Tian H. (2024) UTILIZING MACHINE LEARNING ALGORITHMS TO RESOLVE SPORTS CONTRACT DISPUTES. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 24 (97) pp. 106-120.
DOI: https://doi.org/10.15366/rimcafd2024.97.008

ORIGINAL

UTILIZING MACHINE LEARNING ALGORITHMS TO RESOLVE SPORTS CONTRACT DISPUTES

Hongqiao Tian

Power China Construction Group Co., Ltd, 100120, Beijing, China
Email: 17772929999@163.com

Recibido 08 de diciembre de 2023 **Received** December 08, 2023
Aceptado 07 de julio de 2024 **Accepted** July 07, 2024

ABSTRACT

This study explores the application of machine learning, specifically Support Vector Machine (SVM) models, in resolving sports contract disputes. Text mining techniques, including TF-IDF (Term Frequency-Inverse Document Frequency), were employed to process and extract significant features from unstructured contract dispute judgment documents collected from the China Judgments Online. The dataset encompassed various dispute categories such as athlete and club contract disputes, advertising endorsement disputes, transfer contract disputes, event organization and venue disputes, and copyright disputes. The SVM model, utilizing the nu-SVC variant with an RBF kernel, demonstrated high precision across most categories, effectively handling the high-dimensional and complex nature of legal texts. Key metrics such as precision, recall, and F-measure were used to evaluate model performance. The results highlight the robustness and accuracy of machine learning in classifying and analyzing sports contract disputes, providing a valuable tool for legal professionals to enhance dispute resolution processes. This study underscores the potential of integrating advanced computational techniques with legal analysis to improve the efficiency and effectiveness of resolving contractual conflicts in the sports industry.

KEYWORDS: Sports contract disputes, Machine learning, Support Vector Machine, TF-IDF.

1. INTRODUCTION

The rapid development of China's sports industry has led to a significant increase in the number and complexity of sports contracts. These contracts, encompassing athlete agreements, sponsorship deals, transfer agreements, and broadcast rights, play a crucial role in the functioning of the sports ecosystem (Lee, 2024). However, with the proliferation of these contracts, disputes have become increasingly common, posing significant challenges to stakeholders such as athletes, clubs, agents, sponsors, and governing bodies. This study aims to explore the application of machine learning algorithms in resolving sports contract disputes in China, offering a more efficient and accurate approach compared to traditional methods.

