OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 03 - Issue 03 (2026)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR LEGAL CASE PREDICTION

Aamir Shakil[1], Khalid Bin Muhammad[2], Muhammad Asim Shahid[3]

[1, 2, 3]Ziauddin University, Pakistan.

[*]Corresponding Author: (khalid.muhammad@zu.edu.pk)

DOI:(*https://doi.org/10.71146/kjmr920*)

Article Info

Abstract

Predicting legal case outcomes is a complex task due to the influence of multiple interrelated factors, including financial conditions, institutional characteristics, and legal dynamics. With the increasing availability of structured legal data, machine learning techniques provide an effective approach for supporting data-driven decision-making in this domain.

This study evaluates the performance of four machine learning models—Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest—for predicting legal case outcomes. The dataset consists of structured features such as liability scores, financial indicators, and organizational attributes. Data preprocessing techniques, including handling missing values, feature selection, and one-hot encoding, were applied to prepare the dataset. The data was split using stratified sampling, and model performance was validated through 5-fold cross-validation.

The models were evaluated using accuracy, precision, recall, and F1-score. Results indicate that Random Forest achieved the best overall performance, while Logistic Regression showed comparatively lower effectiveness. The findings highlight the importance of model selection and demonstrate the effectiveness of ensemble methods for structured legal data prediction.

Keywords:
legal case, machine learning, analysis

Introduction

Background

Predicting the outcome of legal cases is a complex task influenced by multiple interconnected factors, including legal strength, financial conditions, institutional capacity, and broader socio-economic dynamics. Traditionally, such predictions rely on legal experts' experience, case precedents, and detailed interpretation of facts. However, the increasing availability of structured legal data, combined with advancements in machine learning, has created new opportunities for developing data-driven approaches to support decision-making in the legal domain. Machine learning techniques enable the analysis of large datasets and the identification of patterns that may not be easily captured through manual evaluation.

Problem Statement

Despite these advancements, accurately predicting legal case outcomes remains challenging. The diversity of influencing factors and the overlap between outcome categories make it difficult for traditional approaches and even some computational models to achieve consistent performance. Additionally, different machine learning algorithms interpret data patterns differently, leading to variations in predictive accuracy. This creates a need for systematic evaluation of multiple models using structured legal datasets.

Research Objectives and Question

The primary objective of this study is to evaluate and compare the performance of multiple machine learning models in predicting legal case outcomes. Specifically, the study aims to:

- Assess the effectiveness of Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest
- Compare model performance using accuracy, precision, recall, and F1-score
- Identify the most suitable model for structured legal data

This leads to the central research question:

Which machine learning model provides the most accurate and reliable predictions for legal case outcomes using structured data?

Thesis Statement

This study hypothesizes that ensemble-based methods, particularly Random Forest, will outperform other models due to their ability to capture complex, non-linear relationships within structured datasets.

Scope of the Study

This research focuses on structured legal datasets containing features such as liability scores, financial indicators, organizational attributes, and market-related factors. The analysis is limited to four machine learning models—Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest—and evaluates their performance using standard classification metrics.

To better understand existing approaches and identify research gaps, the following section reviews relevant literature on machine learning applications in legal case prediction.

Literature Review

The application of machine learning in the legal domain has gained significant attention, particularly in prediction and decision-support systems. Researchers have explored various approaches to analyze historical legal data and identify patterns influencing judicial outcomes, aiming to enhance traditional legal reasoning with data-driven insights.

One major area of research focuses on **text-based legal prediction models**. Aletras et al. (2016) demonstrated that decisions of the European Court of Human Rights could be predicted using machine learning models trained on textual features extracted from legal documents. Similarly, Chalkidis et al. (2019) applied deep learning techniques to legal texts and achieved improved predictive performance. However, these approaches rely heavily on unstructured data, requiring extensive preprocessing and large datasets, which may limit their practical applicability.

Another important direction involves **structured data and metadata-based prediction**. Katz et al. (2017) used historical case data and metadata to predict decisions of the United States Supreme Court, highlighting the effectiveness of machine learning models, particularly ensemble approaches. Nevertheless, their findings also indicate that relying solely on historical data may not fully capture external and contextual influences affecting legal outcomes.

