*Turkish Journal of Civil Engineering, 2026, 105-136, Paper 833, **Research Article***

Early Prediction of Construction Disputes: Decision Support Systems with Machine Learning Techniques

Mahmut SARI[1]
Savaş BAYRAM[2*]
Emrah AYDEMIR[3]

ABSTRACT

This study aims to predict the outcomes of construction disputes before they proceed to litigation and to foster a constructive environment between parties. Within the scope of the study, a total of 24 legal factors; 14 legal factors were identified through extensive literature review and 10 legal factors were identified through content analysis. These legal factors were used in three stages: Pre-Litigation (A, B) and Post-Litigation. Legal factors with significant relationships were tested with 24 different machine learning algorithms. NB Tree, Logit Boost and LMT algorithms achieved 63.79%, 63.66% and 86.90% accuracy for models A, B and C, respectively.

Keywords: Construction disputes, machine learning algorithms, dispute resolution.

1. INTRODUCTION

The construction industry, which has an interdisciplinary working environment by nature, hosts different occupational groups together to create a unique structure. Besides, the sector provides economic mobility due to obligations such as the high quantity and diversity of resources. Moreover, the construction industry can direct the economy in developing countries. The sector represents 13% of the global gross domestic product (GDP), a volume of 10 trillion dollars (Barbosa, 2017). It is estimated that the industry will reach a value of $ 14 trillion by 2025. This situation indicates that stakeholders in the construction sector must continuously improve, modernize, and develop to maintain their competitive edge in both national and international markets [1]. The sector's significant impact on employment is

Note:

- This paper was received on January 13, 2025 and accepted for publication by the Editorial Board on October 19, 2025.
- Discussions on this paper will be accepted by May 31, 2026.

• https://doi.org/10.18400/tjce.1618975

1 Kırsehir Ahi Evran University, Department of Civil Engineering, Kırşehir, Türkiye
mahmutsari@ahievran.edu.tr - https://orcid.org/0000-0002-9298-5518

2 Erciyes University, Department of Civil Engineering, Kayseri, Türkiye
sbayram@erciyes.edu.tr - https://orcid.org/0000-0002-0153-6750

3 Sakarya Uni., Sakarya Business School, Dept. of Management Information Systems, Sakarya, Türkiye
emrahaydemir@sakarya.edu.tr - https://orcid.org/0000-0002-8380-7891

* Corresponding author

evidenced by its contribution of 8–15% to the global Gross Domestic Product (GDP) and its employment of approximately 7–9% of the labor force [2-4]. However, the sector also faces unique challenges with economic consequences, such as low productivity, insufficient collaboration, project delays, and cost overruns [5-7]. Due to these challenges, disputes may arise at any stage of project management, from the preliminary design process to project delivery, whenever the interests of the parties fail to align [8, 9]. Additionally, the sector's vulnerability to market fluctuations, such as risks and financial issues observed during the COVID-19 pandemic [10], has made disputes inevitable.

Inevitable disputes cause significant financial and emotional losses for the parties involved [11]. Moreover, due to the complex nature of construction projects, their lengthy durations, and substantial budgets, it is extremely challenging for project managers to foresee or plan for all possible contingencies [12]. To mitigate the adverse effects of disputes, it is essential to address issues just in time. Recent research indicates that the global construction sector faces an average dispute value of $52.6 million, with disputes lasting an average of 15.4 months [13, 14]. In this context, project managers and their teams require systems capable of predicting disputes at both early and advanced stages [15, 16]. Such systems enhance the responsibility of stakeholders to adopt more constructive behaviors in resolving disputes [17]. Zheng [18], used machine learning models to predict court outcomes of disputes in PPP projects. Ayhan, Dikmen [19], developed a model to predict whether a dispute would arise based on attributes collected through surveys. In addition to these, the case study method has predominantly been used in court rulings to identify attributes effective in predicting construction disputes [20-24]. However, is there a more suitable method for analyzing large datasets of similar cases to uncover general patterns? Can disputes be predicted before proceeding to litigation? Could the expertise of Court of Cassation judges be leveraged in such predictive models? Existing studies fall short in adequately addressing these questions.

If alternative dispute resolution methods are not included in construction contracts, resolution through litigation becomes mandatory [23]. Problems such as the excessive workload of courts [25, 26], increasing costs for rights holders [27], and slow judicial processing [23] lead to significant losses for both government agencies and contractor firms in construction projects that require substantial investment amounts. Moreover, although litigation requires a lengthy, detailed, and factual discovery process, it is evident that judges possess profound legal knowledge and experience. The decision-making process is subjective, relying on this deep legal knowledge and experience [19, 28]. There is a need for the use of content analysis, a more systematic and objective method for documenting what judges say and do [29]. Based on this deep legal knowledge, it is necessary to develop decision support models that minimize subjectivity, reduce litigation durations, and rely on experience and past cases. This study aims to fill this gap by predicting construction dispute case outcomes, particularly in the early stages. Within the scope of this purpose, models have been developed to predict the outcomes of construction dispute cases with machine learning methods using legal factors to be identified by content analysis method. The first section of the study provides a broad perspective on prediction studies concerning construction disputes. The second section introduces the research methodology that explains the identification of the legal factors affecting construction disputes in Türkiye, which are referred to and finalized by the Court of Cassation, from the case texts by content analysis method and the establishment of decision prediction models for the dispute stages defined by the identified legal factors. The next section presents the legal factors identified from 933 case files, the significance and power relationship results of these legal factors and the results of the established models. In the last section, the conclusions of the study and suggestions for future research are presented.

2. LITERATURE REVIEW

In recent years, competition in the public and private construction sectors has intensified, making it increasingly difficult to maintain balance within the project management iron triangle of cost, time, and quality. As a result, decreasing profit margins have made disputes inevitable [30, 31]. Litigation is commonly used to resolve disputes [32]. In litigation, parties encounter several obstacles, including lengthy judicial processes, the lack of confidentiality, and the multi-tiered appeal process [33]. Despite its shortcomings and the existence of alternative dispute resolution methods, litigation remains the preferred option in the global construction sector [13, 14]. The underlying reason for this preference is that stakeholders often fail to accurately identify the elements causing the dispute and to select the appropriate alternative dispute resolution methods [34].

The prediction of possible litigation outcomes has recently become a significant topic. As a pioneering study, 102 case files related to construction disputes filed in the Illinois appellate courts between 1982 and 1992 were collected, and 45 input and 8 output variables were identified. Using Artificial Neural Networks (ANN), the possible outcome of litigation was predicted with a success rate of 67% [20]. In subsequent studies, case files and years from the same appellate courts were diversified in terms of both variables and methods. Case-Based Reasoning (CBR) and Boosted Decision Trees (BDT) achieved prediction accuracies of 83% [21] and 90% [35] respectively. The early prediction of disputes allows parties to become aware of possible outcomes at an early stage. This, in turn, encourages them to adopt a more constructive and solution-oriented approach. This opportunity has led to the use of various modern approaches for predicting possible litigation outcomes. Particle Swarm Optimization (PSO) has also been used for such predictions. With 13 attributes, 1,105 judicial decisions were collected, achieving a prediction accuracy of 80% [36]. n another study, the same number of data points and attributes were used, and CBR achieved a prediction accuracy exceeding 80% [37]. Various methods have been used in the same area of research. Chen and Hsu [38] employed a hybrid ANN-CBR model. The data were based on the decisions of high and appellate courts across 48 states in the United States. The hybrid ANN-CBR model not only used ANN to predict the possible outcomes of judicial authority but also employed CBR to generate early warnings and provide information about previous cases. Consequently, an accuracy rate of 85% was achieved with the hybrid ANN-CBR model. However, some limitations regarding the effectiveness and generalizability of the model used in this study are noteworthy. The hybrid model was not compared with other artificial intelligence methods, and the data source relied solely on 48 U.S. states, limiting the model's applicability and reproducibility in different judicial systems. In this study, the possible outcomes of disputes are predicted with 24 different algorithms by using the decision texts of the highest decision authority of a country. The results obtained will be compared and a study with applicability and reproducibility to different legal systems will be produced.

Pulket and Arditi [39], utilized Ant Colony Optimization (ACO). A dataset of 151 appellate court rulings from Illinois, dated between 1987 and 2005, was used. ACO achieved a prediction accuracy rate of 92%, which is higher than previous studies: 67% with ANN [20], 83% with CBR [21] and 90% with BDT [35]. Pulket and Arditi [40], developed a Universal Prediction Model (UPM) based on the same court files as in their previous study. A code was created using WEKA to automate the knowledge analysis procedure. It was noted that UPM is both adaptable and extensible. In another study, Arditi and Pulket [41] developed an Integrated Prediction Model (IPM) to achieve better performance than their previous results. The data were collected from 132 appellate court cases in Illinois, dated between 1992 and 2000, and a prediction accuracy rate of 91% was achieved. However, due to the study's focus on a single state and the limited number of cases, its results are insufficiently generalizable to different judicial systems or broader geographical contexts. Unlike this limited geographical and legal context, this study will work with a dataset that provides geographical and legal coherence.

Mahfouz and Kandil [23], conducted a detailed evaluation to understand the contributions of legal factors to classification and prediction. They proposed a model that automatically predicts the outcomes of litigation decisions related to differing site conditions (DSC). A total of 400 case files from between 1912 and 2007 were collected, and the third-degree polynomial SVM model outperformed others with a prediction accuracy of 98%. However, despite the impressive 98% accuracy rate of the third-degree polynomial SVM model with a limited dataset, the study's prominent shortcomings include the lack of explanation regarding the hyperparameter settings affecting model performance and the methodological details of comparisons with other methods.

In relevant studies, the performance of hybrid classification techniques, in addition to single machine learning (ML) techniques, has begun to attract attention. Chou, Tsai [42] used 645 dispute cases from public-private partnerships (PPP) in the Taiwan Public Construction Commission (TPCC) database to predict disputes at early stages using single and hybrid classification techniques. Techniques such as Multilayer Perceptron (MLP), Decision Tree (DT), SVM, NB, and k-Nearest Neighbors (k-NN) were used together. The MLP+MLP and DT+DT models emerged as the best models with prediction accuracies of 97% and 96%, respectively. In another study, Ayhan, Dikmen [19], Ayhan et al. (2021) predicted the occurrence of disputes using machine learning techniques on survey data. A conceptual model and questionnaire defining the attributes that cause disputes were developed. Using 108 survey data, the occurrence of the dispute was predicted and an average prediction accuracy of 91% was obtained. However, the study did not address how its results could be applied or contribute to decision-making processes, leaving its practical applicability unexplored. Within the scope of this study, the gap identified in the literature regarding the practical application of prediction models will be addressed, enabling the transformation of these models into usable applications.

