have introduced biases, potentially affecting our results. To face this, we applied a voting ensemble anomaly detection methodology and supervision of the results to ensure optimal data quality.

Another important subject of consideration is causal inferences [59]. The association rules we identified, particularly those linking salaries and recovery durations, should not be misinterpreted as causal relationships. Instead, they can be regarded as correlated evidence in the NBA. Factors such as a player's role, the minutes they played, and the intensity of their in-game activities also influence recovery times, warranting further exploration in future research.

Our study's focus was on NBA basketball, but also raises questions about generalizability in other sports. The unique physical demands and economic structures of professional basketball in the NBA suggest that our findings might not directly apply to other sports or leagues, each of which has its standards and protocols.

5. Conclusions and Future Work

Our research elaborates into the complexities of NBA player injuries and recovery times utilizing advanced techniques in DM and ML. Through the integration of methods such as DBSCAN, isolation forest, and the Z score in anomaly detection, we were able to gain a nuanced understanding of this convoluted domain. Our findings revealed that a sheer 1.04% of our dataset comprised unusual recovery durations, often resulting from extended recovery periods due to factors such as returning after prolonged absences. Consequently, our research emphasizes the need for a comprehensive approach that accounts for various influencing factors.

The application of association rule mining, specifically via the Apriori algorithm [60], has further enriched our understanding by revealing significant correlations between injury types and recovery durations and between sociodemographic and financial implications. For instance, injuries such as thigh, foot, and groin injuries resulted in shorter recovery periods, highlighting the efficacy of NBA treatment and rehabilitation protocols. Furthermore, our analysis has illuminated the pivotal significance of rest periods in player recuperation, as evidenced by the high probability of players returning within 0–10 days from "Rest" injuries. This finding suggests a potential strategic advantage in implementing rest periods for maintaining player health and career longevity.

Regarding the physical aspects of sports injuries, our research also examined the financial consequences, revealing a significant link between a player's salary level and their recovery period. This finding implies that in addition to health considerations, financial factors may play a role in determining the length of time professional athletes take to recover. These observations highlight the need for further investigation into the intersection of financial pressures and player well-being, especially in the context of injury management and recovery decision-making.

Additionally, our findings suggest that recovery duration is associated with players' physical characteristics, such as height and weight. This emphasizes the significance of individualized treatment and rehabilitation plans that consider the specific physical profile of each athlete. Such personalized approaches can result in more effective and customized strategies for injury recovery.

The findings of this investigation hold the promise of influencing the field of sports medicine and athletic training across diverse sporting domains. Future research involving the use of DM and ML could concentrate on studies aimed at comprehending the enduring consequences of injuries and recovery procedures on the professional lives of athletes. Such studies could offer valuable insights into the efficacy of existing rehabilitation protocols and the general path of athletes following an injury.

Furthermore, examining the economic dimensions of sports injuries could provide a comprehensive understanding of how financial incentives and contractual obligations influence injury management in professional sports, facilitating more ethical and player-centered approaches to injury recovery. Moreover, the employment of novel technologies

such as wearable devices [61] and AI-powered analytics for monitoring and managing player health represents a promising avenue for future research [62].

Hence, this state-of-the-art usage could lead to more precise and proactive injury prevention and recovery strategies, enhancing athlete welfare and optimizing performance in professional sports. In essence, this investigation not only contributes to a broader understanding of injury and recovery dynamics in the NBA but also sets the groundwork for future research and practical applications in sports health and management [63]. The interplay between player health, economic factors, and sports performance remains a critical area for ongoing exploration, as there is potential to significantly enhance athlete welfare and optimize performance in professional sports.

Author Contributions: Conceptualization: G.P.; methodology: G.P. and V.S.; software: G.P. and V.S.; validation: C.T., G.P. and V.S.; formal analysis: G.P. and V.S.; investigation: G.P. and V.S.; resources: C.T., G.P. and V.S.; data curation: G.P. and V.S.; writing—original draft preparation: G.P.; writing—review and editing: C.T., G.P. and V.S.; visualization: G.P.; supervision: C.T.; project administration: C.T. and G.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data partly available within the manuscript.

