Journal of Sciences, Islamic Republic of Iran 35(2): 147 - 157 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Classifying Divorce Cases in Iranian Judiciary Courts Using Machine Learning: A Predictive Perspective

E. Tabrizi[1*] and M. A. Farzammehr[2]

[1] *Department of Mathematics, Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Islamic Republic of Iran.*
[2] *Judiciary Research Institute, Tehran, Islamic Republic of Iran*

Received: 3 October 2024 / Revised: 1 January 2025 / Accepted: 22 January 2025

Abstract

This study develops a machine learning model to predict the classification of divorce cases in Iranian Judiciary Courts based on socioeconomic factors. Using data collected between 2011 and 2018 and various machine learning algorithms, the study evaluates the performance of predictive models through a rigorous 10-fold cross-validation process. Results highlight the Random Forest and Neural Network classifiers as the most accurate. Key socioeconomic factors influencing divorce cases, such as unemployment rate and urbanization rate, are identified. The findings provide actionable insights for policymakers to develop data-driven strategies for social policy and resource allocation.

Keywords: Divorce Cases; Data Mining; Machine Learning Techniques; Iran; Judiciary.

Introduction

In recent years, Iran has witnessed a significant increase in divorce rates, which has become a major concern for policymakers and society as a whole. According to the latest available data from the Statistical Center of Iran, the divorce rate in the country has risen from 8.7 per 1,000 marriages in 2006 to 20.8 per 1,000 marriages in 2020 (Statistical Center of Iran, 2020).

Given the social and economic impacts of divorce on families and society, it is crucial to predict divorce trends in civil courts in order to anticipate the demand for legal services and allocate resources accordingly. Accurately predicting divorce cases in civil courts can assist policymakers and court officials in planning for future caseloads, allocating resources, and developing effective policies and programs to support families undergoing divorce.

Predictive modeling using machine learning techniques offers a promising approach to forecasting divorce trends in civil courts, as it can consider a wide range of socioeconomic factors and identify the most important predictors of divorce rates. This information can then be used to inform policy decisions and develop targeted interventions to support families at risk of divorce (1).

Machine learning algorithms use statistical and computational techniques to identify patterns and relationships in large datasets and use these patterns to make predictions on new data. In this case, the algorithm would utilize historical data on divorce rates and socioeconomic factors to build a predictive model capable of forecasting future trends in divorce rates (2,

* Corresponding Author: Tel: +98 21-88329220; Email: elham.tabrizi@khu.ac.ir

3).

Machine learning offers several advantages over traditional methods, such as improved accuracy and the ability to handle large and complex datasets. The use of machine learning in predictive modeling can also identify important predictors that may not be obvious using conventional methods and can provide insights into the underlying factors driving divorce trends in Iran.

Overall, the use of machine learning in predictive modeling can help policymakers and practitioners in Iranian civil courts better understand and prepare for future changes in the volume of divorce cases, as well as develop more effective policies and interventions to address the social and economic issues associated with divorce.

The main innovation of this study lies in the use of a unique and confidential dataset, which was obtained through a challenging data acquisition process, to predict divorce cases in Iranian Judiciary Courts. This dataset's sensitivity and rarity allow for insights that would not be achievable using publicly available data. The study comprehensively compares ten classification algorithms using a rigorous 10-fold cross-validation approach. Among these, the Random Forest and Neural Network classifiers demonstrate superior performance. These findings provide valuable insights for data-driven decision-making in social policy planning and resource allocation.

Literature Review

Divorce is a complex phenomenon that has been widely researched in the social sciences. Numerous studies have examined the factors contributing to divorce rates, including sociocultural, economic, and demographic factors (4, 5, 6). Among these, socioeconomic factors have been found to play a significant role in predicting divorce rates (5).

One of the most commonly studied socioeconomic factors is education. Studies have found that higher levels of education are associated with lower divorce rates (4, 6). This may be because education provides individuals with the skills and resources necessary to maintain stable and healthy relationships.

Another important socioeconomic factor is income. Several studies have found that lower income is associated with higher divorce rates (5, 7). Financial strain can create tension and conflict within a marriage, which may contribute to divorce.

In addition to socioeconomic factors, demographic factors such as age and gender have also been found to be associated with divorce rates. For example, studies have found that a younger age at marriage is associated with higher divorce rates (4, 6). Gender has also been found to play a role, with some studies indicating that women are more likely to initiate divorce than men (8, 5).

Predicting divorce rates is an area of research that has recently been explored using machine learning algorithms. Several studies have implemented machine learning algorithms to predict divorce rates based on a variety of factors (9, 10). These algorithms have shown promising results and may be useful in predicting divorce rates in different populations and contexts.

Another study focused on divorce case prediction using machine learning algorithms, exploring the use of machine learning techniques to identify factors influencing divorce cases and predict divorce outcomes (3). Additionally, researchers investigated the prediction of divorce among Malaysian women using machine learning techniques, examining variables such as demographic, socioeconomic, and behavioral factors (10).

Recent research has further expanded the application of machine learning in divorce prediction. For instance, explainable machine learning techniques were used to predict divorce, emphasizing interpretability through Local Interpretable Model-agnostic Explanations (LIME) (11). Similarly, ensemble learning models, including Support Vector Machines, Linear Models, and Neural Networks, have been compared to determine the most accurate predictors of divorce (12). Another study applied various machine learning algorithms to predict divorce cases in Ha'il, Saudi Arabia, highlighting the role of regional demographic and socioeconomic data in model accuracy (13). Research on union dissolution in Germany employed machine learning techniques to identify key predictors and trends that contribute to relationship instability (14).

These studies underscore the effectiveness of machine learning in analyzing divorce trends across different regions and populations, offering valuable insights for policymakers and researchers. The integration of recent machine learning advancements enhances our understanding of the multifaceted nature of divorce and provides new avenues for targeted interventions.

Overall, the literature suggests that a variety of sociocultural, economic, and demographic factors contribute to divorce rates. While higher levels of education and income may protect against divorce, younger age at marriage and financial strain may increase the likelihood of divorce.