In another study, a predictive approach was developed using machine learning models to identify the main causes of disputes in PPP projects in China and to predict court outcomes [18]. A total of 171 PPP cases from China Judgments Online between 2013 and 2018 were collected, and 17 legal factors were identified using the case analysis approach. Nine ML models were developed and validated. The best prediction accuracy of 96% was achieved with an ensemble model that included Gradient Boosting Decision Tree (GBDT), k-NN, and MLP. This study also demonstrated that the 17 identified legal factors are acceptable for representing and predicting court outcomes in PPP disputes. However, in this and other studies [20-24] the case analysis method used to identify legal factors, while noted for its effectiveness in providing contextual information, has been frequently criticized for being open to subjective interpretations and for limiting the generalizability of results [43]. In Türkiye, there have been numerous efforts to identify the causes of construction disputes using content analysis and statistical methods [9, 11, 44, 45]. Among the prominent studies is one that examined construction dispute files referred to regional courts and the court of cassation using content analysis methods to identify the causes of disputes and the duration of proceedings [46]. Another study identified the operational barriers to the adoption of smart contracts in construction projects through an extensive literature review and determined their importance and weight using the Fuzzy Analytic Hierarchy Process (AHP) [47]. Another study addressing readability issues in construction contracts examined the identification of these causes and the processes leading to disputes [48]. In Türkiye, the 166 decisions of the Directorate of High Technics Board, an institution to which the public party can apply in public construction disputes, were examined using qualitative research methods, and the causes of disputes were categorized [49]. Another prominent study carefully examined whether the defendant made a defense in construction dispute decisions brought before the Court of Cassation, extracted attributes from 2,563 decisions using Natural Language Processing (NLP) methods, and predicted judicial decisions with 87.38% accuracy using Machine Learning methods [50]. The content analysis method for identifying legal factors allows for the systematic and objective examination of datasets and produces more consistent and reproducible results, especially when working with large datasets [51]. Additionally, the coding-based structure of the content analysis method facilitates better organization and analysis of data while enhancing methodological consistency [52]. For this reason, in this study, it is aimed to identify the legal factors affecting the outcome of the litigation from the case texts by using the content analysis method instead of the case study analysis method and to present more objective and generally valid results by increasing the scientific validity of the legal factors.

Findings from the literature review indicate that the number of studies focusing on disputes in the construction sector and the methods used in these studies has increased recently. However, it has been observed that these studies often utilize project data representing a limited geographical-legal region and involving restricted contract, project, and dispute types. This threatens the applicability and reproducibility of the study results in different legal systems. Although some studies have addressed international disputes and a broader range of contract types, these studies have relied on survey data rather than litigation decisions. Moreover, the practical applicability of the study results has not been sufficiently detailed, making it difficult to understand their impact on real cases. The ability of stakeholders to comprehend and interpret judicial decisions in line with their interests is another issue that has not been emphasized in the literature. The absence of a comparative prediction model representing the entire construction sector for each stage of the litigation process, including pre-litigation (early stage) and post-litigation, constitutes a gap in the literature. In this context, the six main issues identified in the literature are: focusing on limited geographical-legal regions, selecting limited or potentially unreliable data collection tools such as surveys, using subjective and non-generalizable methods to identify the legal factors causing disputes, comparing only a limited number of algorithms in predictions, centering on limited project, contract, and dispute types, and focusing on models targeting a single stage of disputes. For this reason, this study aims to: (i) focus on a country (Türkiye) with legal coherence instead of a specific region; (ii) eliminate reliability issues by collecting statistically representative litigation decision texts directly from courts; (iii) use a more systematic and objective method to identify the legal factors causing disputes; (iv) perform predictions using and comparing 24 different machine learning algorithms; and (v) predict court outcomes at different stages of disputes.

3. METHODOLOGY

In this study, firstly, the legal factors affecting construction disputes will be determined. In determining these legal factors, a comprehensive literature review and content analysis of the

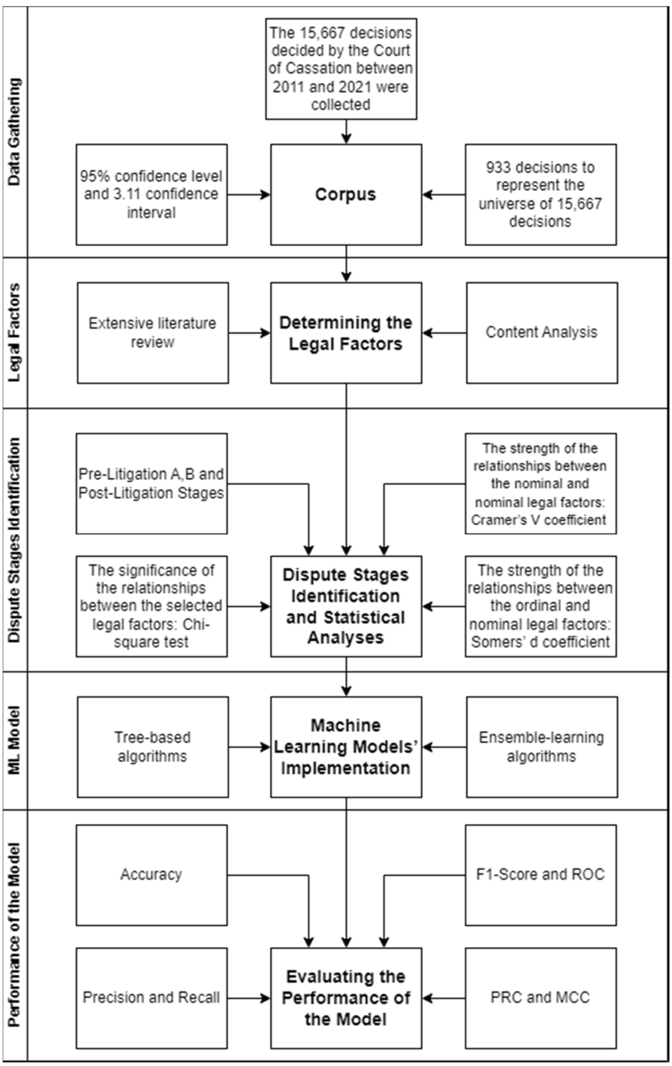

Figure 1 - Research flowchart of the study.

texts of the decisions given by the Court of Cassation, which is the highest decision-making authority in construction disputes in Türkiye, will be used. By defining the different stages of the disputes, the legal factors that have significant relationships with the case outcomes will be identified using statistical methods. Using 24 different machine learning methods, models predicting case outcomes will be developed and compared. The methodology of the study consists of five main stages presented in Figure 1.

3.1. Data Gathering

The data used in this study consists of construction dispute decisions of the Court of Cassation collected from the online legal database named 'Hukukturk' [53]. The data obtained includes cases decided between 2011 and 2021. These decisions were collected from the 6th Civil Chamber of the Court of Cassation, as it is the highest court in Türkiye to decide on construction disputes. The total number of decisions published by the Court of Cassation between 2011 and 2021 is 15,667 [54]. The target data set number of this study was determined based on the population-sample relationship. The targeted sample number was calculated as 933 decisions to represent the universe of 15,667 decisions with 95% confidence level and 3.11 confidence interval. The 933 decisions of the Court of Cassation represent all the decisions analysed. There is no decision that has not been analysed. The 933 decision texts used in this study were compiled from texts published in the public database after being anonymised by the relevant departments of the Court of Cassation; therefore, the raw decisions have been stripped of personal data from the publication stage onwards. No identifying information or distinguishing features that could reveal individual cases were used in the analyses. In addition, the content of the decisions varies between 2 and 20 pages. In the study, the files that did not contain at least half of the legal factors determined within the scope of the research, such as some approval decisions with very short text content, were excluded from the data set. Therefore, a total of 15,667 files were analysed by the researchers, but 933 files were used in the study. When the decisions of the Court of Cassation are analysed, it can be stated that the lack of a general-valid decision writing method applied by all Court of Cassation judges is among the reasons for this situation. Therefore, the judgements representing the 'widest range' were used in the whole sample.

3.2. Determining the Legal Factors and Transformation

In order to identify the legal factors that are effective in the occurrence of construction disputes, firstly a wide literature review was conducted. After the literature review, in the second step, the case files were analysed using the content analysis method, which is a more systematic and objective method than the case analysis method, and the process was meticulously designed in order to identify the factors that are effective in the occurrence of disputes. Holsti [55] stated that content analysis is the process of evaluating scattered data according to their scope. Smith [56] defined content analysis as a tool that collects data, puts them into a standard form and compares/transforms them with other forms. The aim of content analysis is to clarify the themes in the text and to reveal the relationships between concepts, as well as to bring together similar data around certain concepts and themes and to organise them in a way that the reader can understand [57, 58]. In the stage of identifying the legal factors by content analysis, the opinions of experts selected from different professional groups specialised in construction disputes were taken in the development of the coding framework. The expert group has a wide professional diversity, with 50% (4) academicians, 25% (2) lawyers and 25% (2) civil engineers. The first two authors of this study conducted the coding process of the content analysis and acted as coders. Other experts provided support in the development of the coding framework and made significant contributions, particularly in the selection and definition of subcategories of factors. This approach combined both practical and theoretical expertise and provided a solid basis for the analysis. In this study, similar data gathered around certain concepts and themes in judicial decisions were coded at a level that researchers and readers can understand. The coding process was structured to meet the reliability and validity standards. In ensuring validity and reliability, care was taken to analyse the case files based on the consensus of the coders. In order to increase internal validity, coding and categories were supported by the relevant literature. In order to ensure external validity, 746 decision texts (80%) were tested on the remaining 187 decision texts (20%) and the consistency of the results was observed.

3.3. Dispute Stages Identification and Statistical Analyses

Discrepancies in claims and misalignment of interests among parties in the construction sector often lead to disputes [8, 16, 59]. The process preceding litigation is defined as the "Early Stage". Interventions made during the early stages of disputes are effective in both preventing the escalation of disputes to litigation and resolving disputes that do proceed to litigation [60]. In this study, the processes of construction disputes are systematically addressed in three main stages: “Pre-Litigation A,” “Pre-Litigation B,” and “Post-Litigation.” These terms are clearly defined in this section of the study and are used consistently throughout the subsequent sections. At the early stage, parties may wish to understand the potential outcomes of all stages of litigation, both in the short and long term. Therefore, the early stage has been designed to consist of two sub-stages. "Pre-Litigation A" refers to the stage where parties learn the potential outcomes of their disputes before the First Instance Court, while "Pre-Litigation B" refers to the stage where parties learn the potential outcomes of their disputes before the final decision-making authority, the Court of Cassation. At the initial stages, when parties become aware of potential outcomes that may harm their interests in the short or long term, they are more likely to remain open to negotiation and adopt a more conciliatory attitude [17, 18]. In disputes brought to litigation, it is also important for parties to understand the potential outcomes of decisions made by the highest decision-making authority [18, 20, 21, 35-37, 39-41]. ‘Post-Litigation’ is defined as the stage where the parties learn the possible outcomes of the decision to be given by the Court of Cassation. The clear definition of the early stages forms the conceptual basis for the analyses to be conducted in the subsequent sections of the study. In addition, the decision process of the Court of Cassation, which is the highest decision-making authority, is at least as long as the Civil Court of the First Instance and the appeal process. Therefore, even if it is a late stage [38], this stage, which is popularised by the studies in the literature, is also defined in the study. Throughout this study, the terms “Pre-Litigation A,” “Pre-Litigation B,” and “Post-Litigation” will be used consistently in the same spelling without abbreviations in the text, figures, and tables. Furthermore, the term “early stage” in the text covers the Pre-Litigation A and Pre-Litigation B stages.