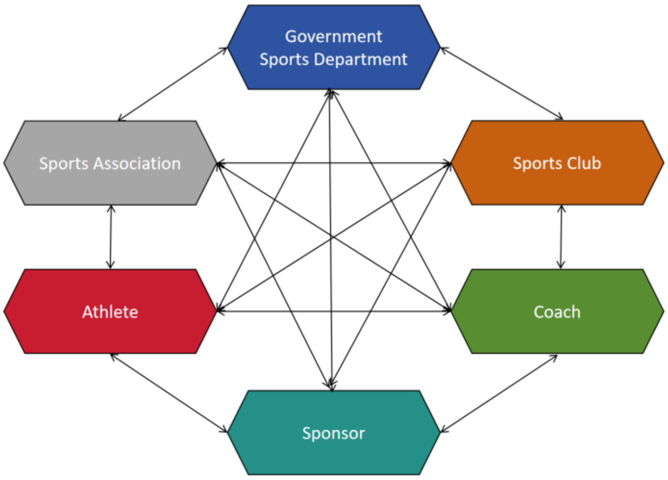

Figure 1: Stakeholders in Sports Contract Disputes

The Figure 1 illustrates the complex relationships among key stakeholders in sports contract disputes, including government sports departments, sports associations, sports clubs, coaches, sponsors, and athletes. These relationships form a network of interconnected contractual obligations and interests, where disputes can arise from various interactions. China's sports industry has experienced exponential growth, driven by the commercialization of sports events, media rights, and global sponsorship deals (Fischer, 2023). Major sports leagues, such as the Chinese Super League (CSL) and the Chinese Basketball Association (CBA), have seen substantial investments and increased visibility, attracting top athletes and significant media attention. This growth has resulted in a complex web of contractual relationships that govern various aspects of the industry. For instance, athlete contracts dictate terms related to salaries, playing time, training schedules, and injury management. Sponsorship agreements involve substantial financial commitments and stipulate promotional activities and brand endorsements. Transfer agreements outline the terms under which athletes move between clubs, often involving significant transfer fees and conditions. Broadcast contracts define the rights and obligations related to the media coverage of sports events. Despite their importance, these contracts are prone to disputes due to their inherent complexity and the dynamic nature of the sports industry. Disputes may arise from unclear contract terms, changes in circumstances, or breaches of contract. Traditional methods of dispute resolution, such as negotiation, arbitration, and litigation, often prove to be time-consuming, costly, and sometimes insufficiently specialized to address the unique aspects of sports contracts in China. Sports contract disputes in China can be broadly categorized into several types: Athlete Contract Disputes: These involve disagreements between athletes and clubs over issues such as compensation, playing time, training requirements, and injury management (Nafziger, 2022). Sponsorship Contract Disputes: These arise between sponsors and athletes or clubs, typically over sponsorship amounts, promotional obligations, and breach of sponsorship terms (Star & Vynckier, 2023). Transfer Contract Disputes: These occur during athlete transfers between clubs and involve disagreements over transfer fees, contract terms, and conditions of the transfer (Sroka, 2024). Media Rights Disputes: These involve conflicts over broadcast rights, including issues related to broadcast fees, content restrictions, and rights violations (Arnold & Ginsburg, 2020).

The causes of sports contract disputes in China are multifaceted and are influenced by the complex relationships among various stakeholders: Ambiguous Contract Terms: The complexity of sports contracts often leads to ambiguities, resulting in different interpretations and potential conflicts among stakeholders such as athletes, clubs, and sponsors. Changes During Contract Execution: The sports industry is highly dynamic, with factors such as athlete performance, club financial health, and market conditions constantly changing, leading to disputes during contract execution. For example, a club's financial instability can affect its ability to fulfill contractual obligations to athletes and coaches. Breach of Contract: Any party's failure to fulfill contractual obligations can lead to disputes. Common breaches include delayed salary payments, unauthorized contract terminations, and failure to deliver on sponsorship commitments. Such breaches can disrupt the relationships between clubs, athletes, and sponsors. Legal and Regulatory Differences: China has a unique legal framework governing sports contracts, and differences in interpretation and enforcement can lead to disputes, especially in international contracts involving Chinese parties. The involvement of government sports departments and sports associations adds another layer of complexity to these legal issues. Traditional methods for resolving sports contract disputes in China, while established, have notable limitations:

- Time-Consuming: Negotiation and litigation processes can be lengthy, delaying the resolution of disputes and disrupting the involved parties' regular activities. For instance, ongoing disputes between athletes and clubs can hinder performance and morale.

- Resource Intensive: These methods often require significant financial and human resources, adding to the burden on the disputing parties. This is particularly challenging for smaller clubs and less financially robust stakeholders.

• Lack of Specialization: Given the specialized nature of sports contracts, general legal practitioners may lack the specific expertise needed to effectively resolve disputes. Stakeholders such as coaches and sponsors may find it difficult to navigate these complexities without specialized legal support.

• Enforcement Challenges: Especially in international contexts, differences in legal systems and enforcement mechanisms can complicate the execution of dispute resolutions. This can be problematic for international sponsors and athletes dealing with Chinese sports entities.