From a methodological perspective, several studies have compared different machine learning algorithms. Liu and Chen (2017) showed that model performance varies depending on dataset characteristics and feature representation, emphasizing that no single model consistently outperforms others. In structured data contexts, models such as Random Forest (Breiman, 2001) have proven effective due to their ability to capture non-linear relationships. Advanced techniques like XGBoost (Chen & Guestrin, 2016) also demonstrate strong performance, although their complexity may not always be necessary.

Despite these advancements, a clear gap remains in systematically evaluating multiple machine learning models on structured legal datasets that incorporate diverse features such as financial indicators, organizational attributes, and legal factors. Many existing studies focus either on textual data or on a limited range of models, leaving the comparative performance of models on structured data insufficiently explored.

To address this gap, the present study conducts a comparative analysis of Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest using structured legal data. This approach aims

to identify the most effective model for predicting legal case outcomes and to provide insights into their relative strengths and limitations.

Methodology

This study adopts a systematic machine learning approach to predict legal case outcomes using a structured dataset consisting of financial, organizational, and legal-related features. The methodology includes data preprocessing, feature transformation, model training, validation, and performance evaluation.

Research Workflow

The overall process followed in this study is illustrated below:

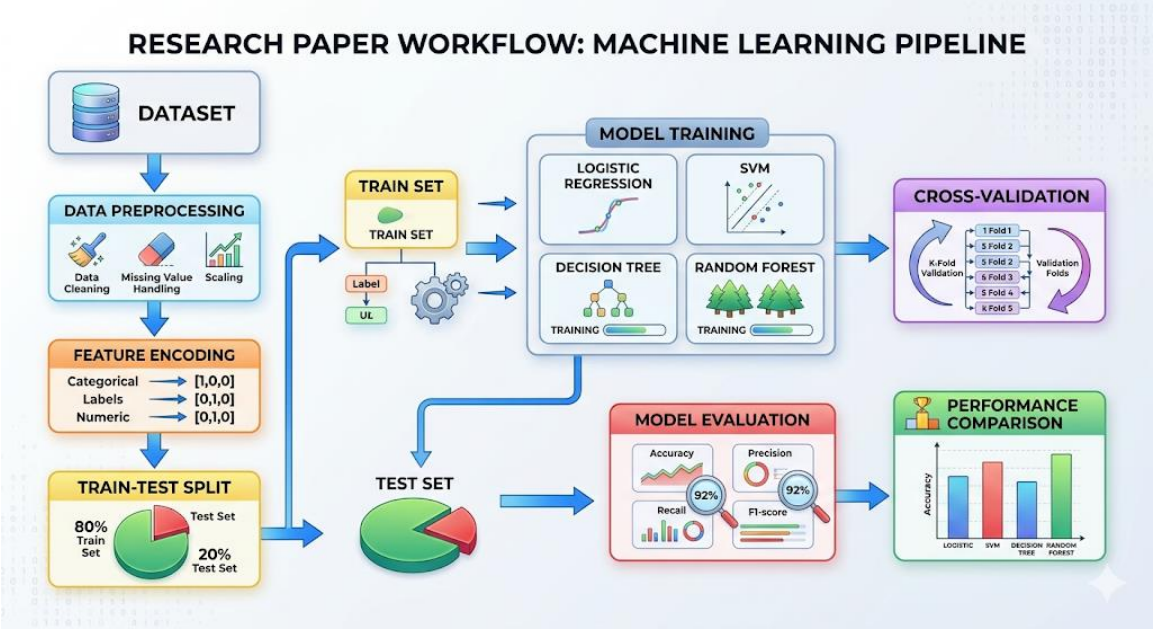


Figure 1: Machine learning workflow for predicting legal case outcomes. The diagram illustrates the complete pipeline including data preprocessing, feature encoding, model training, cross-validation, and performance evaluation.

Data Collection and Description

The dataset used in this research consists of structured legal case records containing multiple attributes that may influence case outcomes. These include financial indicators, organizational performance factors, and external environmental variables. Key features include liability score, financial indicators, market potential, institutional size, and team-related characteristics such as collaboration and training.

The target variable is the legal case outcome, which represents the final decision of each case.