Legal factors appropriate to the stages under consideration were selected. To investigate the significance of the relationships among the selected legal factors, the widely recognized chi-square test method was employed. This test evaluates whether the relationship between two categorical variables is independent, thereby testing the presence of a potential association between them. Pearson [61] developed the chi-square test, which is based on comparing observed frequencies with expected frequencies to analyze relationships between categorical variables. This difference is crucial in determining whether there is a statistically significant relationship between two categorical variables in any dataset. In recent years, the chi-square test has been favored in various applications within the field of civil engineering. For instance, Olanusi and Samuel [62] used this method to analyze the impact of architectural design on church attendance. Similarly, Ihekuna and Dozie [63] discussed the applicability of the chi-square test in time series analysis. Additionally, Ayhan, Dikmen [19] emphasized the importance of the chi-square test in predicting the emergence of construction disputes, particularly in the context of the significance of relationships among categorical variables. These similar studies demonstrate that the chi-square test is a robust tool for analyzing nominal and ordinal data.

The strength of the relationship becomes important when a significant relationship between variables is identified. Lieberson [64] defines Cramer's V coefficient, proposed for evaluating the strength of significance relationships, as a measure for assessing the strength of relationships between nominal variables. Cramer's V coefficient produces values ranging from 0 to 1. The closer the value is to 1, the stronger the relationship between the variables; the closer it is to 0, the weaker the relationship [65]. Additionally, Akoglu [66] classified the strength of significance relationships as follows: very weak ($0 < x < 0.05$), weak ($0.05 < x < 0.10$), moderate ($0.10 < x < 0.15$), strong ($0.15 < x < 0.25$), and very strong ($x > 0.25$). Cramer's V coefficient evaluates the strength of the relationship between variables independently of sample size. Additionally, Somers' d coefficient is used to determine the strength and direction of the significance relationship between ordinal and nominal variables. Somers' d is particularly useful for assessing asymmetric relationships and determining how well one variable predicts another [67]. In the findings section, the significance relationships for all stages were first examined using the chi-square test, and the strength of these relationships was revealed using Cramer's V and Somers' d coefficients based on the appropriate data type. The results of the significance relationships will be compared with the "CorrelationAttributeEvaluation" feature selection method in the WEKA program to validate their accuracy.

3.4. Machine Learning Models' Implementation

For all dispute stages, machine learning models were developed using (i) tree-based algorithms and (ii) ensemble learning algorithms with legal factors that had significant relationships. In the models established for the study, the 10-fold cross-validation method was used to split the training and test data. This method divides the dataset into 10 equal parts, selecting each part as the test data and the remaining 9 parts as the training data. This process is repeated 10 times, and the average of the results obtained from each iteration is calculated [68] Thus, all data were used as both training and test data. It has been observed that classification models designed for binary classification with imbalanced datasets can end up with significant errors in predicting the minority class due to the dominance of the majority class [69]. Imbalance in the dataset can be mitigated through methods such as reducing the majority class to the level of the minority class (undersampling) or increasing the minority class to the level of the majority class by synthetic data replication (oversampling) [70]. However, synthetic data in legal texts can lead to inaccurate results in cases where context is critical [71]. Therefore, to prevent overfitting, stabilize imbalanced data, and avoid bias, the dominant class in all dispute stages was randomly equalized to the minority class using the undersampling method. This process was implemented in Python using the RandomUnderSampler. The best-performing algorithm results for the three stages, based on the balanced datasets, were shared.

In the literature, there are many algorithms used in forecasting studies. In addition, new algorithms are being developed every day. Since some algorithms are not used in decision prediction studies, they are not preferred in this study. In this study, tests were carried out with 24 different machine learning algorithms that are widely used in the literature. The algorithms used are given below.

- Tree-based algorithms are: 1. Alternating Decision Tree (AD Tree), 2. Best First Tree (BF Tree), 3. REP Tree, 4. Random Tree, 5. Simple Cart, 6. J48, 7. Random Forest, 8. Functional Tree (FT), 9. Naive Bayes Tree (NBTree), 10. Decision Stump, 11. Logistic Model Trees (LMT), 12. LAD Tree.
- Ensemble-learning algorithms are: 1. Iterative Classifier Optimizer, 2. UltraBoost, 3. RandomSubSpace, 4. MultiBoostAB, 5. RandomCommittee, 6. FilteredClassifiers, 7. AdaBoost, 8. Ensemble Selection, 9. Bagging, 10. AttributeSelectedClassifier, 11. LogitBoost, 12. ClassificationViaRegression.

Due to the large number of algorithms considered, only the LMT, NB Tree and Logit Boost algorithms that outperform the other algorithms are described. Also, each machine learning algorithm has its own parameters. Testing these algorithms with different parameter values brings a great loss in terms of both time and workload. In addition, there is no common understanding or acceptance in the literature about changing these parameters. For this reason, the parameters used in the study were tested with the default values accepted by the software.

3.4.1. LMT

LMT is a tree-based algorithm that uses the superior features of logistic regression and decision trees together. The algorithm achieves high accuracy by using the logistic regression model at each leaf node in combination with the data partitioning capability of the decision tree method, and it is seen that high accuracy is achieved when used on nominal data [72].

3.4.2. Logit Boost

Boosting is an algorithm, in which the weights of the prediction values of each classification algorithm are varied during the training process, and the prediction values of each trained classification algorithm are considered equal and averaged or majority voting is accepted [73-75]. Logit Boost was introduced as a solution to the overfitting problem in noisy datasets using the AdaBoost algorithm [76]. The algorithm is based on the proposal of increasing the weights of the data-causing overfitting by using the logistic loss function.

3.4.3. NBTree

NB Tree is a hybrid classifier algorithm obtained using decision trees with NB, representing probabilistic information in a pure approach. This contributes to shortening the computation time. The decision tree nodes of the algorithm contain univariate separation similar to decision trees. On the other hand, its leaves contain an NB classifier [77].

3.5. Evaluating the Performance of the Models

Various scientific evaluations have been performed related to the models that reach the specified target [78]. These evaluations are supplied with performance indicators such as accuracy, precision, recall, F1-score, Receiver Operating Characteristics (ROC), Precision-recall Curve (PRC), and Matthews Correlation Coefficient (MCC). In this study, to measure the performance of the models, performance information was defined using the confusion matrix presented in Table 1.

Table 1 - Structure of the confusion matrices [79].

		Predicted label	
		Label=1	Label=0
True label	Label=1	TP (True Positive)	FN (False Negative)
	Label=0	FP (False Positive)	TN (True Negative)

- Accuracy is a measure that defines the level of matching of the predictions performed by the model with the true predictions. It is calculated as the ratio of the number of correct classifications to the number of events. The accuracy equation is as follows [80]:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{1}$$

- Precision is the ratio of the number of first-class predicted true positive values to the number of first-class predicted events. The precision equation is as follows [81]:

$$Precision = \frac{TP}{(TP+FP)} \tag{2}$$

- Recall is the ratio of the number of correctly predicted true positive events to the number of positive events. The recall equation is as follows [82]:

$$Recall = \frac{TP}{(TP+FN)} \tag{3}$$

- The F1-score is determined by evaluating the precision and recall together. The F-measure is calculated by taking the harmonic mean of both components. The F-Score equation is as follows [83]:

$$F1\ Score\ = 2 * \frac{Precision * Recall}{Precision + Recal} \tag{4}$$

- Within the ROC, a table is created from the specificity and sensitivity values corresponding to different values obtained from continuous test measurements. The ROC curve, which is a probability curve, is created by intersecting these table values with the values of specificity (TP rate) on the x-axis, and sensitivity (FP rate) on the y-axis [84]. The performance evaluation of the calculated ROC value is expressed as: 0.50-0.60 is unsuccessful, 0.60-0.70 is poor, 0.70-0.80 is average, 0.80-0.90 is good, and 0.90-1.00 is excellent [85].

- PRC is a curve obtained by intersecting the precision value on the x-axis, and the recall value on the y-axis. It is generally used to check the classification performance of skewed data [86].

- MCC is a measure of the quality of classification using the complete confusion matrix data [87]. In the case of unbalanced data, potentially unreliable high prediction values may occur. Here, the MCC performance indicator is used rather than performance indicators such as Precision, Recall, and F1-Score for each class. This performance indicator is called the "Error Matrix", which allows a more reliable and accurate evaluation. It has a value range of (-1) to (+1). The closeness to (-1) means that the actual and prediction values are opposite to each other. The closeness to (+1) on the other hand means that the classification has achieved accurate success and is away from randomness [88, 89]. The equation of the MCC is as follows:

$$MCC = \frac{(TPxTN) - (FPxFN)}{\sqrt{(TP+FP)x(TPxFN)x(TN+FP)x(TN+FN)}} \tag{5}$$

4. FINDINGS

The first step in identifying the legal factors influencing the occurrence of construction disputes involved conducting an extensive literature review. As a result of the literature review, 14 legal factors were identified. These are:

1. Project Location [19, 42],
2. Project Type [20, 21, 35-41],
3. Contract Type [19-21, 35-41],
4. Construction Type [19, 23, 38, 42],
5. Project Value [19-21, 35-38, 42],
6. Project Duration [19, 38, 42],
7. The type of Plaintiff [20, 21, 35-41],

8. The type of Defendant [20, 21, 35-41],
9. Existence of counter action [20, 21],
10. Existence of a Third Party [20, 21],
11. The type of a Third Party [20, 21],
12. The type of Main Dispute [9, 18],
13. The type of Sub-dispute [9, 18],
14. The type of favor of the decision of a Court of Cassation [20, 21, 35, 39-41],

Content analysis of 933 case texts from the Turkish Court of Cassation, which included words providing information about the project, dispute, and litigation process, led to the identification of 10 legal factors. These legal factors were determined as: Existence of a joining of cases, the type of plaintiff (joining of cases), main dispute value, stage of dispute occurrence, existence of sub-dispute, sub-dispute value, existence of dissenting opinion, relevant law articles in decisions, the type of decision of a civil court of the first instance, and the type of favor of the decision of a civil court of the first instance. The coding framework used for decision texts was prepared at a level that even individuals with basic training in construction law could understand. Based on this framework, coders categorized sentences, words, and phrases from the decision texts with input from experts. Examples of these codes and categories include terms such as "300-day duration project," "dispute over unpaid work fees," and "preliminary acceptance dispute," which were coded and categorized as legal factors like project location, contract type, main dispute value, and project duration. All identified factors were converted into nominal or ordinal data, depending on the appropriate data type for each factor. For example, Project Type (Nominal: 1 - Public Investment, 2 - Private Sector Investment) and Project Duration (Ordinal: 1 - PD < 365 Days, 2 - 365 Days ≤ PD < 1095 Days, 3 - 1095 Days ≤ PD < 1825 Days, 4 - PD ≥ 1825 Days).

In light of the findings, the types of disputes were identified as follows: 233 cases (24.97%) involving “Defective and Flawed Work”, 227 cases (24.33%) involving “Increase and Decrease in the Quantity of Work”, 201 cases (21.54%) involving “Financial”, 173 cases (18.54%) involving “Duration”, 72 cases (7.72%) involving “Determination and Administrative Practices”, and 27 cases (2.89%) involving “Corruption”. It was observed that the construction projects with disputes were evenly distributed between public and private sector investments, and more than 30% of the disputed projects had a low project budget of approximately €14,271. The majority of disputes occurred before substantial completion (72.67%) and involved a single dispute (84.78%). In most disputes, the plaintiff was from the private sector (91.32%), while the defendant was almost equally distributed between public and private sectors. Additionally, the majority of disputed projects (81.99%) involved Turnkey lump sum contract, and 25% of the disputed projects had durations of less than three years.