References

1. Brefeld, U.; Davis, J.; Van Haaren, J.; Zimmermann, A. (Eds.) *Machine Learning and Data Mining for Sports Analytics*; Springer: Cham, Switzerland, 2022. [CrossRef]
2. Rossi, A.; Perri, E.; Pappalardo, L.; Cintia, P.; Iaia, F. Relationship between External and Internal Workloads in Elite Soccer Players: Comparison between Rate of Perceived Exertion and Training Load. *Appl. Sci.* **2019**, *9*, 5174. [CrossRef]
3. Mehrotra, K.G.; Mohan, C.K.; Huang, H. Clustering-Based Anomaly Detection Approaches. In *Anomaly Detection Principles and Algorithms*; Springer: Cham, Switzerland, 2017; pp. 41–55.
4. Wang, X.; Huang, D.; Zhao, X. Design of the Sports Training Decision Support System Based on Improved Association Rule, the Apriori Algorithm. *Intell. Autom. Soft Comput.* **2020**, *26*, 755–763. [CrossRef]
5. Bahnert, A.; Norton, K.; Lock, P. Association between post-game recovery protocols, physical and perceived recovery, and performance in elite Australian Football League players. *J. Sci. Med. Sport* **2013**, *16*, 151–156. [CrossRef] [PubMed]
6. Maffulli, N.; Longo, U.G.; Gougoulias, N.; Caine, D.; Denaro, V. Sport injuries: A review of outcomes. *Br. Med. Bull.* **2011**, *97*, 47–80. [CrossRef] [PubMed]
7. Nwachukwu, B.U.; Anthony, S.G.; Lin, K.M.; Wang, T.; Altchek, D.W.; Allen, A.A. Return to play and performance after anterior cruciate ligament reconstruction in the National Basketball Association: Surgeon case series and literature review. *Physician Sportsmed.* **2017**, *45*, 303–308. [CrossRef] [PubMed]
8. Truong, L.K.; Mosewich, A.D.; Miciak, M.; Pajkic, A.; Le, C.Y.; Li, L.C.; Whittaker, J.L. Balance, reframe, and overcome: The attitudes, priorities, and perceptions of exercise-based activities in youth 12–24 months after a sport-related ACL injury. *J. Orthop. Res.* **2022**, *40*, 170–181. [CrossRef] [PubMed]
9. Hewett, T.E.; Myer, G.D.; Ford, K.R.; Paterno, M.V.; Quatman, C.E. Mechanisms, prediction, and prevention of ACL injuries: Cut risk with three sharpened and validated tools. *J. Orthop. Res.* **2016**, *34*, 1843–1855. [CrossRef]
10. Laver, L.; Kocaoglu, B.; Cole, B.; Arundale, A.J.H.; Bytomski, J.; Amendola, A. *Basketball Sports Medicine and Science*; Springer: Berlin/Heidelberg, Germany, 2020. [CrossRef]
11. Krosshaug, T.; Nakamae, A.; Boden, B.P.; Engebretsen, L.; Smith, G.; Slauterbeck, J.R.; Hewett, T.E.; Bahr, R. Mechanisms of Anterior Cruciate Ligament Injury in Basketball. *Am. J. Sports Med.* **2007**, *35*, 359–367. [CrossRef]
12. Kalaian, S.A.; Kasim, R. Predictive Analytics. In *Handbook of Research on Organizational Transformations through Big Data Analytics*; IGI Global: Hershey, PA, USA, 2015; pp. 12–29. [CrossRef]
13. Jauhiainen, S.; Kauppi, J.-P.; Leppänen, M.; Pasanen, K.; Parkkari, J.; Vasankari, T.; Kannus, P.; Äyrämö, S. New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes. *Int. J. Sports Med.* **2021**, *42*, 175–182. [CrossRef]
14. Terner, Z.; Franks, A. Modeling Player and Team Performance in Basketball. *Annu. Rev. Stat. Appl.* **2021**, *8*, 1–23. [CrossRef]