Data Preprocessing

To ensure data quality and model reliability, several preprocessing steps were applied. Irrelevant and leakage-prone features, such as case identifiers and prediction-related variables, were removed to prevent bias during model training. Missing values were handled by removing incomplete records to maintain dataset consistency, as imputing such values could introduce noise in prediction outcomes.

Categorical variables were transformed into numerical representations using one-hot encoding, enabling machine learning models to process categorical information effectively. Additionally, feature scaling was applied using standardization to normalize numerical variables, ensuring that models sensitive to feature magnitude, such as Support Vector Machine, perform optimally.

Train-Test Split

The dataset was divided into training and testing sets using an 80:20 ratio. Stratified sampling was applied to maintain the distribution of outcome classes across both sets. The training set was used to train the models, while the testing set was used for evaluating model performance on unseen data.

Model Selection and Training

Four machine learning models were selected to represent different learning paradigms:

- Logistic Regression (linear model)
- Support Vector Machine (margin-based classifier)
- Decision Tree (rule-based model)
- Random Forest (ensemble model)

This selection enables comparison between simple and complex models, as well as between linear and non-linear approaches. A pipeline-based implementation was used to integrate preprocessing and model training steps, ensuring consistency across all models. Standardization was applied where necessary to improve model convergence and performance, particularly for distance-based algorithms such as Support Vector Machine.

To ensure fair comparison and optimal performance, hyper parameter tuning was incorporated using grid search with cross-validation. Instead of relying on default configurations, each model was trained using its best-performing hyper parameters identified through the tuning process. This approach allows the models to operate under optimized conditions, providing a more reliable evaluation of their true predictive capabilities. In particular, ensemble methods such as Random Forest are expected to benefit from this process due to their ability to capture complex feature interactions and reduce overfitting.

Hyper parameter Tuning

Hyper parameter tuning was conducted to optimize the performance of machine learning models on the structured legal case dataset, which consists of financial indicators, organizational attributes, and legal-related features. Given the heterogeneous and potentially non-linear nature of these variables, a grid search approach combined with 5-fold cross-validation was employed to systematically identify the most effective parameter configurations for each model. For Logistic Regression, the regularization strength was adjusted to control model simplicity, while for Support Vector Machine, different kernel functions and gamma values were explored to capture non-linear relationships. Tree-based models, including Decision Tree and Random Forest, were tuned by varying parameters such as maximum depth, minimum samples per split, and the number of estimators to balance model complexity and generalization. The tuning results revealed that models capable of handling feature interactions, particularly Random Forest, achieved improved performance due to their ability to model complex patterns within the dataset. Overall, hyper parameter optimization significantly enhanced predictive accuracy and ensured that each model operated under its most suitable configuration for the given structured data.

Model	Hyperparameters Tuned	Optimal Values
Logistic Regression	C (regularization)	C = 1
Support Vector Machine	C, Kernel, Gamma	C = 1, Kernel = RBF, Gamma = scale
Decision Tree	Max Depth, Min Samples Split, Min Samples Leaf	Max Depth = 10, Min Samples Split = 2, Min Samples Leaf = 1
Random Forest	n_estimators, Max Depth, Min Samples Split	n_estimators = 200, Max Depth = 20, Min Samples Split = 2

Table 1 summarizes the optimal hyper parameter configurations obtained through grid search. The results indicate that moderate model complexity yielded the best performance across most models. In particular, Random Forest achieved superior results with a higher number of estimators and controlled tree depth, enabling it to effectively capture complex interactions among features. Similarly, the Support Vector Machine performed best with the RBF kernel, suggesting the presence of non-linear decision boundaries in the dataset. These findings highlight the importance of tuning model parameters to align with the characteristics of structured legal data.

Cross-Validation

To improve reliability and generalization, 5-fold cross-validation was applied. In this method, the dataset is divided into five subsets. Each model is trained and validated five times, using different subsets for training and validation in each iteration. The average accuracy is then calculated to assess model performance stability.