In conclusion, the reviewed literature suggests that machine learning algorithms can be effectively used to predict divorce rates in the context of the Iranian

Judiciary Courts. Future research should focus on refining prediction models, integrating domain knowledge, and addressing legal and ethical concerns associated with using machine learning algorithms to predict divorce cases in order to improve the accuracy of these predictions.

Data Collection and Preprocessing

This study is based on a unique dataset collected from the Iranian Statistics Center and the Judiciary Statistics and Information Technology Center, covering data from 31 provinces in Iran over the years 2011 to 2018. The dataset includes 217 instances and 49 features, comprehensively reflecting socioeconomic, demographic, and legal factors influencing divorce cases.

The dataset is summarized in two tables (Tables 1 and 2), which present the mean and standard deviation of key variables to provide an overview of their distribution across provinces. These descriptive statistics serve to introduce the dataset's features rather than being used directly in the predictive model. Variables such as unemployment rate, literacy rate, urbanization rate, and legal case types (e.g., drug-related or theft cases) are included, allowing a nuanced understanding of the factors impacting divorce trends.

In addition to the collected variables, a new nominal variable, "Divorce Category," was created to classify the divorce cases into three levels: 'Low', 'Medium', and 'High'. This classification was based on the percentage of divorce cases relative to all legal cases in each province. If the percentage was less than 33%, the "Divorce Category" was assigned as 'Low.' Values between 33% and 66% were classified as 'Medium,' and those above 66% as 'High.' This categorization ensures that the target variable captures the variance in divorce rates across provinces while remaining interpretable.

Table 1 focuses on demographic and socioeconomic indicators, presenting variables such as unemployment rate, population distribution by age and location (urban/rural), literacy rates, and labor force participation. The table highlights variations across gender and urban-rural divides.

Table 2 includes economic and legal factors such as GDP, inflation rate, urbanization rate, Gini coefficient, and the number of legal cases related to specific issues (e.g., drugs, theft). These variables provide insight into both the socioeconomic environment and the judicial

Table 1. Means (Dispersions) of attributes

Variable	Unemployment rate	Population aged 15 and over	Literacy rate in the population aged 6 and over	Participation rate
Rural (male)	8.01(0.38)	230948.19(0.69)	81.63(0.04)	73.21(0.08)
Rural (female)	10.77(0.78)	217943.16(0.77)	70.57(0.08)	16.16(0.44)
Rural (male & female)	8.24(0.37)	448891.80(0.72)	76.20(0.06)	44.35(0.12)
Urban (male)	10.96(0.30)	697480.57(1.24)	91.46(0.02)	66.72(0.06)
Urban (female)	25.07(0.29)	697610.89(1.23)	85.03(0.04)	14.09(0.22)
Urban (male & female)	13.41(0.27)	1395091.95(1.24)	88.08(0.03)	40.39(0.07)
Rural & Urban (male)	10.00(0.29)	887083.81(1.03)	88.21(0.03)	68.62(0.06)
Rural & Urban (female)	19.79(0.34)	945841.74(0.99)	80.25(0.06)	14.83(0.25)
Rural & Urban (male & female)	11.67(0.25)	1715253.78(1.06)	84.27(0.05)	41.57(0.08)

Table 2. Continued

Variable	Mean (Dispersion)
Number of cases related to drugs	15396.70(1.08)
Number of cases related to alcoholic beverages	1671.79(1.44)
The number of cases related to theft that require punishment	20972.88(1.43)
Gini coefficient - rural areas	0.30(0.11)
Gini coefficient - urban areas	0.32(0.11)
Consumer price index = annual inflation rate (total index)	78.10(0.30)
Gross domestic product (at market price in billion rials)	222667.91(1.43)
Total added value of 18 sectors (at market price in billion rials)	219306.10(1.43)
Average age of men's first marriage	26.76(0.04)
Average age of women's first marriage	23.31(0.04)
Share of Provinces in total migration	3.23(1.02)
Urbanization rate	66.09(0.18)
Internet Penetration Rate for Population aged 15 to 24	37.54(0.45)

workload within provinces.

The dataset underwent preprocessing to ensure its suitability for machine learning. All features were examined for their relevance to divorce prediction, and noisy or redundant features were removed to improve model performance. The final dataset was divided into training and testing subsets, enabling robust evaluation of machine learning models. Additionally, to enhance computational efficiency, feature selection techniques were applied, prioritizing high-impact variables based on statistical measures such as information gain and chi-square tests.

By structuring the dataset in this way, we ensured that the predictive model leveraged granular, high-quality data that accurately reflects the diverse socioeconomic and judicial conditions across Iran's provinces.

Materials and Methods

Machine learning is a subfield of artificial intelligence that focuses on developing algorithms and statistical models that enable computers to automatically improve their performance on specific tasks based on data inputs. In other words, rather than being explicitly programmed to perform a task, the computer learns to perform the task by analyzing patterns in data inputs.

Several machine learning techniques are employed in this study, including Neural Network, Naïve Bayes, Adaptive Boosting (AdaBoost), Gradient Boosting (GraBoost), Random Forest, Decision Tree, k-Nearest Neighbor (kNN), Stochastic Gradient Descent (SGD), and Support Vector Machine (SVM). These techniques are used to create predictive models for divorce case levels based on socioeconomic factors.

Each technique has its own unique approach to analyzing patterns in data and making predictions. For example, Decision Tree uses a hierarchical structure of nodes to make decisions based on a set of conditions, while SVM is a supervised classification algorithm that constructs a separating hyperplane in high-dimensional space for classification. A good separation is obtained by the hyperplane that maximizes the distance to the nearest training data point of any class (1). Neural Networks use layers of interconnected nodes to process complex data inputs and make predictions.

The selection of classification algorithms in this study was conducted with a focus on achieving methodological diversity, computational efficiency, interpretability, and effectiveness for structured datasets. Ten algorithms were chosen to represent a broad spectrum of classification paradigms while maintaining a balance between simplicity and complexity.