15. Sarlis, V.; Chatziilias, V.; Tjortjis, C.; Mandalidis, D. A Data Science approach analysing the Impact of Injuries on Basketball Player and Team Performance. *Inf. Syst.* **2021**, *99*, 101750. [CrossRef]

16. Cohan, A.; Schuster, J.; Fernandez, J. A deep learning approach to injury forecasting in NBA basketball. *J. Sports Anal.* **2021**, *7*, 277–289. [CrossRef]

17. Kester, B.S.; Behery, O.A.; Minhas, S.V.; Hsu, W.K. Athletic performance and career longevity following anterior cruciate ligament reconstruction in the National Basketball Association. *Knee Surg. Sports Traumatol. Arthrosc.* **2017**, *25*, 3031–3037. [CrossRef]

18. Khan, M.; Ekhtiari, S.; Burrus, T.; Madden, K.; Rogowski, J.P.; Bedi, A. Impact of Knee Injuries on Post-retirement Pain and Quality of Life: A Cross-Sectional Survey of Professional Basketball Players. *HSS J.* **2020**, *16*, 327–332. [CrossRef]

19. Harris, J.D.; Erickson, B.J.; Bach, B.R.; Abrams, G.D.; Cvetanovich, G.L.; Forsythe, B.; McCormick, F.M.; Gupta, A.K.; Cole, B.J. Return-to-Sport and Performance after Anterior Cruciate Ligament Reconstruction in National Basketball Association Players. *Sports Health* **2013**, *5*, 562–568. [CrossRef]

20. Iwamoto, J.; Ito, E.; Azuma, K.; Matsumoto, H. Sex-specific differences in injury types among basketball players. *Open Access J. Sports Med.* **2015**, *6*, 1–6. [CrossRef]

21. Taylor, J.B.; Ford, K.R.; Nguyen, A.D.; Terry, L.N.; Hegedus, E.J. Prevention of Lower Extremity Injuries in Basketball: A Systematic Review and Meta-Analysis. *Sports Health* **2015**, *7*, 392–398. [CrossRef]

22. Trojian, T.H.; Cracco, A.; Hall, M.; Mascaro, M.; Aerni, G.; Ragle, R. Basketball injuries: Caring for a basketball team. *Curr. Sports Med. Rep.* **2013**, *12*, 321–328. [CrossRef] [PubMed]

23. Minhas, S.V.; Kester, B.S.; Larkin, K.E.; Hsu, W.K. The Effect of an Orthopaedic Surgical Procedure in the National Basketball Association. *Am. J. Sports Med.* **2016**, *44*, 1056–1061. [CrossRef] [PubMed]

24. Afara, I.O.; Sarin, J.K.; Ojanen, S.; Finnilä, M.A.J.; Herzog, W.; Saarakkala, S.; Korhonen, R.K.; Töyräs, J. Machine Learning Classification of Articular Cartilage Integrity Using Near Infrared Spectroscopy. *Cell. Mol. Bioeng.* **2020**, *13*, 219–228. [CrossRef] [PubMed]

25. Maffulli, N.; Longo, U.G.; Spiezia, F.; Denaro, V. Sports Injuries in Young Athletes: Long-Term Outcome and Prevention Strategies. *Phys. Sportsmed.* **2010**, *38*, 29–34. [CrossRef] [PubMed]

26. Lu, Y.; Pareek, A.; Lavoie-Gagne, O.Z.; Forlenza, E.M.; Patel, B.H.; Reinholz, A.K.; Forsythe, B.; Camp, C.L. Machine Learning for Predicting Lower Extremity Muscle Strain in National Basketball Association Athletes. *Orthop. J. Sports Med.* **2022**, *10*, 232596712211117. [CrossRef]