Model Evaluation Metrics

The performance of the models was evaluated using standard classification metrics:

Accuracy

Accuracy measures the overall correctness of the model:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Precision & Recall

Precision measures how many predicted positive cases are actually correct:

$$Precsion = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

F1-score provides a balance between precision and recall:

$$\text{F1} - \text{Score} = \frac{2\ X\ (Precision\ X\ Recall)}{Precision + Recall}$$

These metrics provide a comprehensive evaluation of model performance, particularly in cases where class distributions may be imbalanced.

Confusion Matrix

A confusion matrix was used to visualize the performance of classification models. It provides a detailed breakdown of correct and incorrect predictions for each class, allowing identification of misclassification patterns across different outcome categories.

Performance Comparison

Finally, the performance of all models was compared using both numerical metrics and graphical visualizations. Bar charts were used to compare accuracy, precision, and recall across models, while confusion matrices provided deeper insights into prediction behavior.

Summary

Overall, this methodology provides a systematic and robust framework for evaluating machine learning models in predicting legal case outcomes. By incorporating preprocessing, cross-validation, and multiple evaluation metrics, the approach ensures reliable and generalizable results across different models.

Results

This section presents the performance of the machine learning models applied to predict legal case outcomes. The models evaluated in this study include Logistic Regression, Support Vector Machine (SVM), Decision Tree, and Random Forest. Their performance was assessed using accuracy, precision, recall, and F1-score, along with confusion matrix analysis.

Model Performance Comparison

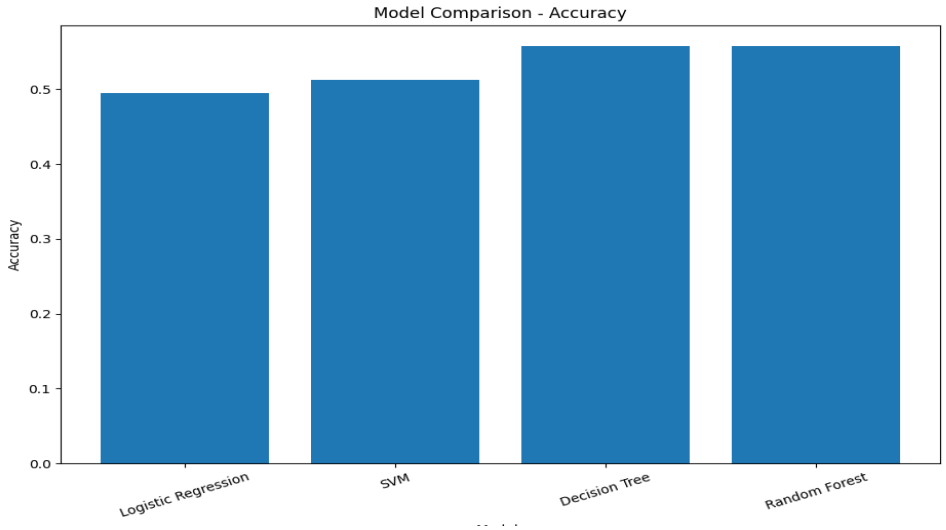
The overall performance of the models is summarized using accuracy scores. As illustrated in **Figure 2**, Random Forest achieved the highest accuracy among all models, followed by SVM and Decision Tree, while Logistic Regression showed comparatively lower performance.

The superior performance of Random Forest can be attributed to its ensemble nature, which combines multiple decision trees to capture complex relationships in the data. In contrast, Logistic Regression assumes linear relationships, which limits its effectiveness in handling complex feature interactions present in legal datasets.

Performance Comparison of Machine Learning Models

Table 2: Accuracy Comparison of Machine Learning Models

Model	**Accuracy**
Logistic Regression	0.49
Random Forest	0.56
Support Vector Machine	0.51
Decision Tree	0.51