For each defined stage, the significance relationship between the selected input legal factors and output legal factors was determined using the chi-square test. The strength of the significance relationship between nominal-nominal legal factors was analyzed using Cramer’s V coefficient, while the strength of the significance relationship between ordinal-nominal legal factors was analyzed using Somers’ d coefficient (Table 2).

Table 2 - *Chi-squared test and strenght of assosiciation results the legal factors of stage models.*

Attributes Codes	Attributes	Type of data	*p*-value			Strenght of association						Correlation Attribute Value	Selection status in the final model
						Somers'd			Cramer's V				
			Pre-Litigation A	Pre-Litigation B	Post-Litigation	Pre-Litigation A	Pre-Litigation B	Post-Litigation	Pre-Litigation A	Pre-Litigation B	Post-Litigation		
A1	Project Location	Nominal	0.316	0.320	0.320	-------	-------	-------	-------	-------	-------	0.014 +- 0.007	No
A2	**Project Type**	Nominal	**0.040**	**0.038**	**0.038**	-------	-------	-------	**0.067**	**0.068**	**0.068**	**0.067 +- 0.012**	**Yes**
A3	Contract Type	Nominal	0.354	0.301	0.301	-------	-------	-------	-------	-------	-------	0.026 +- 0.008	No
A4	Construction Type	Nominal	0.697	0.498	0.498	-------		-------	-------	-------	-------	0.022 +- 0.005	No
A5	Project Value	Ordinal	0.092	0.252	0.252	-------	-------	-------	-------	-------	-------	0.053 +- 0.008	No
A6	**Project Duration**	Ordinal	**0.015**	**0.011**	**0.011**	**0.170**	**0.175**	**0.175**	-------	-------	-------	**0.077 +- 0.006**	**Yes**
A7	**The type of Plaintiff**	Nominal	**0.000**	**0.000**	**0.000**	-------	-------	-------	**0.233**	**0.229**	**0.229**	**0.229 +- 0.010**	**Yes**
A8	**The type of Defendant**	Nominal	**0.000**	**0.000**	**0.000**	-------	-------	-------	**0.225**	**0.224**	**0.224**	**0.224 +- 0.011**	**Yes**
A9	Existence of Counter Action	Nominal	-------	-------	0.115	-------	-------	-------	-------	-------	-------	0.052 +- 0.006	No
A10	Existence of a Joining of Cases	Nominal	-------	-------	0.662	-------	-------	-------	-------	-------	-------	0.016 +- 0.011	No
A11	The type of Plaintiff (Joining of cases)	Nominal	-------	-------	0.092	-------	-------	-------	-------	-------	-------	0.024 +- 0.008	No
A12	**Existence of a Third Party**	Nominal	-------	-------	**0.000**	-------	-------	-------	-------	-------	**0.146**	**0.146 +- 0.013**	**Yes**
A13	The type of a Third Party	Nominal	-------	-------	0.537	-------	-------	-------	-------	-------	-------	0.040 +- 0.013	No
A14	**The type of Main Dispute**	Nominal	**0.000**	**0.000**	**0.000**	-------	-------	-------	**0.282**	**0.276**	**0.276**	**0.103 +- 0.005**	**Yes**
A15	Main Dispute Value	Ordinal	0.268	0.304	0.304	-------	-------	-------	-------	-------	-------	0.064 +- 0.012	No

A16	**Stage of Dispute Occurrence**	Nominal	**0.000**	**0.000**	**0.000**	-------	-------	-------	**0.150**	**0.147**	**0.147**	**0.077 +- 0.015**	**Yes**
A17	Existence of Sub-dispute	Nominal	0.117	0.135	0.135	-------	-------	-------	-------	-------	-------	0.049 +- 0.012	No
A18	The type of Sub-dispute	Nominal	0.188	0.191	0.191	-------	-------	-------	-------	-------	-------	0.041 +- 0.011	No
A19	Sub-dispute Value	Ordinal	0.229	0.152	0.152	-------	-------	-------	-------	-------	-------	0.027 +- 0.013	No
A20	Existence of Dissenting Opinion	Nominal	-------	-------	0.590	-------	-------	-------	-------	-------	-------	0.018 +- 0.011	No
A21	**1- None**	Nominal	-------	-------	**0.020**	-------	-------	-------	-------	-------	**0.076**	**0.076 +- 0.006**	**Yes**
	2- Law No. 6100	Nominal	-------	-------	**0.008**	-------	-------	-------	-------	-------	**0.087**	**0.087 +- 0.011**	**Yes**
	3- Law No. 6098	Nominal	-------	-------	**0.037**	-------	-------	-------	-------	-------	**0.068**	**0.068 +- 0.008**	**Yes**
	4- General specifications for the construction affairs	Nominal	-------	-------	**0.000**	-------	-------	-------	-------	-------	**0.183**	**0.183 +- 0.009**	**Yes**
	5- Law No. 4735	Nominal	-------	-------	0.691	-------	-------	-------	------	-------	-------	0.017 +- 0.012	No
	6- Law No. 818	Nominal	-------	-------	0.244	-------	-------	-------	-------	-------	-------	0.038 +- 0.01	No
	7- Law No. 1086	Nominal	-------	-------	**0.001**	-------	-------	-------	-------	-------	**0.105**	**0.105 +- 0.012**	**Yes**
	8- Law No. 2004	Nominal	-------	-------	**0.040**	-------	-------	-------	-------	-------	**0.067**	**0.067 +- 0.009**	**Yes**
	9- Law No. 6102	Nominal	-------	-------	0.445	-------	-------	-------	-------	-------	-------	0.025 +- 0.01	No
	10- Law No. 492	Nominal	-------	-------	0.194	-------	-------	-------	-------	-------	-------	0.042 +- 0.01	No
	11- Law No. 5302	Nominal	-------	-------	0.545	-------	-------	-------	-------	-------	-------	0.02 +- 0.012	No
	12- Law No. 3194	Nominal	-------	-------	**0.037**	-------	-------	-------	-------	-------	**0.068**	**0.068 +- 0.005**	**Yes**
	13- Other laws	Nominal	-------	-------	0.131	-------	-------	-------	-------	-------	-------	0.049 +- 0.012	No
A22	**Type of Order of a Civil Court of the First Instance**	Nominal	-------	-------	**0.000**	-------	-------	-------	-------	-------	**0.784**	**0.450 +- 0.005**	**Yes**
A23	The Type of Favor of the Decision of a Civil Court of the First Instance	Nominal	**Output (Pre-Litigation A)**										
A24	The Type of Favor of the Decision of a Court of Cassation	Nominal	**Output (Pre-Litigation B and Post-Litigation)**										

For Pre-Litigation A, 14 legal factors are selected as input and 1 legal factor is selected as output (Table 2). Chi-square test results show that among 14 input legal factors, A2, A6, A7, A8, A14, A16 are statistically significant ($p<0.05$) with the output legal factor (Table 2). Considering the significance strengths, A14 (The type of Main Dispute) (Cramer's V: 0.282) is very strong, A7 (The type of Plaintiff) (Cramer's V: 0.233) and A8 (The type of Defendant) (Cramer's V: 0.225) have strong significance, A16 (Stage of Dispute Occurrence) (Cramer's V: 0.150) has medium significance, A2 (Project Type) (Cramer's V: 0.067) and A6 (Project Duration) (Somer's d: 0.170) have weak significance (Table 2). For Pre-Litigation B, 14 legal factors were selected as input and 1 legal factor was selected as output (Table 2). When the chi-square test results are analysed, it is seen that A2, A6, A7, A8, A14, A16 legal factors are statistically significant ($p<0.05$) with the output legal factor as in Pre-Litigation A. Considering the significance strengths, A14 (The type of Main Dispute) (Cramer's V: 0.276) is very strong, A7 (The type of Plaintiff) (Cramer's V: 0.229) and A8 (The type of Defendant) (Cramer's V: 0.224) have strong significance, A16 (Stage of Dispute Occurrence) (Cramer's V: 0.147) has medium significance, A2 (Project Type) (Cramer's V: 0.068) and A6 (Project Duration) (Somer's d: 0.175) have weak significance (Table 2).

For Post-Litigation, 34 legal factors were selected as input and 1 legal factor was selected as output (Table 2). Chi-square test results show that among 34 input attributes, A2, A6, A7, A8, A12, A14, A16, A21-1,2,3,4,7,8,12 and A22 attributes are statistically significant ($p<0.05$) with the output attribute (Table 2). Considering the significance strengths, the output attribute and A22 (Type of Order of a Civil Court of the First Instance) (Cramer's V: 0.784) and A14 (The type of Main Dispute) (Cramer's V: 0.276) are very strong, A7 (The type of Plaintiff) (Cramer's V: 0.229), A8 (The type of Defendant) (Cramer's V: 0.224) and A21-4 (General specifications for the construction affairs) (Cramer's V: 0.183) strong, A16 (Stage of Dispute Occurrence) (Cramer's V: 0.147), A12 (Existence of a Third Party) (Cramer's V: 0.146) and A21-7 (Law No. 1086) (Cramer's V: 0.105) medium, A2 (Project Type) (Cramer's V: 0.068), (A21-1 None) (Cramer's V: 0.076), A21-2 (Law No. 6100) (Cramer's V: 0.087), A21-3 (Law No. 6098) (Cramer's V: 0.068), A21-8 (Law No. 2004) (Cramer's V: 0.067), A21-12 (Law No. 3194) (Cramer's V: 0.068) and A6 (Project Duration) (Somer's d: 0.175) were found to have weak significance powers (Table 2).

The "CorrelationAttributeEvaluation" feature selection method was applied in the WEKA program to the legal factors in the Post-Litigation stage, focusing on the significance relationship between input legal factors and the output legal factor. The results obtained are consistent with the significance relationships identified using the chi-square test (Table 2). In the Pre-Litigation stages, as in the Post-Litigation stage, it is clearly observed that the legal factors with significant relationships are similar. It is evident that similar legal factors consistently influence output legal factors regardless of the stages. In this context, it is observed that Pre-Litigation A and B can predict the potential outcomes of disputes at earlier stages with fewer legal factors (Table 2). Being aware of the potential outcomes of disputes at an earlier stage is more valuable for the parties. Therefore, the use of Pre-Litigation A and B offers numerous advantages, such as time, cost, and quality gains, compared to the Post-Litigation stage.