The study included Decision Tree-Based Methods (Decision Tree, Random Forest), which provide interpretability and robustness against overfitting, particularly for high-dimensional data. Linear Models such as Stochastic Gradient Descent (SGD) and Multinomial Logistic Regression were selected to establish baseline comparisons and address linearly separable datasets. Support Vector Machines (SVMs) were included for their ability to handle both linear and non-linear data using kernel methods. Neural Networks, specifically Multilayer Perceptron (MLP), were employed to capture non-linear relationships and higher-order feature interactions.

Additionally, Instance-Based Learning (k-Nearest Neighbors, kNN) was used for its simplicity and proximity-based classification capabilities. Probabilistic Models like Naïve Bayes were selected for their low computational complexity and probabilistic foundations. Finally, Ensemble Methods (Adaptive Boosting, Gradient Boosting) were included for their iterative boosting strategies that enhance performance by reducing bias and variance.

While the range of classification algorithms is vast, this subset was chosen to provide a meaningful comparison across linear and non-linear models, parametric and non-parametric techniques, and simple versus complex ensemble approaches.

To evaluate the performance of the predictive models, several metrics are used, including AUC, CA, F1, Precision, and Recall. These metrics assess how well the model predicts the outcome of interest (in this case, the level of court divorce cases) based on the input variables. Orange software is used to implement and evaluate the performance of the various machine learning techniques applied in this study.

The dataset is first divided into training and testing sets using a 70:30 split. Each model is trained on the training set and evaluated on the testing set. This process is repeated for each model, and the performance metrics are recorded for each.

To select the best model, the performance metrics for each model are compared. The model with the highest AUC, F1 score, and CA, as well as the highest precision and recall values, is selected. Additionally, the model's ability to generalize to new data is assessed using k-fold cross-validation with 10 folds, where the data is split into 10 subsets, and the model is trained and evaluated on each subset. This helps determine if the model is overfitting to the training data.

The ROC curve is also used to assess the accuracy of a diagnostic test in a categorical case with three levels, such as a disease that can be classified as mild, moderate, or severe. The ROC curve plots the true positive rate against the false positive rate for different cutoff values of the diagnostic test. The area under the ROC curve (AUC) ranges from 0.50 to 1.00, with a higher AUC indicating better discriminative ability across the three levels of disease severity. In this paper, a larger AUC indicates that the diagnostic test can more effectively distinguish between the three levels of divorce volume, allowing for more effective policies and interventions to address the social and economic issues associated with divorce.

Overall, the model evaluation and selection process allowed us to identify the most suitable machine learning technique for predicting the volume of divorce cases in Iranian judiciary courts, ensuring the selected model is both accurate and able to generalize well to new data.

Furthermore, this study conducted experiments with four different feature selection algorithms: ReliefF, Information Gain, Chi-Square, and Gain Ratio. ReliefF is an enhanced version of the Relief statistical model developed by Kononenko in 1994. This algorithm selects features by creating a model based on a sample's proximity to other samples within the same class and its distance from different classes. Compared to Relief, ReliefF is a more robust algorithm, capable of handling missing and noisy data, and is applicable in all situations. It is less biased, allows for feature interaction, and can capture local dependencies missed by other methods. The Information Gain method is commonly used in feature selection to identify the feature set that provides the most knowledge about the classes. This entropy-based algorithm calculates the information gain coefficient for each attribute and selects feature sets with the highest coefficients. The Chi-Square method is a statistics-based algorithm that calculates the chi-square of all attributes and evaluates them individually based on their class. Gain Ratio, on the other hand, is an alternative version of Information Gain that maximizes feature information gain while minimizing the number of feature values, unlike Information Gain, which favors features with a large number of values.

Results

In this study, nine data mining models were utilized for classification purposes. Specifically, the models were applied to classify the outcome into one of three categories: Low, Medium, or High, using a collection of 49 independent variables, a subset of which is enumerated in Tables 1 and 2. To evaluate the performance of these models, 10-fold cross-validation was applied separately to each algorithm using the input dataset. This approach ensured a comprehensive and rigorous evaluation of the models' classification ability.

Model Selection

Table 3 presents a ranking of the best-performing models based on AUC, CA, F1, Precision, and Recall. Additionally, the results for four different target classes—namely, average over classes, Low, Medium, and High—are presented in Table 3. Overall, based on AUC, it can be observed that Random Forest, followed by Neural Network, outperformed the other models on both the training and test data. According to CA, F1, Precision, and Recall, it is evident that while AdaBoost shows better accuracy on the training data compared to Random Forest and Neural Network, Random Forest and Neural Network exhibit higher accuracy on the test data, indicating their greater predictive ability. The results suggest that the best overall performance was achieved when the target class was 'High,' while the worst performance was observed for the 'Medium' class.

As we know, if a model suffers from overfitting and demonstrates a higher dependency on training data, the model evaluation criteria will typically show higher values on the training data, while these values will

Table 3. Performance metrics of the nine data mining models

Model (Average Over Classes) criteria		kNN	Decision Tree	SVM	SGD	Random Forest	Neural Network	Naïve Bayes	Multinomial Logistic Regression	GraBoost	AdaBoost
AUC	Train	0.9525	0.9233	0.9695	0.9274	0.981	0.9718	0.9683	0.9277	0.9727	0.9401
	Test	0.9663	0.9313	0.9571	0.9034	0.9905	0.9787	0.9652	0.9197	0.9813	0.8886
CA	Train	0.8355	0.8553	0.9013	0.9013	0.9145	0.9013	0.8487	0.8092	0.8947	0.9211
	Test	0.8462	0.8923	0.8769	0.8615	0.9385	0.9385	0.8923	0.8	0.8308	0.8308
F1	Train	0.836	0.8562	0.9013	0.901	0.9143	0.901	0.8483	0.8093	0.8942	0.9205
	Test	0.8472	0.894	0.8798	0.8681	0.9393	0.9393	0.8925	0.8074	0.8397	0.8397
Precision	Train	0.8431	0.8584	0.9015	0.903	0.9142	0.9014	0.8505	0.8099	0.8944	0.9233
	Test	0.8488	0.9082	0.9046	0.8859	0.9429	0.9429	0.8944	0.8186	0.8903	0.8903
Recall	Train	0.8355	0.8553	0.9013	0.9013	0.9145	0.9013	0.8487	0.8092	0.8947	0.9211
	Test	0.8462	0.8923	0.8769	0.8615	0.9385	0.9385	0.8923	0.8	0.8308	0.8308