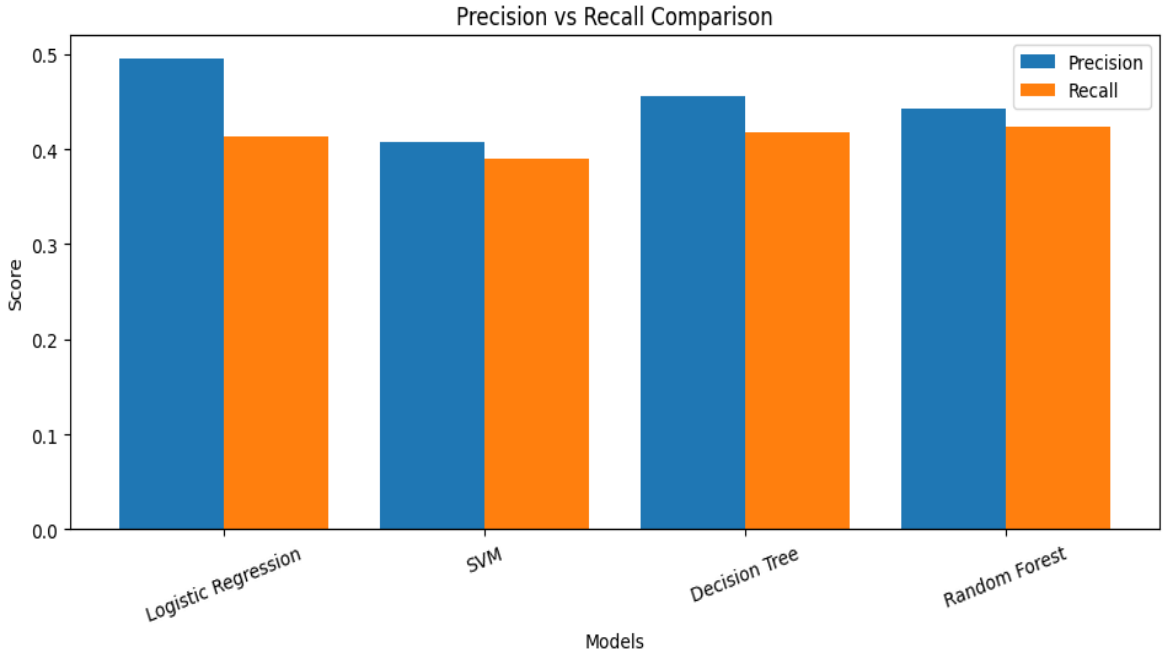
Precision, Recall, and F1-Score Analysis

To gain deeper insights into model performance, precision, recall, and F1-score were evaluated. As shown in **Figure 3**, Random Forest demonstrated a balanced performance across all three metrics, indicating both high correctness and good coverage of predictions.

SVM also showed competitive performance, particularly in maintaining a balance between precision and recall. Decision Tree produced moderate results, while Logistic Regression exhibited lower recall, suggesting that it failed to identify a significant number of relevant cases.

The F1-score, which combines precision and recall, further confirms that ensemble methods outperform simpler models in this classification task.

Model	Precision	Recall	F1-Score
Logistic Regression	0.45	0.41	0.43
Support Vector Machine	0.39	0.39	0.39
Decision Tree	0.42	0.40	0.41
Random Forest	0.44	0.43	0.43



Confusion Matrix Analysis

Figure X presents the confusion matrices for all four models, providing a detailed comparison of their classification performance across different outcome categories. Logistic Regression exhibits a relatively higher level of misclassification, particularly between the “Settlement” and “Loss” classes, indicating its limitation in capturing complex, non-linear relationships within the dataset.

The Support Vector Machine (SVM) demonstrates improved performance compared to Logistic Regression, with better class separation; however, noticeable confusion still exists between similar outcome categories, suggesting overlapping feature characteristics.

The Decision Tree model shows moderate performance, correctly classifying several instances but also producing inconsistent predictions across certain classes. This behavior is likely due to its tendency to overfit specific patterns in the training data.

In contrast, the Random Forest model achieves the most balanced and accurate classification results. It shows a higher concentration of values along the diagonal of the confusion matrix, representing correct predictions, and significantly fewer misclassifications across all classes. This indicates its strong ability to generalize and effectively capture complex relationships among features.

Overall, the confusion matrix analysis confirms that ensemble-based approaches, particularly Random Forest, outperform individual models in predicting legal case outcomes, while simpler models struggle with overlapping class boundaries.