Table 2 reveals not only which attributes are significantly related to the outcomes but also their stage-sensitive effect. While party characteristics (A7-A8) and type of dispute (A14) are dominant in the early stages, trial-based attributes (A22, A21-4) become significantly more important when the dispute is brought to court (Post-Litigation) and when it is considered to be brought to court (Pre-Litigation B). However, some attributes expected to have an effect (A21-5, A21-6) are found to have no effect. Another noteworthy finding is that, in the final stage, Existence of a Third Party (A12) is moderately effective (Cramer's V: 0.146), while The type of a Third Party (A13) attribute is insignificant; that is, the model responds to the existence of a third party in the dispute, not to the identity of the third party. Similarly, attributes such as Existence of Counter Action (A9), Existence of a Joining of Cases (A10), and The type of Plaintiff (Joining of cases) (A11) are ineffective; this indicates that the presence of Counter Action/Joining of Cases/The type of Plaintiff (Joining of cases) in the case does not have a meaningful effect on the outcome. Another important finding is that the project-specific attributes Project Duration (A6) and Project Type (A2) are attributes that support prediction success in a small but consistent manner.

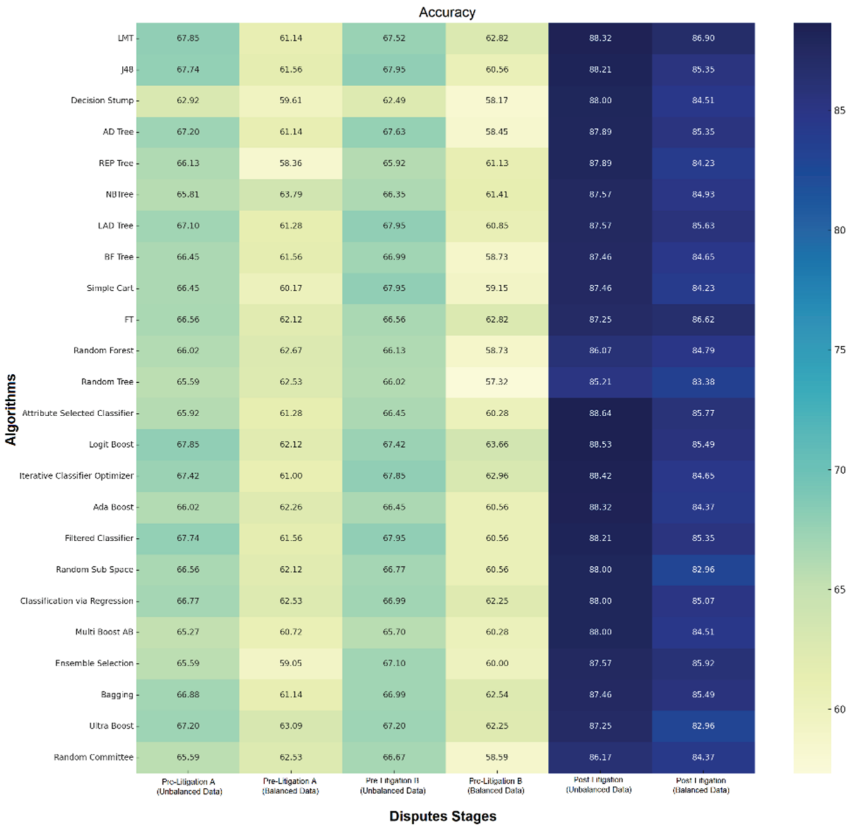

Figure 2 - Accuracy results in 24 algorithms with unbalanced and balanced data sets of all stages

The obtained data revealed an imbalanced dataset; in Pre-Litigation A, for the target legal factor A23 (The Type of Favor of the Decision of a Civil Court of the First Instance), Defendant Favor/Plaintiff Favor = 359/574, and in Pre-Litigation B and Post-Litigation, for the target legal factor A24 (The Type of Favor of the Decision of a Court of Cassation), Defendant Favor/Plaintiff Favor = 578/355. Since this imbalance would lead to overfitting, the number of dominant class data in the target legal factor was reduced to match the minority class data in all three stages using the undersampling method. The accuracy rates of 24 algorithms, tested on both imbalanced and balanced datasets across the three stages, are presented in Figure 2.

In Pre-Litigation A, Logit Boost and LMT algorithms were the most successful on the imbalanced dataset, achieving an accuracy rate of 67.85%. When the dataset was balanced, the NB Tree algorithm became the most successful, with an accuracy rate of 63.79%. In Pre-Litigation B, J48, Filtered Classifier, Simple Cart, and LAD Tree algorithms were the most successful on the imbalanced dataset, with an accuracy rate of 67.95%. When the dataset was balanced, Logit Boost was the most successful algorithm, achieving an accuracy rate of 63.66%. In the Post-Litigation stage, the Attribute Selected Classifier algorithm was the most successful on the imbalanced dataset, with an accuracy rate of 88.64%. When the dataset was balanced, the LMT algorithm became the most successful, achieving an accuracy rate of 86.90% (Figure 2).

The prediction accuracies of 24 machine learning algorithms on imbalanced and balanced (undersampled) datasets are presented in Figure 2. In Pre-Litigation A, accuracies are approximately 65–68% on unbalanced data and 59–63% on balanced data; in Pre-Litigation B, they are concentrated in the range of 66–68% on unbalanced data and 58–63% on balanced data. In the Post-Litigation phase, the accuracy level remains high, ranging from 85–88.6% for unbalanced data and 83–86.9% for balanced data. The slight decrease observed when the imbalanced dataset is converted to a balanced dataset using the undersampling technique is an expected result, as it eliminates the “predict the majority, get high accuracy” bias that arises in imbalanced data by equalizing the number of classes. In other words, training with a balanced dataset corrects the artificial high accuracy caused by imbalance; it reduces the model's tendency to stick to the majority class and provides a more realistic performance measurement, especially for the minority class. This limited decline observed with the transition to a balanced dataset reduces the risk of overfitting to the majority class by tempering the optimism in the overall accuracy metric through the equalization of prediction classes; thus, predictions, especially in the early stages, become more cautious and methodologically sound. Furthermore, the narrow differences between algorithms indicate that performance is sensitive to stage-specific information content rather than algorithm selection.

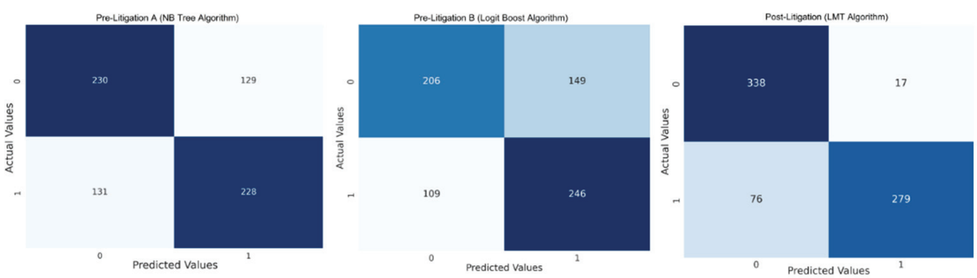

Figure 3 - Confusion matrices of algorithms showing high accuracy on the balanced data set in all stages

Focusing solely on the accuracy of algorithms can obscure class-based errors. Therefore, confusion matrices must be examined. Confusion matrices directly show which types of errors are concentrated in which class, along with the false positive/false negative balance. In a legal context, this is essential for the practical applicability and interpretability of models, as it involves presenting erroneous decision profiles with decision matrices. The confusion matrices of the algorithms with the highest accuracy performance based on a balanced dataset at all stages are shown in Figure 3. In Pre-Litigation A, the NB Tree algorithm made 230 true negative and 228 true positive predictions, providing limited support for parties to foresee the potential decision of the first-instance court. However, 131 false negatives created a risk of overlooking the rightful party's advantage, while 129 false positive predictions reduced the likelihood of settlement between parties. In Pre-Litigation B, the Logit Boost algorithm showed better performance in positive classes with 246 true positive and 206 true negative predictions. Reducing the number of false negatives to 109 increased success, while 149 false positive predictions posed a risk of unnecessary legal proceedings for the parties. In the Post-Litigation stage, the LMT algorithm demonstrated the highest performance, with 338 true negative and 279 true positive predictions. Its low false positive rate (17) and reduced false negatives (76) allowed parties to accurately predict Court of Cassation decisions and make effective strategic decisions. LMT stands out as the most reliable algorithm at this stage.

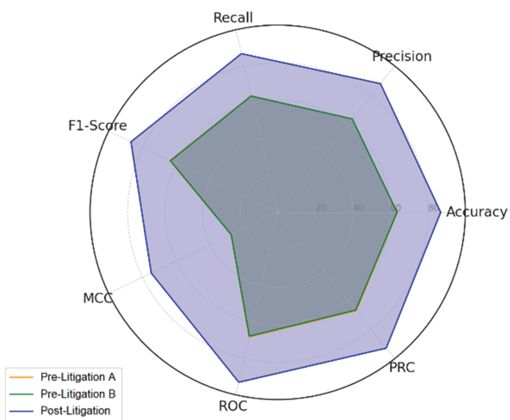

Figure 4 - Radar chart demonstration of the performance values of the algorithms with the highest accuracy in three stages with balanced datasets

The radar chart in Figure 4 shows the multi-criteria performance of algorithms with the highest accuracy across balanced datasets in three stages, without being constrained to a single accuracy value. In Pre-Litigation A, the NB Tree algorithm, which achieved the highest accuracy, recorded the following values: Precision: 0.638, Recall: 0.638, F-1 Score: 0.638, MCC: 0.276, ROC: 0.684, and PRC: 0.673. In Pre-Litigation B, the Logit Boost algorithm, which achieved the highest accuracy, recorded the following values: Precision: 0.638, Recall: 0.637, F-1 Score: 0.635, MCC: 0.275, ROC: 0.679, and PRC: 0.668. S Finally, in the Post-Litigation stage, the LMT algorithm, which achieved the highest accuracy, recorded the following values: Precision: 0.879, Recall: 0.869, F-1 Score: 0.868, MCC: 0.748, ROC: 0.931, and PRC: 0.928. In the post-litigation phase, all performance values of the LMT algorithm with the highest accuracy demonstrate a superior profile. The performance values of the most successful algorithms in pre-litigation A and B largely overlap, indicating consistent performance in the early stages.

5. DISCUSSION

The literature review indicates that previous studies primarily focused on the late stages of construction disputes. The originality of this study lies in its integrated evaluation of three distinct stages: Pre-Litigation A, Pre-Litigation B, and Post-Litigation, allowing for assessment in both early and late stages of disputes. The 86.90% accuracy obtained in the Post-Litigation stage is higher than the results of studies such as Arditi, Oksay [20], Arditi and Tokdemir [21], Chau [36], Chau [37], Chen and Hsu [38], but lower than those of Zheng [18], Ayhan, Dikmen [19], Mahfouz and Kandil [23], Arditi and Pulket [35], Pulket and Arditi [39], Pulket and Arditi [40], Arditi and Pulket [41], Chou, Tsai [42]. Some of the mentioned studies [35, 39-41] demonstrate the potential of artificial intelligence methods in predicting construction case outcomes but have certain limitations. For example, the datasets used in these studies were limited to 151 cases from the state of Illinois, raising questions about the generalizability of their results to other legal systems or broader geographical contexts. Although various methods (CBR, BDT, ACO) were compared, the methodological consistency between the accuracy rates and performance metrics of these models was not sufficiently clarified. The small dataset size may not be adequate for complex algorithms, and the reproducibility of model performance on larger and more diverse datasets needs to be tested. Although the (UPM) was presented as adaptable and extensible, it was not clarified how this flexibility would be achieved with different datasets. Additionally, while data analysis automation was achieved through WEKA, the error rates and limitations of the process were not detailed. In conclusion, while these studies offer significant innovations, improvements are needed in terms of data diversity, methodological details, and generalizability. However, even though these studies did not address the distribution of contract types, higher accuracy might have been achieved due to the use of data representing a limited region or contract types specific to a particular area.