Table 3. Continued

Model (Low) **criteria**		**kNN**	**Decision Tree**	**SVM**	**SGD**	**Random Forest**	**Neural Network**	**Naïve Bayes**	**Multinomial Logistic Regression**	**GraBoost**	**AdaBoost**
AUC	Train	0.9459	0.9327	0.9652	0.9073	0.981	0.9635	0.972	0.9262	0.982	0.928
	Test	0.9556	0.9038	0.9477	0.8782	0.9877	0.9773	0.9586	0.9103	0.9921	0.8462
CA	Train	0.875	0.9276	0.9276	0.9079	0.9474	0.9079	0.875	0.8684	0.9408	0.9539
	Test	0.8615	0.9231	0.8769	0.8923	0.9385	0.9385	0.9077	0.8308	0.8769	0.8769
F1	Train	0.7765	0.8791	0.8791	0.8542	0.913	0.8511	0.7816	0.7778	0.9011	0.9195
	Test	0.8235	0.8936	0.8261	0.8571	0.92	0.92	0.88	0.7843	0.8182	0.8182
Precision	Train	0.8462	0.8889	0.8889	0.82	0.913	0.8333	0.8293	0.7955	0.9111	0.9756
	Test	0.84	1	0.95	0.913	0.9583	0.9583	0.9167	0.8	1	1
Recall	Train	0.7174	0.8696	0.8696	0.8913	0.913	0.8696	0.7391	0.7609	0.8913	0.8696
	Test	0.8077	0.8077	0.7308	0.8077	0.8846	0.8846	0.8462	0.7692	0.6923	0.6923
AUC	Train	0.9349	0.8813	0.9573	0.885	0.9736	0.9593	0.9587	0.8916	0.9804	0.9183
	Test	0.9384	0.909	0.9104	0.8599	0.9811	0.951	0.9272	0.8221	0.965	0.8662
CA	Train	0.8355	0.8553	0.9013	0.9013	0.9145	0.9013	0.8487	0.8158	0.8947	0.9211
	Test	0.8615	0.8923	0.8769	0.8615	0.9385	0.9385	0.8923	0.8	0.8308	0.8308
F1	Train	0.7967	0.8136	0.8696	0.8624	0.885	0.8649	0.8067	0.7586	0.8571	0.8947
	Test	0.6897	0.7742	0.7647	0.7273	0.8667	0.8667	0.7586	0.5806	0.7027	0.7027
Precision	Train	0.7424	0.7869	0.8621	0.9038	0.8929	0.8889	0.7742	0.7458	0.8727	0.8947
	Test	0.6667	0.7059	0.65	0.6316	0.8125	0.8125	0.7333	0.5294	0.5652	0.5652
Recall	Train	0.8596	0.8421	0.8772	0.8246	0.8772	0.8421	0.8421	0.7719	0.8421	0.8947
	Test	0.7143	0.8571	0.9286	0.8571	0.9286	0.9286	0.7857	0.6429	0.9286	0.9286
.AUC	Train	0.9859	0.9436	0.998	0.995	0.9991	1	0.9937	0.9599	1	0.9755
	Test	0.997	0.975	1	0.96	1	1	0.999	0.999	0.982	0.9475
CA	Train	0.9605	0.9276	0.9737	0.9934	0.9671	0.9934	0.9737	0.9342	0.9539	0.9671
	Test	0.9692	0.9692	1	0.9692	1	1	0.9846	0.9692	0.9538	0.9538
F1	Train	0.9375	0.8842	0.9592	0.9899	0.9495	0.9899	0.9592	0.898	0.9307	0.9515
	Test	0.96	0.9615	1	0.9583	1	1	0.9804	0.9583	0.9388	0.9388
Precision	Train	0.9574	0.913	0.9592	0.98	0.94	0.98	0.9592	0.898	0.9038	0.9074
	Test	0.96	0.9259	1	1	1	1	0.9615	1	0.9583	0.9583
Recall	Train	0.9184	0.8571	0.9592	1	0.9592	1	0.9592	0.898	0.9592	1
	Test	0.96	1	1	0.92	1	1	1	0.92	0.92	0.92

significantly decrease on the test data. Based on Table 3, all models generated stable classification results, and there were no clear signs of overfitting.

Finally, the experimental results demonstrate that all ten prediction models achieved good performance. Furthermore, the AUC values of the ten models were all greater than 0.88.

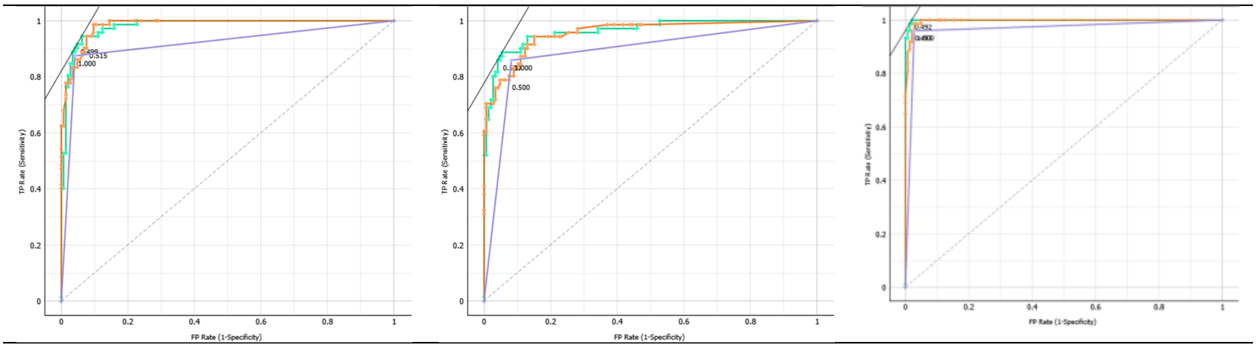
In the following, the success of some proposed methods is evaluated using various evaluation criteria, including the utilization of ROC curves. ROC curves offer a graphical representation of the performance of classification models. By examining these curves, we can select a set of candidate models, compare them, and report the best one based on their ROC curve performance. Figure 1 presents the ROC curves obtained for the Random Forest, Neural Network, and AdaBoost models. Upon examining the figure:

1. If the target class is 'Low', it becomes evident that the area under the ROC curve is highest for the Random Forest and Neural Network algorithms (AUC: 0.984, 0.981).
2. If the target class is 'Medium', the Neural Network and then Random Forest algorithms (AUC: 0.964, 0.952) outperform the AdaBoost algorithm.
3. If the target class is 'High', the best model is Neural Network (AUC: 1).

This indicates the superior classification performance of the Neural Network and Random Forest algorithms. Conversely, the AdaBoost method yielded the poorest result, with an AUC of 0.887.

Note that ROC curves provide valuable visualization of the trade-off between sensitivity and specificity for each model, facilitating a straightforward comparison of their classification performance. The outstanding performances of the Neural Network and Random Forest algorithms, as highlighted by their large AUC values, underscore their efficacy as two powerful classification techniques.

Table 4 presents the confusion matrix for three classification models: Random Forest, Neural Network, and AdaBoost. The confusion matrix provides insights into the performance of these models, based on the training data. Among them, the Neural Network and Random Forest stand out, accurately classifying 200 and



a. Target class = 'Low' | b. Target class = 'Medium' | c. Target class = 'High'

Figure 1. Roc curves for nine machine learning models across 3 classes

Table 4. Confusion Matrix using the Three Machine Learning Models

Model	Level	Low	Medium	High	Correct	Incorrect
Neural	Low	65	7	0	200	17
Network	Medium	7	61	3		
	High	0	0	74		
Random	Low	69	3	0	198	19
Forest	Medium	9	58	4		
	High	0	3	71		
AdaBoost	Low	63	9	0	195	22
	Medium	6	61	4		
	High	0	3	71		

198 out of 217 instances, with only 17 and 19 misclassifications, respectively. It is worth noting that the models generally exhibit a higher number of false negatives than false positives, indicating a higher rate of Type II errors (incorrectly classifying positive cases) compared to Type I errors (incorrectly classifying negative cases). In practical terms, this suggests that the models are more likely to correctly identify negative cases but may be more prone to misidentifying positive cases. The confusion matrix serves as a valuable tool for assessing and comparing the performance of the classification models in terms of their predictive accuracy.

The classification results obtained using the Neural Network and Random Forest algorithms, as presented in Table 4, demonstrate the distribution of items across the 'Low', 'Medium', and 'High' classes. Among the 72 items classified as 'Low', 65 and 69 were correctly assigned to this class, while 7 and 3 were erroneously classified as 'Medium', respectively. Notably, none of the items were misclassified as 'High'.

Similarly, out of the 71 items classified as 'Medium', 61 and 58 were accurately categorized as such. However, in the Neural Network model, 7 were mistakenly labeled as 'Low' and 3 as 'High'. In the Random Forest model, 9 were mistakenly labeled as 'Low' and 4 as 'High'.

For the Neural Network model, all 74 items classified as 'High' were correctly identified, with none misclassified as 'Low' or 'Medium'. In contrast, within the Random Forest model, 71 items were correctly assigned to the 'High' class, but 3 items were mistakenly labeled as 'Medium'.

Ultimately, when examining all the models, it becomes evident that the Random Forest and Neural Network models outperform others in accurately predicting the classification of divorce cases within the Iranian Judiciary Courts.

Feature Selection

Through the application of the model, various experiments were conducted to determine the most compelling feature sets. Testing was performed on a range of results, spanning from the most impactful single feature to all 49 features. This was achieved by employing five distinct feature selection algorithms, with the subsequent classification outcomes meticulously recorded. In this subsection, we aim to rank features by assigning scores based on their correlation with the discrete target variable, utilizing relevant internal scoring methods such as information gain, chi-square, and others.

The sequencing of the influential features obtained through different feature selection methods is presented.

		#	Info. gain	Gain ratio	Gini	ANOVA	χ²	ReliefF	FCBF
1	N Population aged 15 and over (rural-male)		0.558	0.279	0.233	112.472	78.130	0.160	0.452
2	N Population aged 15 and over (rural - male and female)		0.430	0.215	0.191	91.718	66.187	0.144	0.000
3	N Population aged 15 and over (rural-female)		0.369	0.185	0.170	68.763	55.933	0.134	0.000
4	N Population aged 15 and over (urban and rural-female)		0.917	0.459	0.388	97.580	134.085	0.109	0.000
5	N Population aged 15 and over (urban and rural- male)		0.805	0.402	0.338	90.842	127.328	0.095	0.000
6	N Population aged 15 and over (urban-female)		0.969	0.485	0.392	67.346	141.324	0.089	1.178
7	N Population aged 15 and over (urban-male and female)		0.950	0.475	0.382	66.143	139.916	0.089	0.000
8	N Population aged 15 and over (urban-male)		0.950	0.475	0.382	64.946	139.916	0.088	0.000
9	N Population aged 15 and over (urban and rural-male and female		0.684	0.342	0.286	82.361	114.295	0.087	0.000
10	N Share of Provinces in total migration		0.714	0.357	0.274	73.825	118.406	0.085	0.663
11	N Urbanization rate		0.122	0.061	0.058	10.552	16.133	0.080	0.000
12	N Number of cases related to drugs		0.561	0.280	0.242	55.726	101.665	0.070	0.000
13	N Gross domestic product (at market price in billion rials)		0.532	0.266	0.206	38.553	85.987	0.063	0.000
14	N Literacy rate in population aged 6 years and older (urban- female)		0.146	0.073	0.065	18.030	23.472	0.063	0.000
15	N Total added value of 18 sectors (at market price in billion rials)		0.525	0.263	0.203	38.991	83.833	0.062	0.000
16	N Average age of women's first marriage		0.029	0.015	0.013	0.822	3.976	0.061	0.000
17	N Average age of men's first marriage		0.131	0.065	0.061	5.397	16.375	0.053	0.000
18	N The number of cases related to theft that require punishment		0.693	0.347	0.288	41.768	117.146	0.053	0.000
19	N Literacy rate in the population aged 6 years and older (urban-male and female)		0.172	0.086	0.078	20.887	31.111	0.052	0.000
20	N Number of cases related to alcoholic beverages		0.682	0.341	0.290	64.689	116.670	0.051	0.614