27. Jauhiainen, S.; Kauppi, J.P.; Krosshaug, T.; Bahr, R.; Bartsch, J.; Äyrämö, S. Predicting ACL Injury Using Machine Learning on Data From an Extensive Screening Test Battery of 880 Female Elite Athletes. *Am. J. Sports Med.* **2022**, *50*, 2917–2924. [CrossRef] [PubMed]

28. Sarlis, V.; George, P.; Christos, T. Sports Analytics and Text Mining NBA Data to Assess Recovery from Injuries and Their Economic Impact. *Computers* **2023**, *12*, 261. [CrossRef]

29. Rehman, S.U.; Asghar, S.; Fong, S.; Sarasvady, S. DBSCAN: Past, present and future. In Proceedings of the Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014), Bangalore, India, 17–19 February 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 232–238.

30. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 413–422.

31. Rousseeuw, P.J.; Hubert, M. Anomaly detection by robust statistics. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1236. [CrossRef]

32. Borgelt, C.; Kruse, R. Induction of Association Rules: Apriori Implementation. In *Compstat*; Physica-Verlag HD: Heidelberg, Germany, 2002; pp. 395–400.

33. Zhang, C.; Zhang, S. (Eds.) Negative Association Rule. In *Association Rule Mining*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 47–84.

34. Chomątek, Ł.; Sierakowska, K. Automation of Basketball Match Data Management. *Information* **2021**, *12*, 461. [CrossRef]

35. swar. nba_api. 2023. Available online: https://github.com/swar/nba_api (accessed on 15 November 2023).

36. ESPN. NBA Stats. Available online: https://www.espn.com/nba/stats (accessed on 15 November 2023).

37. Glez-Peña, D.; Lourenço, A.; López-Fernández, H.; Reboiro-Jato, M.; Fdez-Riverola, F. Web scraping technologies in an API world. *Brief Bioinform.* **2014**, *15*, 788–797. [CrossRef]

38. Ochieng, P.J.; London, A.; Krész, M. A Forward-Looking Approach to Compare Ranking Methods for Sports. *Information* **2022**, *13*, 232. [CrossRef]

39. Zadeh, L.A. Fuzzy logic. *Computer* **1988**, *21*, 83–93. [CrossRef]

40. Alexandridis, G.; Varlamis, I.; Korovesis, K.; Caridakis, G.; Tsantilas, P. A Survey on Sentiment Analysis and Opinion Mining in Greek Social Media. *Information* **2021**, *12*, 331. [CrossRef]

41. Li, L.; Pratap, A.; Lin, H.-T.; Abu-Mostafa, Y.S. Improving Generalization by Data Categorization. In *Knowledge Discovery in Databases: PKDD 2005, Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, Porto, Portugal, 3–7 October 2005*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 157–168.

42. Vatsalan, D.; Bhaskar, R.; Gkoulalas-Divanis, A.; Karapiperis, D. Privacy Preserving Text Data Encoding and Topic Modelling. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1308–1316.

43. Brunnermeier, M.K.; Sannikov, Y. On the Optimal Inflation Rate. *Am. Econ. Rev.* **2016**, *106*, 484–489. [CrossRef]

44. Ali, T.; Asghar, S.; Sajid, N.A. Critical analysis of DBSCAN variations. In Proceedings of the 2010 International Conference on Information and Emerging Technologies, Karachi, Pakistan, 14–16 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1–6.

45. Birant, D.; Kut, A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data Knowl. Eng.* **2007**, *60*, 208–221. [CrossRef]

46. Li, C.; Guo, L.; Gao, H.; Li, Y. Similarity-Measured Isolation Forest: Anomaly Detection Method for Machine Monitoring Data. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]

47. Ferragut, E.M.; Laska, J.; Bridges, R.A. A New, Principled Approach to Anomaly Detection. In Proceedings of the 2012 11th International Conference on Machine Learning and Applications, Boca Raton, FL, USA, 12–15 December 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 210–215.