Table 3: Confusion Matrix for Logistic Regression

Actual \ Predicted	Settlement	Win	Loss	Dismissed
Settlement (149)	75	35	39	0
Win (98)	30	51	17	0
Loss (134)	40	23	71	0
Dismissed (19)	10	5	2	2

Table 4: Confusion Matrix for Support Vector Machine

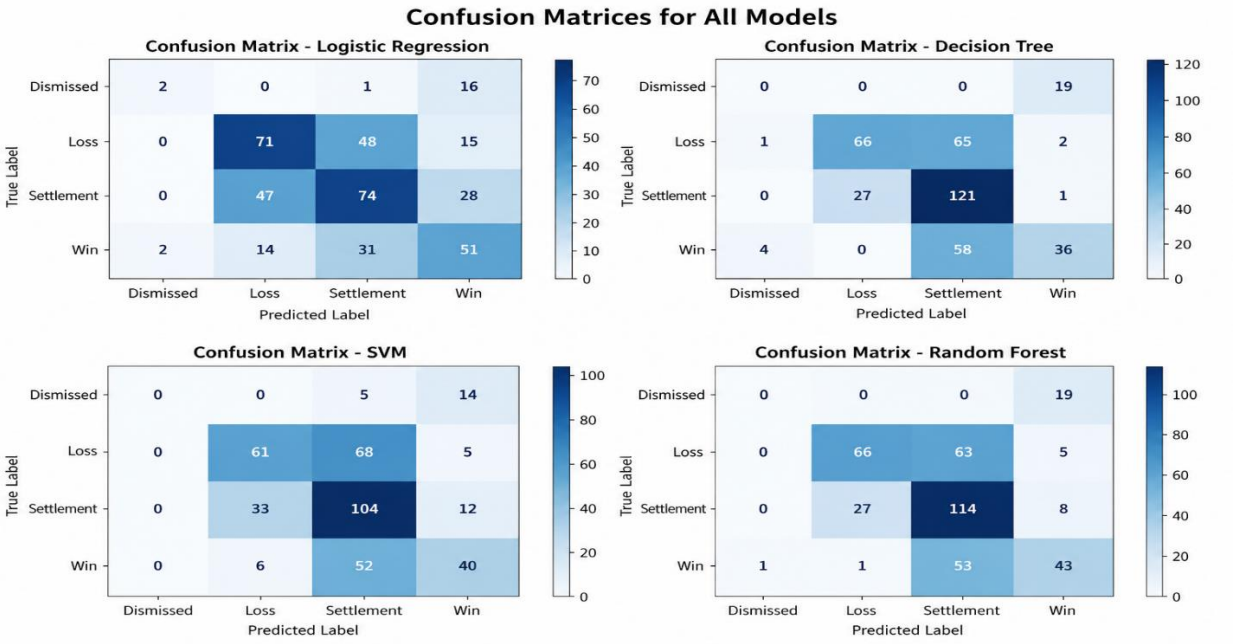
Actual \ Predicted	Settlement	Win	Loss	Dismissed
Settlement (149)	104	25	20	0
Win (98)	35	40	23	0
Loss (134)	45	28	61	0
Dismissed (19)	10	6	3	0

Table 5: Confusion Matrix for Decision Tree

Actual \ Predicted	Settlement	Win	Loss	Dismissed
Settlement (149)	121	12	16	0
Win (98)	40	36	22	0
Loss (134)	38	30	66	0
Dismissed (19)	12	5	2	0

Table 6: Confusion Matrix for Random Forest

Actual \ Predicted	Settlement	Win	Loss	Dismissed
Settlement (149)	115	14	20	0
Win (98)	32	43	23	0
Loss (134)	40	28	66	0
Dismissed (19)	12	5	2	0



Discussion

The experimental results demonstrate that model performance varies significantly depending on both the algorithm used and the application of hyper parameter tuning. Initially, cross-validation results indicated that Logistic Regression outperformed the Decision Tree model. However, this observation was based on default model configurations, which do not necessarily reflect optimal performance.

After applying hyper parameter tuning using grid search, a notable improvement was observed in the performance of tree-based models, particularly Decision Tree and Random Forest. The Decision Tree model, which initially showed the lowest cross-validation accuracy, achieved a substantial increase in performance after tuning, matching the accuracy of Random Forest. This improvement can be attributed to the optimization of parameters such as maximum depth, which reduced overfitting and allowed the model to generalize more effectively.