When reviewing relevant studies, it is observed that researchers have also used the legal factors identified in the current study, including A2 (Project Type), A3 (Contract Type), A5 (Project Value), A7 (The Type of Plaintiff), A8 (The Type of Defendant), A9 (Existence of Counter Action), A12 (Existence of a Third Party), and A13 (The Type of a Third Party) [18-21, 23, 35-42]. In the current study, the fact that the A3 (Contract Type) legal factor does not show a significant relationship in all stages differs from some studies related to decision prediction [35, 39-41]. This difference may stem from the use of specific contract types in a particular region, but the distributions of contract types were not provided in these studies. Additionally, the findings indicate a significant relationship between the A5 (Project Value) and A12 (Existence of a Third Party) legal factors and the A24 (The Type of Favor of the Decision of a Court of Cassation) output legal factor. However, no significant relationship was observed between the A13 (The Type of a Third Party) and A9 (Existence of Counter Action) legal factors. This finding differs from studies predicting the formation of disputes and focusing on specific types of disputes [19-21, 35, 38, 42]. This difference may be attributed to the limited or non-existent geographical and legal coherence in data sources, the use of survey data, and a focus on specific project types in other studies.

Similarly, Chau [36], Chau [37] utilized the A3 (Contract Type), A5 (Project Value), A7 (The type of Plaintiff) and A8 (The type of Defendant) legal factors identified in the current study and employed canonical models and inductive reasoning indexing methods to reveal relationships between variables in their studies. The generalizability of the models used in these studies has been a significant point of criticism, as the datasets in both studies were limited to cases from Hong Kong, thereby restricting the applicability of the results to other regions or legal systems. Moreover, the process of selecting factors was not adequately explained, and the extent to which specific features contributed to predictive performance was not detailed. This lack of methodological transparency undermines the reproducibility of the results. While the high performance of the models raises the possibility of overfitting, it suggests that these results may be specific to the dataset and may prove inadequate for new or unique scenarios. In particular, the complex multifactorial relationships inherent in construction cases were not sufficiently addressed by the simplistic structure of the models used, limiting their explainability and practical applicability.

Although comparing the current study with one that focuses on a single dispute type may not be entirely appropriate, Mahfouz and Kandil [23], selected and utilized the A4 legal factor identified in the current study from among the attributes involved in disputes over changing site conditions. The fact that the A4 (Construction Type) legal factor does not show a significant relationship across stages differs from the study on project disputes in public-private partnership (PPP) projects [23]. This difference may be due to the influence of the infrastructure type of PPP projects on decision prediction. In the current study, which encompasses all types of disputes, the effectiveness of construction type in decision prediction diminishes. Although the study demonstrates better predictive performance, it remains limited in methodological diversity and scope. It used only three models (SVM, Naive Bayes, and Decision Trees), and their performance was not compared with a broader range of algorithms. The exclusion of widely used tree-based and ensemble methods, which are common today, limits the opportunity to improve prediction accuracy and achieve more generalizable results. Additionally, although the model performances appear high, the limited dataset (only 400 case files from 1912–2007) restricts the reproducibility and generalizability of the results in other contexts. For the mentioned study, although the RBF algorithm is ahead with its accuracy (%96.0), they obtained a higher value than the precision (%86.4). This situation opens the door for precision to come to the forefront rather than accuracy success in studies where sensitive predictions such as judgement decisions are made. The fact that the precision indicator is low despite the high accuracy success in the related study highlights the possibility that the algorithm's positive predictions are not positive enough due to the class imbalance in the data set. This issue comes to mind since there is no precise information about the distribution of the data set in the study.

In their study, Chou, Tsai [42] used the A2 (Project Type) and A4 (Construction Type) legal factors, which exhibit a high level of statistical significance in the current study, as well as the A5 (Project Value) legal factor, which does not show a significant relationship in the current study. This difference may be attributed to the limited or non-existent legal coherence in data sources, the use of survey data, and a focus on specific project types in other studies. In the study of Chou, Tsai [42], it is seen that we obtained a higher accuracy than all algorithms obtained from single and combined clustering and classification techniques. In addition, although we achieved lower accuracy than the combination of MLP, DT and classification techniques, we outperformed 7 out of 10 techniques in terms of ROC value.

The reason for the low ROC value despite the high accuracy success may be the inability of the model to discriminate between certain classes. In addition, it may be misleading to look only at the accuracy success in the case of an unbalanced data set in which the prediction classes are distributed 76%-24% and the ROC value is lower than the expected values.

Ayhan, Dikmen [19] selected the A1 (Project Location), A3 (Contract Type), A4 (Construction Type), A5 (Project Value), and A6 (Project Duration) legal factors from among the attributes in their study. Following chi-square testing, they used A1 (Project Location), A5 (Project Value), and A6 (Project Duration) legal factors in their final models. The fact that the A1 (Project Location) legal factor does not show a significant relationship in the stages of the current study differs from a study focusing on construction disputes in different countries. This difference may also stem from the limitation of our study focusing on disputes within a single country. Although the relevant study used survey data and thus may not be directly comparable with the current study, it is observed that we achieved higher accuracy and precision values in some of the models used by Ayhan, Dikmen [19] for predicting dispute occurrence.

Zheng [18] identified attributes using the case study method and utilized the A18 (The Type of Sub-dispute) and A14 (The Type of Main Dispute) legal factors from our study in their model design. The findings indicate no significant relationship between the A18 (The Type of Sub-dispute) legal factor and the output legal factor. This difference may stem from the fact that the presence of a sub-dispute does not influence the outcome as much as the presence of the main dispute. Although the study demonstrated an impressive performance with an accuracy rate of 96.42%, focusing solely on accuracy does not provide a sufficient metric for understanding the predictive capability of the model. Accuracy can be misleading for imbalanced datasets. It is suggested that further analysis of the contextual significance of performance criteria should be conducted within the study. For example, while composite metrics such as the F1 score may be high, the impact of class distribution on model performance has not been discussed in detail. The limited dataset of 171 case files used in the study underscores the need for a more comprehensive test dataset to evaluate the overall performance of the model. The fact that it deals with a specific and unique type of contract type may have contributed to its high success in judgement prediction. However, the sample of the current study includes both public sector (64.60%) and private sector (35.40%) investments. Although previous studies addressing a specific type of contract, dispute subject or region may achieve high accuracy, they may not perform as well for construction disputes from a general perspective.

Decision prediction models should not focus solely on the accuracy of predictions. In addition to the accuracy of the models' predictions, attention should also be paid to the interpretability of the predictions by the disputing parties and legal practitioners. It is also essential for users to understand the relationship between the reasons behind the model's prediction [90]. It is particularly important for parties to know the critical attributes that influence the strategic decisions they will make before taking potential disputes to formal adjudication and to be able to interpret the extent to which the decision is influenced. Considering these requirements, the current study transparently presents the reasoning behind the predictions by revealing the attribute-output relationships in all models through statistical analyses and providing the relationship levels. For example, in Pre-Litigation A and B, which are processes with dispute prevention potential, such as predicting disputes at an early stage, the type of

Main Dispute (A14) stands out as the strongest determining attribute. Furthermore, the attributes Type of Plaintiff (A7), Type of Defendant (A8), Project Type (A2), and Project Duration (A6) are seen as decisive attributes in descending order. This finding shows that disputes, claimant and defendant types are decisive in determining the court decision from the perspective of project type and project duration. The parties make their predictions with an awareness that the decisive attributes are effective in the decision. In light of all the findings, the current prediction models produce predictions that are far from being "black box" predictions, transparently presenting which attribute is effective at what level and at which stage. Thus, the outputs of models for all stages, like many studies in the literature, enable disputing parties and legal practitioners to interpret predictions not only through probability values but also through the reasoning that shapes the decision.

The study analysed construction dispute court decisions that had been referred to the official judiciary and ruled upon, using qualitative research methods, thereby identifying characteristics specific to construction disputes. The identified attributes were selected based on their significance and strength using statistical methods, and dispute stages were established. This provides construction stakeholders and legal professionals with an interpretable framework for decision support logic in early/advanced stages. Each legal system has its own unique differences. The applicability of the current study to other legal systems can be explained by the originality potential and replicability of the study method. Generalising the study to different legal systems can be achieved by applying the study method itself. The analysis of dispute texts in different legal systems by experts using qualitative research methods will result in the creation of a universe of attributes specific to each legal system and influencing the decision, revealing statistically significant relationships and strengths with output attributes. Prediction models can be developed by determining stages according to the needs of the relevant legal system with attributes that have a meaningful relationship.

The study analysed construction dispute court decisions that had been referred to the official judiciary and ruled upon, using qualitative research methods, thereby identifying characteristics specific to construction disputes. The identified attributes were selected based on their significance and strength using statistical methods, and dispute stages were established. This provides construction stakeholders and legal professionals with an interpretable framework for decision support logic in early/advanced stages. Each legal system has its own unique differences. The applicability of the current study to other legal systems can be explained by the originality potential and replicability of the study method. Generalising the study to different legal systems can be achieved by applying the study method itself. The analysis of dispute texts in different legal systems by experts using qualitative research methods will result in the creation of a universe of attributes specific to each legal system and influencing the decision, revealing statistically significant relationships and strengths with output attributes. Prediction models can be developed by determining stages according to the needs of the relevant legal system with attributes that have a meaningful relationship.

It is of great importance to examine the use of prediction models developed in this study for three different stages of dispute resolution in the construction industry and by stakeholders in the construction sector. It is highly valuable for academic studies to produce concrete outputs for the sector, i.e. to be patentable or commercialisable. To understand the potential of the current study in this regard, it would be beneficial to look at the development of artificial intelligence-supported systems in the field of law. This process, which began in the 1990s in America and Europe with the modelling of court outcomes and the analysis of case law, has undergone rapid transformation and progress over the years. Although artificial intelligence-supported legal systems, which have become an indispensable part of corporate law firms, have received the desired attention in America and Europe, they are still undiscovered in Türkiye. Thanks to rapid developments in artificial intelligence in recent years, these legal systems have become faster, more controllable, and more scalable than the current legal bureaucracy, capable of automatically analysing court decisions and providing legal insights. Preparing claims and defence texts during the trial process is a long and laborious process for the parties involved. Relativity AI, which quickly scans thousands of documents and identifies key evidence in a case, marking the relevant sections, was valued at $3.6 billion as of 2021. It has over 300,000 users in 49 countries [91]. Pre/Dicta, used by many large law firms in the United States, predicts how a judge will rule with 86% accuracy using 120 attributes [92]. Theo AI, which attracted $4.2 million in seed funding in May 2025 following a $2.2 million pre-seed investment round in November 2024, is also noteworthy for its system that predicts court outcomes or whether a settlement will be reached based on past case data [93]. Artificial intelligence-supported systems in the field of law have been developing rapidly in recent years and have become indispensable in the international arena. Current systems focus on resolving disputes in the construction sector that have been brought before the official courts in favour of the parties. Unlike existing systems, the current study enables parties to prevent and resolve disputes without going to formal court (Pre-Litigation A and B). It also enables the shortening of the lengthy and costly process by presenting the outcome of the trial to the parties during the formal trial process (Post-Litigation). The commercialisation of the proposed models is possible through the expansion of past case data, the attainment of reasoned decisions, and the extraction of judge profiles.