48. Ghafari, S.M.; Tjortjis, C. A survey on association rules mining using heuristics. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1307. [CrossRef]

49. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago de Chile, Chile, 12–15 September 1994; pp. 487–499.

50. Du, J.; Zhang, X.; Zhang, H.; Chen, L. Research and improvement of Apriori algorithm. In Proceedings of the 2016 Sixth International Conference on Information Science and Technology (ICIST), Dalian, China, 6–8 May 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 117–121.

51. Dasseni, E.; Verykios, V.S.; Elmagarmid, A.K.; Bertino, E. Hiding Association Rules by Using Confidence and Support. In *Information Hiding, Proceedings of the 4th International Workshop, IH 2001, Pittsburgh, PA, USA, 25–27 April 2001*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 369–383.

52. Scheffer, T. Finding Association Rules That Trade Support Optimally against Confidence. In Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery, Freiburg, Germany, 3–5 September 2001; pp. 424–435.

53. McNicholas, P.D.; Murphy, T.B.; O'Regan, M. Standardising the lift of an association rule. *Comput. Stat. Data Anal.* **2008**, *52*, 4712–4721. [CrossRef]

54. Fujita, M.; McGeer, P.C.; Yang, J.C.-Y. Multi-Terminal Binary Decision Diagrams: An Efficient Data Structure for Matrix Representation. *Form Methods Syst. Des.* **1997**, *10*, 149–169. [CrossRef]

55. Huyghe, T.; Alcaraz, P.E.; Calleja-González, J.; Bird, S.P. The underpinning factors of NBA game-play performance: A systematic review (2001–2020). *Phys. Sportsmed.* **2022**, *50*, 94–122. [CrossRef]

56. Lian, J.; Sewani, F.; Dayan, I.; Voleti, P.B.; Gonzalez, D.; Levy, I.M.; Musahl, V.; Allen, A. Systematic Review of Injuries in the Men's and Women's National Basketball Association. *Am. J. Sports Med.* **2022**, *50*, 1416–1429. [CrossRef]

57. Matthew, B. *Financial Management in the Sport Industry*; Routledge: London, UK, 2016. [CrossRef]

58. Mihajlovic, M.; Cabarkapa, D.; Cabarkapa, D.V.; Philipp, N.M.; Fry, A.C. Recovery Methods in Basketball: A Systematic Review. *Sports* **2023**, *11*, 230. [CrossRef]

59. Pearl, J. Causal inference in statistics: An overview. *Stat Surv.* **2009**, *3*, 96–146. [CrossRef]

60. Yakhchi, S.; Ghafari, S.M.; Tjortjis, C.; Fazeli, M. ARMICA-Improved: A New Approach for Association Rule Mining. In *Knowledge Science, Engineering and Management, Proceedings of the 10th International Conference, KSEM 2017, Melbourne, VIC, Australia, 19–20 August 2017*; Springer: Cham, Switzerland, 2017; pp. 296–306.

61. Ren, B.; Wang, Z.; Ma, K.; Zhou, Y.; Liu, M. An Improved Method of Heart Rate Extraction Algorithm Based on Photoplethysmography for Sports Bracelet. *Information* **2023**, *14*, 297. [CrossRef]

62. Xiao, J.; Tian, W.; Ding, L. Basketball Action Recognition Method of Deep Neural Network Based on Dynamic Residual Attention Mechanism. *Information* **2022**, *14*, 13. [CrossRef]

63. Pintér, G.; Felde, I. Analyzing the Behavior and Financial Status of Soccer Fans from a Mobile Phone Network Perspective: Euro 2016, a Case Study. *Information* **2021**, *12*, 468. [CrossRef]