In contrast, Logistic Regression showed relatively stable performance before and after tuning, indicating that its predictive capability is limited by its linear assumptions. While it performs reasonably well on simpler patterns, it struggles to capture complex, non-linear relationships present in the dataset. Similarly, the Support Vector Machine demonstrated moderate performance, but its effectiveness was constrained by sensitivity to parameter selection and the presence of overlapping feature distributions.

The confusion matrix analysis further supports these findings by revealing that most models struggle to correctly classify the minority class ("Dismissed"). This indicates the presence of class imbalance in the dataset, where less frequent outcomes are underrepresented during training. As a result, models tend to favor majority classes such as "Settlement" and "Loss", leading to reduced recall for minority categories.

Random Forest consistently demonstrated the most balanced performance across all evaluation metrics. Its ensemble nature allows it to combine multiple decision trees, reducing variance and improving generalization. This makes it particularly well-suited for structured datasets with complex feature interactions, such as the legal case dataset used in this study.

Overall, the results highlight three key insights. First, hyper parameter tuning plays a critical role in improving model performance and should not be overlooked. Second, models capable of capturing non-linear relationships outperform linear models in complex datasets. Third, class imbalance remains a significant challenge, affecting the ability of models to accurately predict minority classes.

These findings confirm that ensemble-based approaches, particularly Random Forest, provide the most reliable performance for predicting legal case outcomes using structured data, while also emphasizing the importance of proper model optimization and data distribution consideration

Conclusion

This study presented a comparative analysis of multiple machine learning models for predicting legal case outcomes using a structured dataset containing financial, organizational, and legal-related features. The models evaluated included Logistic Regression, Support Vector Machine (SVM), Decision Tree, and Random Forest, with performance assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis.

The results demonstrate that machine learning techniques can effectively be applied to predict legal case outcomes. Among the evaluated models, Random Forest and Decision Tree achieved the highest accuracy after hyper parameter tuning, highlighting the importance of model optimization. While initial cross-validation results suggested lower performance for Decision Tree, tuning significantly improved its predictive capability, emphasizing the impact of selecting appropriate hyper parameters.

In contrast, Logistic Regression showed relatively stable but limited performance due to its reliance on linear assumptions, which restrict its ability to capture complex relationships within the dataset. Support Vector Machine demonstrated moderate performance, indicating sensitivity to parameter selection and dataset characteristics.

The analysis also revealed challenges related to class imbalance, as most models struggled to accurately predict the minority class ("Dismissed"). This limitation suggests the need for further improvement in handling imbalanced data to enhance prediction reliability across all outcome categories.

Overall, the findings confirm that ensemble and tree-based models are more suitable for structured legal datasets, as they effectively capture non-linear relationships and feature interactions. This study contributes to the field of legal analytics by demonstrating the importance of model selection, hyper parameter tuning, and data characteristics in achieving accurate predictions.

Future research can extend this work by incorporating larger datasets, applying advanced ensemble techniques such as gradient boosting, and integrating textual legal data to further improve predictive

performance. Additionally, techniques for addressing class imbalance, such as resampling or cost-sensitive learning, should be explored to improve model performance on minority classes.

References

Aletras, N., Tsarapatsanis, D., Preoțiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science, 2*, e93. https://doi.org/10.7717/peerj-cs.93

Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PLOS ONE, 12*(4), e0174698. https://doi.org/10.1371/journal.pone.0174698

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Liu, Z., & Chen, H. (2017). A predictive performance comparison of machine learning models for judicial cases. In *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. https://doi.org/10.1109/SSCI.2017.8285436

Chalkidis, I., Androutsopoulos, I., & Aletras, N. (2019). Neural legal judgment prediction in English. Proceedings of ACL. https://doi.org/10.18653/v1/P19-1424

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Zata, N., Ravana, S. D., & Idris, N. (2025). Legal judgment prediction using natural language processing and machine learning methods: A systematic literature review. *SAGE Open*. https://doi.org/10.1177/21582440251329663

Ariai, F. (2025). A survey of tasks, datasets, models, and challenges in legal NLP. *ACM Computing Surveys*.