6. CONCLUSION

In this study, a comprehensive literature review was conducted to identify the legal factors influencing construction disputes, resulting in the identification of 14 legal factors. Additionally, 933 decisions from the Turkish Court of Cassation on construction disputes between 2011 and 2021 were analyzed using content analysis, leading to the identification of 10 more legal factors. As a result, a total of 24 legal factors were determined through the literature review and analyses. To enhance internal validity in content analysis, coding and categories were supported by the relevant literature. To ensure external validity, the study's findings (746 decision texts) were tested on a separate set of Court of Cassation decisions (187 decision texts), and consistency was observed in the results. Three distinct dispute stages were defined based on whether construction disputes were brought to court. These stages are as follows: (i) Pre-Litigation A: The stage where the potential decision of the Civil Court of the First Instance is predicted before the dispute is brought to litigation, (ii) Pre-Litigation B: The stage where the potential decision of the Court of Cassation is predicted before the dispute is brought to litigation, and (iii) Post-Litigation: The stage where the dispute is brought to litigation, and the decision of the highest court, the Court of Cassation, is predicted. Appropriate legal factors were selected for the defined stages. The legal factors used in the predictive models were determined by evaluating the significance relationships and their strength between input and target legal factors at each stage. A total of 24 classification algorithms from tree-based and ensemble algorithm families were considered for building predictive models. To prevent overfitting, the 10-fold cross-validation method was first used to split the training and test data. Subsequently, the algorithms were implemented first with imbalanced datasets and then with balanced datasets. For the imbalanced dataset, the best-performing algorithms were: Logit Boost with 67.85% prediction accuracy in Pre-Litigation A, J48 with 67.95% accuracy in Pre-Litigation B, and AttributeSelectedClassifier with 88.64% accuracy in Post-Litigation. For the balanced dataset, the best-performing algorithms were: NB Tree with 63.79% accuracy in Pre-Litigation A, Logit Boost with 63.66% accuracy in Pre-Litigation B, and LMT with 86.90% accuracy in Post-Litigation.

This study differs from previous research in several ways: (i) developing three distinct models for different stages, (ii) having a diverse and up-to-date set of machine learning algorithms, (iii) using nationwide data instead of limited geographically or legally regional data, (iv) basing the models directly on court decisions, (v) using datasets with extensive project, contract, and dispute types, (vi) employing 10-fold cross-validation and undersampling techniques to mitigate the risk of overfitting in binary classification problems, (vii) using the content analysis method, which allows for systematic and objective examination of datasets and produces more consistent and reproducible results, especially with large datasets, and (viii) considering a greater number of legal factors compared to previous studies. The originality of this study lies in the conceptualisation of the attribute universe specific to construction disputes through the analysis of construction dispute court decisions that have been referred to official courts and ruled upon using qualitative research methods, and the selection of these attributes based on their validity (Chi-Square, Cramer's V, Somers' d) and defining stages appropriate to disputes (Pre-Litigation A/B and Post-Litigation), and providing construction stakeholders and legal professionals with an interpretable framework for decision support logic in early/advanced stages. Due to the three-stage decision support framework being the focus of the study, it is intended to produce a baseline that prioritises comparability, reproducibility, and legal interpretability by running 24 machine learning algorithms with their default hyperparameters. Therefore, running the 24 algorithms used in the prediction of the stages with their default hyperparameters is a conscious design choice. Given the scope of the study, the focus of its originality lies beyond the hyperparameter optimisation of the algorithms. In this respect, this priority can be considered one of the study's limitations. On the other hand, the study has certain limitations: (I) it focuses solely on a single country, (II) the inability to access detailed judicial processes due to various reasons (in Türkiye, only reasoned judicial decisions are allowed for analysis), and (III) the lack of a standardized template for writing Court of Cassation decisions, which prevents the identified and selected legal factors from being consistently present in every case file (this limitation may not apply in other legal systems).

The findings of this study provide significant insights for professionals by highlighting the legal factors influencing the occurrence of disputes in the construction sector and presenting predictive models that estimate the potential outcomes of disputes at early and late stages. These results underscore their critical role in reducing the costs associated with construction disputes and improving decision-making processes in the management of construction projects. Additionally, the models developed in this study provide a generalizable framework for researchers and practitioners aiming to improve dispute resolution processes in litigation across different legal and organizational contexts. In 71% of the reasoned decisions in the dataset (659 decisions), references to legal statutes are present. The decisions used in this study were evaluated by the Court of Cassation judges within the framework of the Turkish Code of Obligations No. 6098, which is based on the Swiss Code of Obligations and defines the duties and responsibilities of all stakeholders in construction projects, along with other relevant laws. The Turkish Code of Obligations No. 6098 (122 decisions) and the repealed former Code of Obligations No. 818 (372 decisions) were the most frequently cited statutes. By applying the original methodology steps used in this study, similar studies can be encouraged in countries like Switzerland, Germany, and France, which share the same or similar legal parameters. With the identification and selection of appropriate legal factors and model development for different legal systems, the results obtained can be compared with those of this study to observe similarities and differences. Future studies may achieve better results by using current machine learning algorithms not employed in this study. Furthermore, high prediction accuracy can be achieved by optimising the hyperparameters of the algorithms used in the predictions. Different models can also be developed to predict outcomes at earlier stages. The dataset can be expanded to include decisions from multiple countries with similar legal infrastructures. NLP-based decision prediction studies have become prominent in recent years, especially with legal data. Such studies may focus on broader compilations involving decisions from diverse legal roles or specifically on construction disputes using NLP methods. Legal texts often face limited and imbalanced datasets. To avoid the risk of overfitting and to prevent synthetic data generation, undersampling was applied in this study. For future studies, oversampling can be employed to balance the dataset and increase data quantity, allowing the development of predictive models whose results can be compared with those of this study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data and WEKA files will be made available on reasonable request.

Acknowledgment

This work, created from the Ph.D. thesis of Mahmut SARI, was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) through the Innovative Solutions Research Projects Support Program in Social Sciences and Humanities (3005) under grant 122G126.

References

[1] Giang, D.T.H. and L. Sui Pheng, Role of construction in economic development: Review of key concepts in the past 40 years. Habitat International, 2011. 35(1): p. 118-125.

[2] Ringen, K. and A. Englund, The Construction Industry. Annals of the New York Academy of Sciences, 2006. 1076(1): p. 388-393.

[3] Iliev, B.Z., World Construction Market. Review of Business and Economics Studies, 2019. 7: p. 32-36.

[4] Lean, C.S., Empirical tests to discern linkages between construction and other economic sectors in Singapore. Construction Management and Economics, 2001. 19(4): p. 355-363.

[5] Jaber, F.K., N.A. Jasim, and F.M. Al-Zwainy, Forecasting techniques in construction industry: earned value indicators and performance models. Scientific Review Engineering and Environmental Sciences (SREES), 2020. 29(2): p. 234-243.

[6] Li, J., D. Greenwood, and M. Kassem, Blockchain in the built environment and construction industry: A systematic review, conceptual models and practical use cases. Automation in construction, 2019. 102: p. 288-307.

[7] Mengistu, D.G. and G. Mahesh, Challenges in developing the Ethiopian construction industry. African Journal of Science, Technology, Innovation and Development, 2020. 12(4): p. 373-384.

[8] Fenn, P., D. Lowe, and C. Speck, Conflict and dispute in construction. Construction Management and Economics, 1997. 15(6): p. 513-518.

[9] Sarı, M., B. Sayın, and C. Akçay, Classification and resolution procedure for disputes in public construction projects. Revista de la Construcción. Journal of Construction, 2021. 20(2): p. 259-276.

[10] Dobrucali, E., et al., Exploring the Impact of COVID-19 on the United States Construction Industry: Challenges and Opportunities. IEEE Transactions on Engineering Management, 2022: p. 1-13.

[11] Sarı, M., Savaş Bayram, and E. Aydemir, Construction-Related Disputes: A Comprehensive Bibliometric Investigation in The 8th International Project and Construction Management Conference (IPCMC 2024), S.U. Kerim Koç, Serkan Kıvrak, Editor. 2024: İstanbul. p. 507-518.

[12] Cheung, S.O. and T.W. Yiu, Are Construction Disputes Inevitable? IEEE Transactions on Engineering Management, 2006. 53(3): p. 456-470.

[13] Arcadis, C., Global Construction Disputes Report 2020, in Global Construction Disputes Report. 2020, Arcadis Company

[14] Arcadis, C., Global Construction Disputes Report 2022, in Global Construction Disputes Report 2022.

[15] Tanriverdi, C., et al., Causal mapping to explore emergence of construction disputes. Journal of Civil Engineering and Management, 2021. 27(5): p. 288-302.

[16] Kumaraswamy, M.M., Conflicts, claims and disputes in construction. Engineering, Construction and Architectural Management, 1997. 4(2): p. 95-111.

[17] Heue, L. and S. Penrod, Predicting the Outcomes of Disputes: Consequences for Disputant Reactions to Procedures and Outcomes. Journal of Applied Social Psychology, 1994. 24(3): p. 260-283.

[18] Zheng, X., Liu, Y., Jiang, J., Thomas, L.M., Su, N., Predicting The Litigation Outcome of PPP Project Disputes Between Public Authority and Private Partner Using an Ensemble Model. Journal of Business Economics and Management, 2021. 22: p. 320-345.

[19] Ayhan, M., I. Dikmen, and M.T. Birgönül, Predicting the Occurrence of Construction Disputes Using Machine Learning Techniques. Journal of Construction Engineering and Management, 2021. 147(4).

[20] Arditi, D., F.E. Oksay, and O.B. Tokdemir, Predicting the Outcome of Construction Litigation Using Neural Networks. Computer-Aided Civil and Infrastructure Engineering, 1998. 13(2): p. 75-81.

[21] Arditi, D. and O.B. Tokdemir, Using Case-Based Reasoning to Predict the Outcome of Construction Litigation. Computer-Aided Civil and Infrastructure Engineering, 1999. 14(6): p. 385-393.

[22] Mahfouz, T. and A. Kandil, Factors Affecting Litigation Outcomes of Differing Site Conditions (DSC) Disputes: A Logistic Regression Models (LRM), in Building a Sustainable Future. 2009. p. 239-248.

[23] Mahfouz, T. and A. Kandil, Litigation Outcome Prediction of Differing Site Condition Disputes through Machine Learning Models. Journal of Computing in Civil Engineering, 2012. 26: p. 298-308.

[24] Chaphalkar, N.B., K.C. Iyer, and S.K. Patil, Prediction of outcome of construction dispute claims using multilayer perceptron neural network model. International Journal of Project Management, 2015. 33(8): p. 1827-1835.

[25] Treacy, T.B., Use of Alternative Dispute Resolution in the Construction Industry. Journal of Management in Engineering, 1995. 11(1): p. 58-63.

[26] Justice, T.M.o., General Directorate of Judicial Records and Statistics. 2023.