Figure 2. Effective feature orders obtained according to different feature selection algorithms

The top twenty ranked features are enumerated in Figure 2, along with their corresponding algorithm-assigned scores. For instance, according to the ReliefF algorithm, the most prominent feature is "Population aged 15 and over (rural - male)," while the feature with the least impact based on the same algorithm is "Consumer price index." In contrast, the Information Gain algorithm designates "Population aged 15 and over (urban - female)" as the most influential feature, with "Participation rate (rural - female)" being the least impactful. Similarly, the Gain Ratio algorithm also identifies "Population aged 15 and over (urban - female)" as the most influential feature, with "Participation rate (rural - female)" as the least influential. Lastly, according to the Chi-Square algorithm, "Population aged 15 and over (urban - female)" ranks as the most effective feature, while "Unemployment rate (urban - male & female)" is deemed the least effective.

Note that The red lines represent the calculated importance scores for each feature as determined by these algorithms. Each score indicates the contribution of the respective feature to the predictive accuracy of the classification models used in the study. A higher score above the red line suggests a more significant contribution of the feature to the prediction outcome, while lower scores indicate minimal impact. Each algorithm employs a distinct statistical approach for measuring feature importance, and thus, the scores across different algorithms are not directly comparable but provide complementary insights. Also, The importance of the alcoholic beverages feature fluctuates significantly across the algorithms, suggesting that its influence depends on the methodology used. In methods like Chi-Square, its influence surpasses that of other social indicators, such as drug-related cases and theft-related cases. Despite its relevance, the feature ranks lower than broader socioeconomic metrics such as urbanization rate and migration share, suggesting that structural factors might play a more central role in divorce prediction. The moderate ranking in certain methods, despite its theoretical association with social issues, could be linked to cultural variations in reporting and significance of alcohol consumption in marital disputes within the studied region.

Model and Feature Integration: Performance Evaluation and Selection

After identifying the sequences of impactful features, these attributes undergo classification using various algorithms. As indicated in subsection 5.1, it becomes evident that both the Random Forest and Neural Network algorithms outperform the others in terms of performance.

In the initial phase, by applying the Chi-Square feature selection algorithm, we exclusively use the first

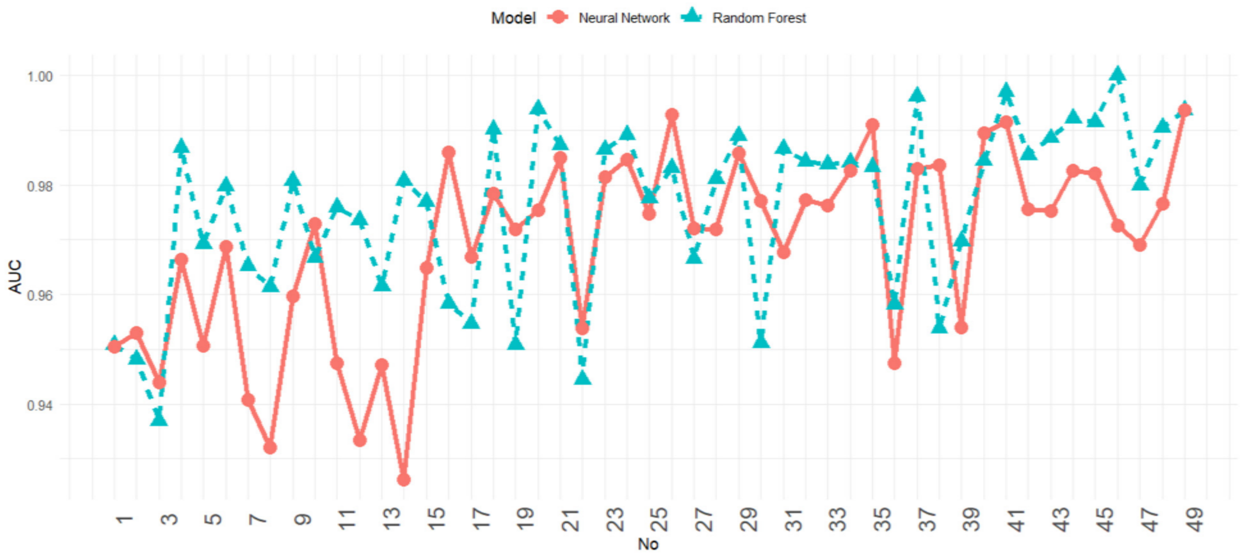


Figure 3. Feature count vs. AUC performance in Random Forest and Neural Network algorithms: unveiling the impact of feature selection