With advancements in artificial intelligence, machine learning algorithms present a promising solution for improving the efficiency and accuracy of sports contract dispute resolution (Chou, Cheng, Wu, & Pham, 2014). Chou et al. found that hybrid machine learning models, particularly those combining multiple classification techniques (Chou, Tsai, & Lu, 2013), outperform single models in predicting public-private partnership (PPP) project disputes, with MLP+MLP and DT+DT models achieving the highest prediction accuracies. Singh et al. provided a comprehensive overview of sports law (Singh & Malik, 2024), emphasizing the need for international standardization, regulation of player contracts, transparency, effective dispute resolution, and adaptation to emerging trends and technology in global and Indian sports leagues. Ayhan et al. utilized machine learning techniques to predict construction disputes (Ayhan, Dikmen, & Talat Birgonul, 2021), identifying key factors through empirical data and Chi-square tests, with support vector machines achieving 90.46% accuracy and ensemble classifiers reaching 91.11% accuracy, offering a conceptual model and an early-warning mechanism for dispute prevention (Singh & Malik, 2024). By analyzing large datasets of sports contracts and dispute cases, machine learning models can:

• Identify Potential Risks: Detect clauses and factors within contracts that are likely to cause disputes, allowing for proactive mitigation. This can help clubs, athletes, and sponsors to avoid common pitfalls.

• Predict Dispute Outcomes: Analyze historical dispute data to forecast likely outcomes of similar disputes, providing valuable insights for stakeholders. This can assist government sports departments and sports associations in making informed decisions.

• Enhance Resolution Efficiency: Automate the analysis and processing of extensive contract and dispute information, significantly reducing the time and effort required for resolution. This benefits all parties involved, including athletes, clubs, coaches, and sponsors.

This study seeks to investigate the practical applications of machine learning algorithms in the resolution of sports contract disputes in China, aiming to demonstrate their potential in transforming traditional dispute resolution methods (Hubáček, Šourek, & Železný, 2019). By leveraging these technologies, stakeholders in the Chinese sports industry can achieve more efficient, accurate, and specialized dispute resolution, ultimately contributing to the stability and growth of the industry (Schwalbe, 2018). The remainder of this paper is structured as follows: Section 2 reviews the existing literature on the application of machine learning in legal contexts and the specific characteristics of sports contract disputes in China. Section 3 outlines the research methodology, including data collection, preprocessing, model selection, and evaluation metrics. Section 4 presents the experimental results and discusses the performance of different machine learning models. Section 5 provides a detailed discussion of the findings, highlighting the advantages and limitations of the proposed approach. Finally, Section 6 concludes the paper with a summary of key insights and suggestions for future research directions.

2. Methodology

Text mining is a method for extracting meaningful knowledge and information from semi-structured or unstructured text data. With the rapid development of network and information technology, informatization and digitization have become significant markers of the era. Information data has become an essential social resource, and in daily life, humans continuously generate and receive vast amounts of information. This information can be useless junk, chaotic, incomplete, or fragmented, making it challenging to utilize effectively. Identifying and converting the necessary information into valuable knowledge from this information flood is a crucial task.In text mining, the Term Frequency-Inverse Document Frequency (TF-IDF) technique is commonly used. TF-IDF measures the importance of words by calculating the term frequency (TF) and the inverse document frequency (IDF) across a document corpus. Due to the unstructured nature and ambiguity of language, processing text information is highly challenging. TF-IDF effectively helps in identifying the most representative words in a text, thus enhancing the efficiency of text mining. Text mining, as a research methodology, addresses these challenges by extracting useful information and knowledge from text data, a process often referred to as "knowledge discovery". This computational technique employs methods from multiple disciplines, including statistics, natural language processing, and machine learning. Techniques like TF-IDF, combined with computer technology and grammatical rules, allow for the extraction of useful information and data from massive amounts of text, thereby uncovering the latent, unknown information within texts. This opens new avenues for research and analysis of various phenomena. This section briefly introduces the designed methodology, which includes three main steps: text preprocessing, feature selection based on TF-IDF, and text classification using SVM. The following sections will describe each step in detail. Figure 2 illustrates the general architecture of the proposed method.