[27] Gill, A., et al., Comparison of the effects of litigation and ADR in South-East Queensland. International Journal of Construction Management, 2015. 15(3): p. 254-263.

[28] Parikh, D., G.J. Joshi, and D.A. Patel, Development of Prediction Models for Claim Cause Analyses in Highway Projects. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 2019. 11(4): p. 04519018.

[29] Hall, M.A. and R.F. Wright, Systematic Content Analysis of Judicial Opinions. Calif. L. Rev., 2008. 96: p. 63-63.

[30] Do, S.T., V.T. Nguyen, and N.H. Nguyen, Relationship networks between variation orders and claims/disputes causes on construction project performance and stakeholder performance. Engineering, Construction and Architectural Management, 2022.

[31] Francis, M., T. Ramachandra, and S. Perera, Disputes in Construction Projects: A Perspective of Project Characteristics. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 2022. 14(2).

[32] Echternach--Jaubert, M., R. Pellerin, and L. Joblot, Litigation management process in construction industry. Procedia Computer Science, 2021. 181: p. 678-684.

[33] Tazelaar, F. and C. Snijders, Dispute resolution and litigation in the construction industry. Evidence on conflicts and conflict resolution in The Netherlands and Germany. Journal of Purchasing and Supply Management, 2010. 16(4): p. 221-229.

[34] Haugen, T. and A. Singh, Dispute Resolution Strategy Selection. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 2015. 7(3).

[35] Arditi, D. and T. Pulket, Predicting the Outcome of Construction Litigation Using Boosted Decision Trees. Journal of Computing in Civil Engineering, 2005. 19(4): p. 387-393.

[36] Chau, K.W. Predicting Construction Litigation Outcome Using Particle Swarm Optimization. in Innovations in Applied Artificial Intelligence. 2005. Springer Berlin Heidelberg.

[37] Chau, K.W. Prediction of Construction Litigation Outcome – A Case-Based Reasoning Approach. in Advances in Applied Artificial Intelligence. 2006. Springer Berlin Heidelberg.

[38] Chen, J.-H. and S.C. Hsu, Hybrid ANN-CBR model for disputed change orders in construction projects. Automation in Construction, 2007. 17(1): p. 56-64.

[39] Pulket, T. and D. Arditi, Construction litigation prediction system using ant colony optimization. Construction Management and Economics, 2009. 27(3): p. 241-251.

[40] Pulket, T. and D. Arditi, Universal Prediction Model for Construction Litigation. Journal of Computing in Civil Engineering, 2009. 23(3): p. 178-187.

[41] Arditi, D. and T. Pulket, Predicting the Outcome of Construction Litigation Using an Integrated Artificial Intelligence Model. Journal of Computing in Civil Engineering, 2010. 24(1): p. 73-80.

[42] Chou, J.-S., C. Tsai, and Y. Lu, Project Dispute Prediction by Hybrid Machine Learning Techniques. Journal of Civil Engineering and Management, 2013. 19(4): p. 505-517.

[43] Hyett, N., A. Kenny Dr, and V. Dickson-Swift Dr, Methodology or method? A critical review of qualitative case study reports. International Journal of Qualitative Studies on Health and Well-being, 2014. 9(1): p. 23606.

[44] Sarı, M. and S. Bayram, İnşaat Anlaşmazlıklarının Önlenmesi ve Çözümünde Etkili Faktörler: Erken Aşamalarda ve Resmi Yargı Sürecindeki Rolü. Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi, 2025. 41(1): p. 133-142.

[45] Sarı, M. and S. Bayram. From courts to boards: Comparative legal analysis of early-stage dispute management in public construction disputes. in 4th International Civil Engineering & Architecture Conference (ICEARC'25). 2025. Trabzon, Türkiye.

[46] Cevikbas, M., Identification of dispute sources in the construction industry via court files. Turkish Journal of Civil Engineering, 2023. 34(2): p. 57-76.

[47] Künkcü, H., et al., Operational barriers against the use of smart contracts in construction projects. Turkish Journal of Civil Engineering, 2023. 34(5): p. 81-106.

[48] Koc, K. and A.P. Gurgun, Causal relationships of readability risks in construction contracts. Teknik Dergi, 2022. 33(2): p. 11823-11846.

[49] ÇAKMAK, P.I., Causes of disputes in the Turkish construction industry: Case of public sector projects. Istanbul Technical University, AZ, 2016. 13(3): p. 109-118.

[50] Sarı, M., S. Bayram, and E. Aydemir, When defendants speak: Quantifying the predictive value of defence arguments in construction litigation. Journal of Construction Engineering, Management & Innovation, 2025. 8(1): p. 64-88.

[51] Finfgeld-Connett, D., Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. Qualitative Research, 2014. 14(3): p. 341-352.

[52] Dumay, J. and L. Cai, A review and critique of content analysis as a methodology for inquiring into IC disclosure. Journal of Intellectual Capital, 2014. 15(2): p. 264-290.

[53] Hukukturk. Database. 2024 [cited 2024 22.03.2024]; Available from: https://www.hukukturk.com/yargitay-kararlari.

[54] Court of Cassation, Average Case Processing Time in 2023. 2023, Court of Cassation. [cited 2023]; Available from: https://www.yargitay.gov.tr/item/51/istatistikler

[55] Holsti, O.R., Content analysis for the social sciences and humanities. First ed. Reading. MA: Addison-Wesley (content analysis). 1969.

[56] Smith, H., Strategies of social research: The methodological imagination (Prentice-Hall methods of social science series). 1975: Prentice-Hall. 509.

[57] Yıldırım, A. and H. Simsek, Sosyal bilimlerde nitel araştırma yöntemleri (11 baski: 1999-2018). 1999.

[58] Falkingham, L.T. and R. Reeves, Context analysis—A technique for analysing research in a field, applied to literature on the management of R&D at the section level. Scientometrics, 1998. 42(2): p. 97-120.

[59] Kisi, K., R. Kayastha, and Y. Chitrakar, Construction Claims and Payment Disputes Analysis: Alternative Dispute Resolution to Litigation. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 2023. 15(1): p. 06522005.

[60] Stipanowich, T.J., Managing construction conflict: unfinished revolution, continuing evolution. Constr. Law., 2014. 34: p. 13.

[61] Pearson, K., X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1900. 50(302): p. 157-175.

[62] Olanusi, J. and S. Samuel, Application of Chi-Square Test to Determine Architectural Impact on Church Patronage. International Journal of Research and Innovation in Social Science, 2023. 7(6): p. 605-616.

[63] Ihekuna, S.O. and K.C. Dozie, Chi-square test in time series data. Asian Journal of Probability and Statistics, 2022. 20(3): p. 49-63.

[64] Lieberson, S., Measuring population diversity. American Sociological Review, 1969: p. 850-862.

[65] Agresti, A., Categorical data analysis. 3rd Edition ed. Vol. 792. 2013: John Wiley & Sons. 742.

[66] Akoglu, H., User's guide to correlation coefficients. Turkish Journal of Emergency Medicine, 2018. 18(3): p. 91-93.

[67] Somers, R.H., A new asymmetric measure of association for ordinal variables. American sociological review, 1962: p. 799-811.

[68] Raschka, S., Model evaluation, model selection, and algorithm selection in machine learning. arXiv.org preprint arXiv:1811.12808, 2018.

[69] He, H. and E.A. Garcia, Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, 2009. 21(9): p. 1263-1284.

[70] Kaur, H., H.S. Pannu, and A.K. Malhi, A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions. ACM Comput. Surv., 2019. 52(4): p. Article 79.

[71] Assunção, G.O., R. Izbicki, and M.O. Prates, Is Augmentation Effective in Improving Prediction in Imbalanced Datasets? Journal of Data Science, 2024: p. 1-16.

[72] Landwehr, N., M. Hall, and E. Frank, Logistic model trees. Machine learning, 2005. 59: p. 161-205.

[73] Sutton, C.D., Classification and regression trees, bagging, and boosting. Handbook of Statistics, 2005. 24: p. 303-329.

[74] Witten, I., E. Frank, and M. Hall, Chapter 8 - Ensemble Learning, in Data Mining: Practical Machine Learning Tools and Techniques (Third Edition), I.H. Witten, E. Frank, and M.A. Hall, Editors. 2011, Morgan Kaufmann: Boston. p. 351-373.

[75] Alpaydin, E., Introduction to machine learning. Fourth ed. 2020: MIT Press.

[76] Friedman, J., T. Hastie, and R. Tibshirani, Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The Annals of Statistics, 2000. 28(2): p. 337-407.

[77] Kohavi, R. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996. Portland, Oregon: AAAI Press.

[78] Uncuoglu, E., et al., Comparison of neural network, Gaussian regression, support vector machine, long short-term memory, multi-gene genetic programming, and M5 Trees methods for solving civil engineering problems. Applied Soft Computing, 2022. 129.

[79] Ting, K.M., Confusion Matrix, in Encyclopedia of Machine Learning, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 209.

[80] Kakas, A., Accuracy, in Encyclopedia of Machine Learning, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 9-10.

[81] Ting, K.M., Precision, in Encyclopedia of Machine Learning, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 780.

[82] Ting, K.M., Precision and Recall, in Encyclopedia of Machine Learning, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 781.

[83] Goutte, C. and E. Gaussier. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. in Advances in Information Retrieval: 27th European Conference on Information Retrieval Research. 2005. Santiago de Compostela, Spain: Springer.

[84] Hoo, Z.H., J. Candlish, and D. Teare, What is an ROC curve? Emergency Medicine Journal, 2017. 34(6): p. 357-359.

[85] Kumar, N. and S. Khatri. Implementing WEKA for medical data classification and early disease prediction. in 3rd International Conference On Computational Intelligence & Communication Technology. 2017. IEEE.

[86] Davis, J. and M. Goadrich. The relationship between Precision-Recall and ROC curves. in 23rd International Conference On Machine Learning. 2006.

[87] Matthews, B.W., Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta, 1975. 405(2): p. 442-451.

[88] Liu, Y., et al., Research on the Matthews correlation coefficients metrics of personalized recommendation algorithm evaluation. International Journal of Hybrid Information Technology, 2015. 8(1): p. 163-172.

[89] Bulut, F., Örnek tabanlı sınıflandırıcı topluluklarıyla yeni bir klinik karar destek sistemi. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 2017. 32(1).

[90] Collenette, J., K. Atkinson, and T. Bench-Capon, Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights. Artificial Intelligence, 2023. 317: p. 103861.

[91] BuiltInChicago. Relativity Reaches $3.6B Valuation After Recent Investment From Silver Lake. 2021 [cited 2025; Available from: https://www.builtinchicago.org/articles/relativity-silver-lake-funding-3b-valuation?utm_source=chatgpt.com.

[92] Evans, D.L., et al., Dispute Resolution Enhanced: How Arbitrators and Mediators Can Harness Generative AI. Dispute Resolution Journal, 2024. 78(1).

[93] Moran, H. Theo Ai Secures 4.2MM Seed Round to Advance AI-Powered Settlement Prediction for Big Law. 2025 [cited 2025; Available from: https://legalfundingjournal.com/theo-ai-secures-4-2mm-seed-round-to-advance-ai-powered-settlement-prediction-for-big-law/.