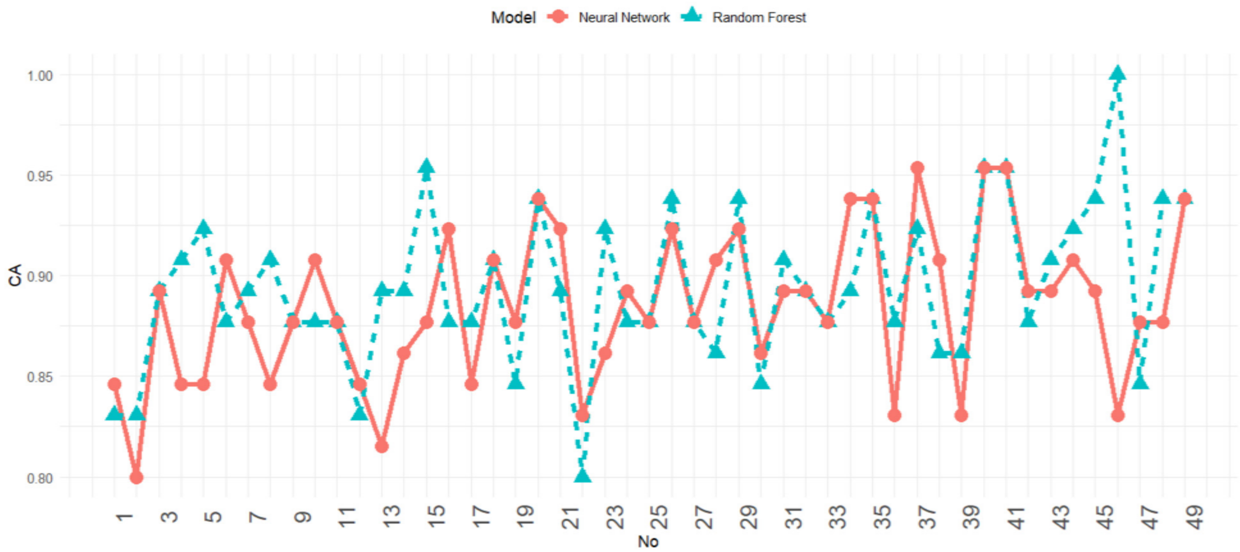


Figure 4. Feature count vs. Classification accuracy (CA) performance in Random Forest and Neural Network algorithms: unveiling the impact of feature selection

identified effective feature for modeling purposes, employing the Random Forest and Neural Network algorithms. Subsequently, we assess the accuracy based on the test data. This process is repeated for the first two features, followed by the first three features, and so on, until all the features have been incorporated. The outcomes of these evaluations are graphically presented in Figures 3 and 4. As shown, the highest success rate was achieved with the Random Forest algorithm, which attained an AUC and CA of 1. This success was achieved using the first 46 features.

Table 5 displays the optimal outcomes obtained through the implementation of the two aforementioned classification algorithms and the Chi-Square feature selection algorithm. Upon examining the table, it is clear that the first 46 features selected by the Chi-Square algorithm, classified by the Random Forest algorithm, yield the highest levels of AUC, CA, F1, Precision, and Recall.

Identifying Key Features through Importance Analysis

Feature importance is a fundamental aspect of data

Table 5. Confusion Matrix using the Three Machine Learning Models

No. of Features	Model	AUC	CA	F1	Precision	Recall
16	Neural Network	0.986	0.923	0.922	0.934	0.923
20	Random Forest	0.994	0.938	0.939	0.948	0.938
21	Neural Network	0.985	0.923	0.924	0.929	0.923
23	Random Forest	0.986	0.923	0.924	0.929	0.923
26	Neural Network	0.993	0.923	0.922	0.926	0.923
29	Random Forest	0.989	0.938	0.938	0.938	0.938
29	Neural Network	0.986	0.923	0.922	0.926	0.923
35	Neural Network	0.991	0.938	0.939	0.94	0.938
37	Random Forest	0.996	0.923	0.921	0.932	0.923
40	Random Forest	0.984	0.954	0.952	0.957	0.954
40	Neural Network	0.989	0.954	0.954	0.956	0.954
41	Random Forest	0.997	0.954	0.954	0.96	0.954
41	Neural Network	0.991	0.954	0.954	0.954	0.954
43	Random Forest	0.989	0.908	0.909	0.924	0.908
44	Random Forest	0.992	0.923	0.922	0.922	0.923
45	Random Forest	0.991	0.938	0.937	0.944	0.938
46	Random Forest	1	1	1	1	1
48	Random Forest	0.99	0.938	0.938	0.943	0.938
49	Random Forest	0.994	0.938	0.939	0.94	0.938
49	Neural Network	0.994	0.938	0.939	0.948	0.938

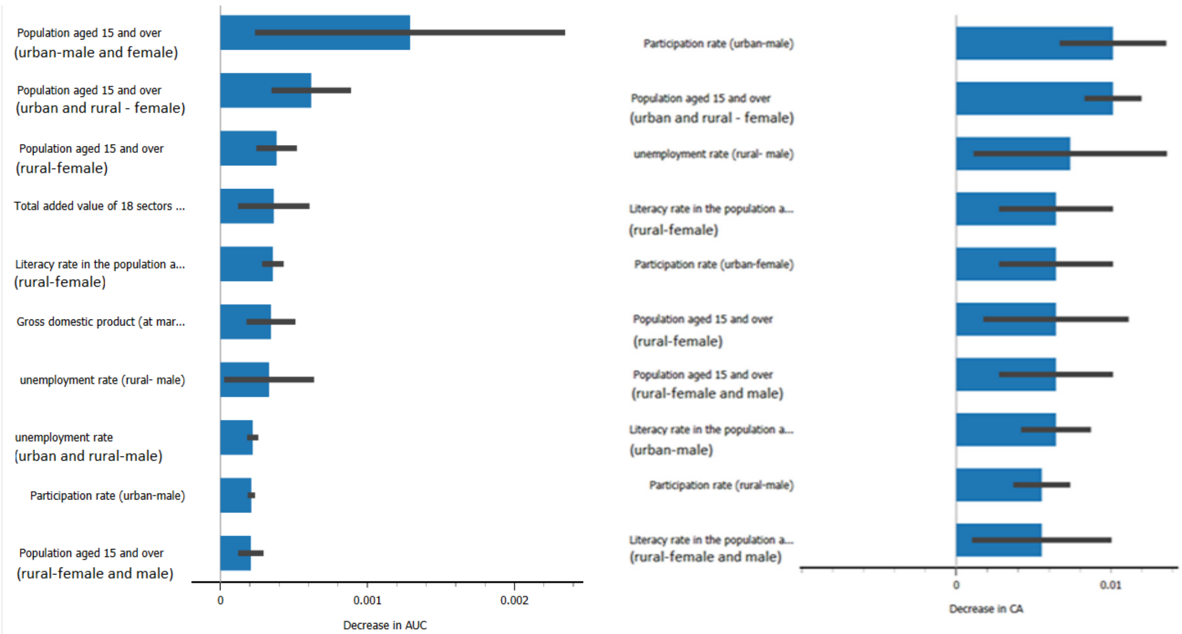


Figure 5. Features importance based on all AUC and CA scores in the Random Forest algorithm (The first 46 features obtained in subsection 5.3 are applied.)