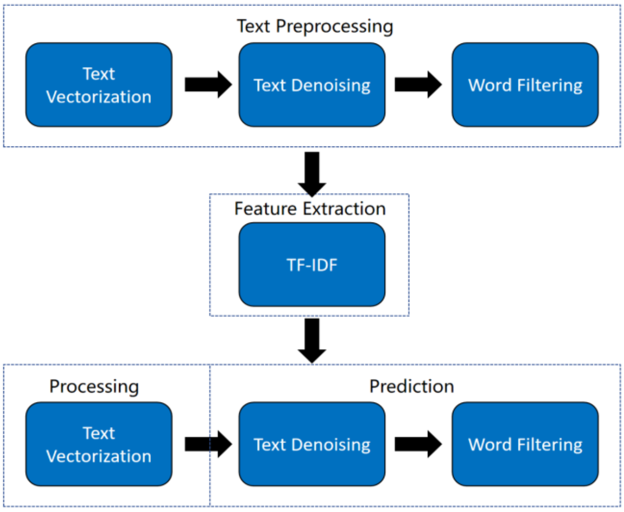

Figure 2: General architecture of the proposed method

2.1 Text Preprocessing

Sports contract dispute litigation documents directly reflect the process and causes of contract disputes. After detailed investigations into contract dispute litigation, court clerks record the investigation and trial results, forming judgment documents. These documents are publicly accessible and highly reliable, making them suitable for text mining. We collected and organized sports contract dispute litigation documents from the China Judgments Online. Using these contract judgment documents, we applied text mining methods to extract contract dispute factors and vectorized the unstructured text type of sports contract dispute litigation documents as the basic data for subsequent analysis. During the collection and screening process, we selected sports contract dispute judgment documents adjudicated and published by the Supreme People's Court. Due to the special nature of the publishing institution, the high quality of the text content was ensured, eliminating the need for corrections of typos in the judgment documents. However, sports contract dispute judgment documents are lengthy, comprising four parts: the header, the parties involved, the facts, and the reasons. The "header" and "parties" sections primarily contain information about the court, case reason, case number, plaintiff, defendant, and legal representatives, which are irrelevant to this study. The "facts" section contains the facts of the case determined by the court, including the contract signing date, contract terms, and the actions of both parties. Although this section includes part of the dispute process, it is recorded chronologically and vaguely, missing most dispute factors. The "reasons" section, however, provides a detailed analysis and judgment of the key points of the dispute, making it suitable for text mining. Therefore, only the "reasons" section of the judgment documents was selected as the research corpus to achieve better text mining results. We used the Jieba segmentation package for word segmentation. Jieba is a widely used Chinese word segmentation tool based on Python. It includes over 300,000 words, meeting basic segmentation needs and supporting custom dictionaries. By loading stop words and merging dictionaries, it can filter or merge selected words. During the segmentation process, Jieba matches the text content with words recorded in the dictionary, segmenting continuous long texts into shorter words. Thus, the quality of the dictionary directly affects the accuracy of segmentation and the extraction of contract dispute factors. Since Jieba's built-in dictionary lacks professional vocabulary related to contract disputes, we established a custom dictionary of contract dispute terms to meet our research needs. Additionally, we created a stop words dictionary to filter out frequently occurring but irrelevant words and a merging dictionary to handle synonyms and phrases.

2.2 Feature Extraction

Features are the basic units used to represent text. We selected contract dispute factor features based on the following criteria: ① the feature words should represent the main contract dispute factors in the judgment documents; ② the number of feature words should be within a reasonable range; ③ the feature words should be easily identifiable. Given the number of collected contract dispute judgment documents, we first set thresholds for specific feature parameters to screen out high-frequency words. We then traced the meanings of these high-frequency words in the judgment documents to identify contract dispute vocabulary, forming an initial set of contract dispute factors. We selected term frequency, document frequency, inverse document frequency, information entropy, and entropy-weighted term frequency as reference parameters for selecting feature words. Term Frequency (TF): Represents the frequency of a word in the text, calculated as:

$$TF_i = n_i \qquad (1)$$

Where n_i represents the total number of occurrences of the contract dispute factor F_i in all contract dispute judgment documents. Therefore, the higher the value ofTF_i, the more significant the role of this word in the collection of judgment documents. The Document Frequency (DF) represents the frequency of a term appearing in multiple contract dispute judgment documents. By utilizing the parameter DF, one can study how many judgment documents in the collection contain the contract dispute factorF_i. The formula is given as:

$$DF_i = |\{j: t_i \epsilon \ d_i\}| \qquad (2)$$

Where $|\{j: t_i \epsilon \ d_i\}|$ represents the number of judgment documents in which the term t_i appears. TF-IDF represents the term frequency-inverse document frequency. The formula is given as:

$$IDF = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \qquad (3)$$

$$TF-IDF = TF \times IDF = n_i \times \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (4)$$

From the formula, it can be seen that $TF-IDF$ is directly proportional to TF and inversely proportional to DF. $TF-IDF$ is often used to indicate the importance of a term in a document corpus. Information Entropy: Suppose the contract dispute factor F_i appears in *m* contract dispute judgment documents. Then the probability distribution of F_i in the contract dispute judgment document corpus can be expressed as P:

$$P_i = \frac{TF_j^i}{\sum TF_j^i} \quad (5)$$

Where TF_j^i represents the frequency of the contract dispute factor F_i in the contract dispute judgment document d_j. *H* represents the information entropy, and through the information entropy parameter *H*, we can study the distribution degree of the contract dispute factor F_i in the contract dispute judgment documents. The formula is:

$$H(F_i) = \sum P_i \log \frac{1}{P_i} = -\sum P_i \log P_i \quad (6)$$

The higher the value of $H(F_i)$, the more random the occurrence of the factor F_i. Entropy-weighted Term Frequency $TF-H$: $TF-H$ represents the entropy-weighted term frequency. The formula is given as:

$$TF-H = TF(F_i) \times H(F_i) = -n_i \times \sum P_i \log P_i \quad (7)$$

From the formula, it is evident that *TF-H* is directly proportional to *TF* and *H*. A higher *TF-H* value indicates that the term appears more uniformly and frequently in the contract dispute judgment document corpus, signifying that the factor F_i has a greater impact on the collection of contract dispute judgment documents.

2.3 Support Vector Machine

Support Vector Machine (SVM) is a classification technique that was first applied by Joachims for text classification. It is a powerful and supervised learning model based on the principle of minimizing structural risk (Barnett & Treleaven, 2018). During the training process, the SVM algorithm creates a hyperplane to separate positive and negative samples, and then it classifies new samples by determining their placement on the hyperplane. In this study, we used the nu-SVC type for the classification task and chose the RBF (Radial Basis Function) as the kernel type, as we wanted to map our training samples nonlinearly to higher-dimensional space. Additionally, we set the parameter nu to its maximum value of 0.5. Sports contracts often include numerous clauses and detailed information, and SVM has shown strong performance in handling high-dimensional data, effectively extracting and classifying important features. SVM maximizes the margin between classes, enabling it to accurately classify different types of dispute clauses (Gu, Foster, Shang, & Wei, 2019).Using the RBF kernel, SVM can map nonlinear problems to higher-dimensional space, showing excellent performance in handling complex contract terms and dispute classifications. SVM also exhibits strong robustness to noise and outliers, maintaining high classification accuracy even with errors or incomplete information in the text. By adjusting parameters such as ν\nuν and γ\gammaγ, SVM can flexibly control the number of support vectors and model complexity, adapting to different scales and complexities of contract text (Miller, 2015).We selected the nu-SVC type of SVM model because it can handle imbalanced data and allows control over the number of support vectors. To handle nonlinear data, we chose the Radial Basis Function (RBF) as the kernel. The RBF kernel is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (8)$$

Where γ is a parameter that controls the influence of individual training samples, x_i and x_j are data points. In the nu-SVC model, the parameter γ\gammaγ ranges from 0 to 1. We set it to the maximum value of 0.5 to balance the number of support vectors and training error. Using the collected sports contract dispute judgment data, we conducted model training. First, we applied the text preprocessing and feature extraction methods described earlier to convert the contract texts into numerical features. Then, these features were input into the SVM model for training. For the classification task, the SVM model seeks to find the optimal hyperplane by solving the following optimization problem:

$$\min_{w,b,\zeta} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\zeta_i \quad (9)$$

Where w is the weight vector, b is the bias term, ζ_i are the slack variables, and C is the penalty parameter. Subject to:

$$y_i(w \cdot \varphi(x_i) + b) \geq 1 - \zeta_i, \zeta_i \geq 0, \forall_i$$

Where y_i is the label of the sample, and $\varphi(x_i)$ is the function that maps the sample to a higher-dimensional space.

3. Experiments and analysis

3.1 Experimental Data

For the purposes of training and validating our SVM model in the resolution of sports contract disputes, we compiled and categorized a dataset of sports contract dispute litigation documents collected from the China Judgments Online. These documents were chosen based on their high reliability and public accessibility, as they were adjudicated and published by the Supreme People's Court. The dataset was categorized into various types of disputes and further divided based on specific dispute factors. Each category and factor was represented with a specific number of training and testing cases to ensure a robust evaluation of the SVM model's performance. The dataset was divided into training and testing sets to facilitate model development and evaluation. The distribution of cases across different dispute categories and factors is summarized in Table 1:

Table 1: Distribution of Training and Testing Cases Across Different Sports Contract Dispute Categories and Factors

DISPUTE CATEGORY	DISPUTE FACTOR	TRAINING CASES	TESTING CASES
ATHLETE AND CLUB CONTRACT DISPUTES	Salary	132	26
	Treatment	47	9
	Contract Termination	223	45
	Breach of Contract	157	31
ADVERTISING ENDORSEMENT CONTRACT DISPUTES	Endorsement Fees	241	48
	Ad Content	76	15
	Contract Duration	28	6
TRANSFER CONTRACT DISPUTES	Transfer Fees	77	15
	Buyout Fees	148	30
	Breach of Contract	57	11
EVENT ORGANIZATION AND VENUE DISPUTES	Venue Rent	43	9
	Event Date	37	7
	Venue Management	92	18
COPYRIGHT CONTRACT DISPUTES	Broadcasting Rights	18	4
	Sponsorship Contracts	67	13
	Copyright Fees	173	35
	Authorization Scope	136	27
	Breach of Contract	215	43
TOTAL		1967	392

This distribution ensures a comprehensive coverage of different types of disputes and factors, allowing for a thorough evaluation of the SVM model's classification capabilities. Each category and factor was carefully annotated to reflect the real-world complexities and nuances of sports contract disputes, providing a solid foundation for model training and testing.

3.2 Evaluation Metrics

To evaluate the performance of our SVM model in classifying sports contract disputes, we used three key metrics: Recall, Precision, and F-measure. These metrics provide a comprehensive assessment of the model's accuracy and reliability in identifying relevant dispute cases. Before defining the evaluation metrics, it is essential to understand the following terms:

TP (True Positives): The number of correctly identified positive cases. FN (False Negatives): The number of positive cases that were incorrectly classified as negative. FP (False Positives): The number of negative cases that were incorrectly classified as positive. TN (True Negatives): The number of correctly identified negative cases. Recall, also known as Sensitivity or True Positive Rate, measures the proportion of actual positive cases that are correctly identified by the model. It is particularly important in scenarios where missing a positive case is costly. The formula for Recall is:

$$Recall = \frac{TP}{TP + FN} \qquad (10)$$

Precision, also known as Positive Predictive Value, measures the proportion of positive cases identified by the model that are actually positive. It reflects the model's accuracy in identifying relevant cases among the predicted positives. The formula for Precision is:

$$Precision = \frac{TP}{TP + FP} \qquad (11)$$

F-measure, or F1 Score, is the harmonic mean of Recall and Precision, providing a balanced metric that considers both false positives and false negatives. It is especially useful when there is an uneven class distribution. The formula for the F-measure is:

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (12)$$

The F-measure ranges from 0 to 1, where a higher value indicates better performance. These metrics are crucial for assessing the effectiveness of the

SVM model in correctly identifying and classifying sports contract disputes, ensuring that the model provides reliable and actionable results in real-world applications.