mining algorithms. It refers to the quantification of the contribution each individual feature makes to the predictive accuracy of the model. This measure not only helps in understanding the underlying relationships within the data but also aids in feature selection, allowing the identification of key attributes that have the most substantial influence on the model's performance.

The results of the feature importance assessment are presented in Figure 5. In this analysis, the provided dataset serves as the basis for computing the significance of each individual feature with respect to predictions. This is achieved by quantifying the increase in the model's prediction error after permuting the values of a specific feature. This permutation effectively severs the inherent connection between the feature and the target, enabling the determination of its true impact on prediction accuracy.

Based on the AUC scores, the most important feature is the percentage of the urban population in the province aged 15 and above. This factor directly correlates with the divorce rate. Therefore, it can be concluded that urbanization has a direct impact on the divorce rate. In other words, the cultural differences between urban and rural lifestyles in the country could lead to variations in divorce rates. Shifting focus from urbanization and rural living to gender, provinces with a higher percentage of women residents tend to have higher divorce rates. Other influential attributes include the unemployment rates in urban and rural areas, as well as the literacy rate among rural women.

Discussion

This paper presents a comprehensive comparison of ten classification algorithms, including Neural Network, Naïve Bayes, Multinomial Logistic Regression, AdaBoost, GraBoost, Random Forest, Decision Tree, kNN, SGD, and SVM, for predicting the 'Divorce Category' attribute with labels 'Low', 'Medium', and 'High'. The experimental results clearly indicate that Random Forest and Neural Network outperform the other algorithms when applied to the divorce dataset. This conclusion is based on rigorous evaluation using 10-fold cross-validation. The findings of this study have significant implications for law enforcement agencies, highlighting the potential advantages of utilizing machine learning algorithms like Random Forest to effectively address divorce-related issues. Notably, the performance of feature selection algorithms has a more positive and favorable impact compared to using all features. This observation is particularly important given the extensive use of numerous features in existing studies on divorce volume diagnosis. Identifying truly effective features has been a persistent challenge, and this study makes valuable contributions to addressing this issue.

Looking ahead, future research plans involve applying spatiotemporal classification algorithms to the divorce dataset and evaluating their prediction performance specifically for Iranian provinces. Additionally, exploring alternative techniques for feature selection and investigating their effects on the prediction performance of different algorithms represent promising avenues for further exploration in this area.

While this study aimed to predict the level of court cases related to divorce using data mining models, the limited quantity of data available may have affected the precision of our models, particularly in predicting the different levels of divorce. To improve the accuracy of such models for categorical variables with multiple levels, future researchers should focus on managing and collecting a larger and more diverse dataset of historical divorce cases, ensuring a sufficient number of cases for each level of the categorical variable. By doing so, they can better train and validate their models, leading to more precise predictions.

References

1. Rosili NAK, Zakaria NH, Hassan R, Kasim S, Rose FZC, Sutikno T. A systematic literature review of machine learning methods in predicting court decisions. IAES Int J Artif Intell. 2021;10(4):1091.
2. Narendran DJ, Abilash R, Charulatha BS. Exploration of classification algorithms for divorce prediction. In: Proceedings of the International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2020. Singapore: Springer; 2021. p. 291-303.
3. Sharma A, Chudhey AS, Singh M. Divorce case prediction using machine learning algorithms. In: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). 2021. p. 214-9.
4. Amato PR. Research on divorce: Continuing trends and new developments. J Marriage Fam. 2010;72(3):650-66.
5. Lundberg S. The added worker effect. In: The Oxford Handbook of Women and the Economy. 2017. p. 316-35.
6. Raley RK, Bumpass LL. The topography of the divorce plateau: Levels and trends in union stability in the United States after 1980. Demogr Res. 2003;8(8):245-60.
7. Musick K, Bumpass L. Reexamining the case for marriage: Union formation and changes in well-being. J Marriage Fam. 2012;74(1):1-18.
8. Leopold T. Gender differences in the consequences of divorce: A study of multiple outcomes. Demography. 2018;55(3):769-97.
9. Afthanorhan A, Mahmud A, Sapri A, Aimran N, Aireen A, Rambli A. Prediction of Malaysian women divorce using machine learning techniques. Malays J Comput. 2022;7(2):1149-61.
10. Acharya UR, Molinari F, Sree SV, Chattopadhyay S, Ng KH, Suri JS. Automated diagnosis of epileptic EEG using entropies. Biomed Signal Process Control. 2012;7(4):401-8.
11. Ahsan MM. Divorce Prediction with Machine Learning: Insights and LIME Interpretability. arXiv preprint arXiv:2310.08620. 2023.
12. Fareed MMS, Khan MA, Javed K, et al. Predicting Divorce Prospect Using Ensemble Learning: Support Vector Machine, Linear Model, and Neural Network. Comput Intell Neurosci. 2022;2022:3687598.
13. Moumen A, Shafqat A, Alraqad T, Alshawarbeh ES, Saber H, Shafqat R. Divorce prediction using machine learning algorithms in Ha'il region, KSA. Scientific Reports. 2024 Jan 4;14(1):502.
14. Arpino B, Le Moglie M, Mencarini L. What tears couples apart: A machine learning analysis of union dissolution in Germany. Demography. 2022 Feb 1;59(1):161-86.