3.3 Analysis of Experimental Results

The experimental results, as summarized in Table 2, provide a detailed overview of the precision of our SVM model in classifying various sports contract dispute factors. While the overall performance is promising, several factors have influenced the results. This section analyzes these factors and their impact on the experimental outcomes.

Table 2: Precision of Testing Cases

DISPUTE CATEGORY	DISPUTE FACTOR	PRECISION OF TESTING CASES (%)
ATHLETE AND CLUB CONTRACT DISPUTES	Salary	88.38
	Treatment	93.53
	Contract Termination	90.15
	Breach of Contract	95.07
ADVERTISING ENDORSEMENT CONTRACT DISPUTES	Endorsement Fees	86.69
	Ad Content	94.63
	Contract Duration	88.03
TRANSFER CONTRACT DISPUTES	Transfer Fees	93.49
	Buyout Fees	94.65
	Breach of Contract	93.02
EVENT ORGANIZATION AND VENUE DISPUTES	Venue Rent	95.07
	Event Date	90.96
	Venue Management	96.06
COPYRIGHT CONTRACT DISPUTES	Broadcasting Rights	87.59
	Sponsorship Contracts	89.87
	Copyright Fees	91.09
	Authorization Scope	93.18
	Breach of Contract	88.92

The quality of the text data and the preprocessing steps are critical to the performance of the SVM model. Accurate text cleaning and segmentation are essential to ensure that the features extracted are representative of the underlying dispute factors. Any noise, such as irrelevant information or typos, can mislead the model, reducing its ability to correctly classify disputes. For example, the lower precision in "Endorsement Fees" (86.69%) compared to "Ad Content" (94.63%) suggests that variability and noise in the text data for endorsement fees may have impacted the model's accuracy. The choice of features used to represent the text data significantly affects the model's performance. In this study, TF-IDF was used for feature extraction, which has proven effective for many text classification tasks. However, the inclusion of more sophisticated features, such as word embeddings, could potentially improve the model's ability to capture semantic nuances. The high precision in categories like "Venue Management" (96.06%) indicates that the chosen features were effective for clearly defined terms, while the slightly lower precision in "Broadcasting Rights" (87.59%) suggests room for improvement with more advanced feature extraction techniques. The performance of the SVM model is highly dependent on the selection of appropriate hyperparameters. In this study, the nu-SVC variant was used with the RBF kernel, and parameters were carefully tuned. Suboptimal parameter settings can lead to underfitting or overfitting, which in turn affects precision. The kernel choice is crucial as well; while the RBF kernel is suitable for capturing non-linear patterns, other kernels might better fit specific dispute categories. The generally high precision values across different factors, such as "Buyout Fees" (94.65%) and "Venue Rent" (95.07%), demonstrate that the model parameters were well-tuned for these categories. The dataset's imbalance, with varying numbers of cases across different dispute categories, can bias the model towards more frequent classes. For instance, categories with fewer training cases, such as "Broadcasting Rights" (87.59%) and "Contract Duration" (88.03%), showed lower precision compared to categories with ample training data, like "Contract Termination" (90.15%) and "Endorsement Fees" (86.69%). Techniques such as resampling, synthetic data generation (e.g., SMOTE), or cost-sensitive learning could be employed to mitigate class imbalance and improve precision. A larger and more diverse training dataset typically enhances the model's generalization ability. The precision in categories like "Endorsement Fees" (86.69%) and "Contract Termination" (90.15%) reflects the benefits of having substantial and varied training data. Conversely, limited training data can lead to overfitting, where the model performs well on training data but poorly on unseen cases. Ensuring that the training data is representative of the real-world distribution of disputes is crucial for robust performance. The relatively high precision in well-represented categories, such as "Breach of Contract" in both Athlete and Club (95.07%) and Copyright (88.92%) disputes, highlights the importance of sufficient training data. Legal texts often contain complex and ambiguous language, which poses a challenge for text classification models. The model's slightly lower precision in categories such as "Salary" (88.38%) and "Breach of Contract" (88.92%) in Copyright disputes may be due to the nuanced and context-dependent nature of these disputes. Incorporating more advanced natural language processing techniques that can capture contextual information and implicit meanings could enhance model performance. While precision is a critical metric, relying solely on it may not provide a complete picture of the model's performance. Including additional metrics such as recall, F-measure, and accuracy would offer a more comprehensive evaluation. Moreover, employing robust validation techniques like k-fold cross-validation can provide more reliable assessments and help in fine-tuning the model to achieve better generalization. External factors, such as changes in laws or regulations, can influence the nature of disputes and the relevance of past cases. Incorporating domain-specific knowledge and legal expertise into the feature selection and interpretation processes can significantly enhance the model's accuracy and relevance.

4. Conclusion

The experimental results demonstrate that the SVM model effectively classifies sports contract disputes with high precision, showcasing the significant role of machine learning in this domain. The model achieved impressive precision in categories such as "Treatment" (93.53%) and "Venue Management" (96.06%), highlighting its ability to handle high-dimensional data and diverse dispute types. The slightly lower precision in some areas, like "Endorsement Fees" (86.69%), indicates opportunities for further refinement. Overall, machine learning, particularly SVM, proves to be a powerful tool for resolving complex contractual disputes in the sports industry, offering reliable and data-driven insights to enhance legal analysis and decision-making.

Reference

Arnold, R., & Ginsburg, J. C. (2020). Comment: Foreign Contracts and US Copyright Termination Rights: What Law Applies? *Columbia Journal of Law & the Arts, 43*(4).

Ayhan, M., Dikmen, I., & Talat Birgonul, M. (2021). Predicting the occurrence of construction disputes using machine learning techniques. *Journal of construction engineering and management, 147*(4), 04021022.

Barnett, J., & Treleaven, P. (2018). Algorithmic dispute resolution—The automation of professional dispute resolution using AI and blockchain technologies. *The Computer Journal, 61*(3), 399-408.

Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Pham, A.-D. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Systems with Applications, 41*(8), 3955-3964.

Chou, J.-S., Tsai, C.-F., & Lu, Y.-H. (2013). Project dispute prediction by hybrid machine learning techniques. *Journal of Civil Engineering and Management, 19*(4), 505-517.

Fischer, M. (2023). Sports Marketing in China: Ball and the Wall. In *Selling to China: Stories of Success, Failure, and Constant Change* (pp. 127-157): Springer.

Gu, W., Foster, K., Shang, J., & Wei, L. (2019). A game-predicting expert system using big data and machine learning. *Expert Systems with Applications, 130*, 293-305.

Hubáček, O., Šourek, G., & Železný, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting, 35*(2), 783-

796.
Lee, J.-K. (2024). Assessing the Specificity of Intellectual Property Rights and Contractual Agreements for E-Sports Athletes in South Korea. *Law and Economy, 3*(3), 12-22.
Miller, T. W. (2015). *Sports analytics and data science: winning the game with methods and models*: FT press.
Nafziger, J. A. (2022). International sports law. In *Handbook on International Sports Law* (pp. 2-34): Edward Elgar Publishing.
Schwalbe, U. (2018). Algorithms, machine learning, and collusion. *Journal of Competition Law & Economics, 14*(4), 568-607.
Singh, A., & Malik, K. (2024). Prospects for Legal Evolution in Sports Contracts: An In-Depth Study.
Sroka, R. (2024). Financial fair play and the Court of Arbitration for Sport. *Journal of Global Sport Management, 9*(2), 285-304.
Star, S., & Vynckier, J. (2023). Commercial contracts: Sponsoring and ticketing. In *Research Handbook on the Law of Professional Football Clubs* (pp. 194-216): Edward Elgar Publishing.