

# Análise de Dados Sob a Perspectiva da Classificação: Desvendando Insights sobre Divórcio Através da Mineração de Dados

Data Analysis from the Classification Perspective: Unveiling Insights on Divorce  
through Data Mining

Janaide Nogueira de Sousa Ximenes

Instituto Federal do Ceará  
Rua Luiz Cunha, 178 - Monte Castelo -  
62.350-000,  
Ubajara – CE – Brasil  
nogueirajanaide@gmail.com

Rhyan Ximenes de Brito

Instituto Federal do Ceará  
CE-187, s/n - Estádio - 62.320-000,  
Tianguá – CE – Brasil  
rxbrito@gmail.com 3rd Author name

## ABSTRACT

This article presents an analysis of divorces using data mining techniques, employing statistical classifiers and artificial neural networks. The objective is to explore and understand the factors influencing divorce occurrences, as well as to develop classification models capable of identifying patterns and predicting future divorces. The obtained results are analyzed and compared, considering performance metrics such as accuracy, precision, recall, and F1-score. The results achieved outperformed previous works using the same database.

## Keywords

Statistical Methods; Analysis of Electoral Data; Electoral Data.

## RESUMO

Este artigo apresenta uma análise dos divórcios por meio de técnicas de mineração de dados, utilizando classificadores estatísticos e redes neurais artificiais. O objetivo é explorar e compreender os fatores que influenciam a ocorrência que levam ao divórcio, além de desenvolver modelos de classificação capazes de identificar padrões e prever divórcios futuros. Os resultados obtidos são analisados e comparados, levando em consideração métricas de desempenho, como acurácia, precisão, *recall* e *F1-score*. Os resultados obtidos alcançaram índices melhores que os trabalhos já realizados com a base de dados utilizada.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

## Palavras-Chave

Análise de dados; Divórcios; Mineração de dados.

## 1. INTRODUÇÃO

Segundo [14] o Brasil atingiu o recorde de 386,8 mil divórcios no ano de 2021 e se o valor for comparado com o ano anterior, estes valores representam um aumento de 16,8%, ou seja, uma diferença de 55,6 mil divórcios.

De acordo com [7] o divórcio mantém sua posição como a segunda fonte mais intensa de estresse enfrentada por uma pessoa, ficando logo atrás da dor provocada pela perda de um filho ou cônjuge. Além disso, as dificuldades familiares e conjugais representam cerca de metade de todas as consultas realizadas por indivíduos em busca de apoio psicoterapêutico [7].

É importante ressaltar que o divórcio é considerado uma das experiências mais desafiadoras e estressantes que uma pessoa pode vivenciar, chegando perto do impacto emocional causado pela perda de um filho ou cônjuge devido à morte [2]. A família desempenha um papel fundamental na sociedade atual, independentemente de sua estrutura, sendo responsável por transmitir aprendizado, conhecimento e educação às crianças. No entanto, é evidente que o aumento significativo dos divórcios no Brasil tem gerado conflitos relacionados à guarda dos filhos, o que pode resultar em traumas duradouros na vida das crianças.

O objetivo deste estudo é realizar uma análise dos divórcios por meio da mineração de dados, com foco na utilização de classificadores estatísticos e redes neurais artificiais. Portanto, a pesquisa tem como objetivo explorar o desempenho e a eficácia de diferentes algoritmos de classificação, como a árvore de decisão, *Support Vector Machines* (SVM), *KNearest Neighbors* (KNN), *Multi-Layer Perceptron* (MLP), *Random Forest* (RF) e *Extreme Learning Machine* (ELM), na previsão e identificação de padrões nos dados sobre divórcio.

Além disso, busca-se utilizar diversas métricas de avaliação, como acurácia, precisão, *recall* e *F1-score*, para comparar e analisar o desempenho dos classificadores. O estudo tem como propósito contribuir para uma melhor compreensão dos fatores que influenciam os divórcios e fornecer *insights* que possam auxiliar na implementação de

políticas públicas e medidas preventivas relacionadas ao tema.

O artigo está organizado em seis seções, a contar da introdução. A seção 2 traz uma análise em trabalhos relacionados da literatura sobre o divórcio na perspectiva da Inteligência Artificial, na seção 3, foi apresentada a fundamentação teórica do trabalho, com uma revisão de literatura relacionada ao divórcio, e sobre o *Knowledge Discovery in Databases* (KDD). Já na seção 4 materiais e métodos, são expostas as ferramentas utilizadas, base de dados, pré-processamento e transformação e, por fim, a mineração dos dados, interpretação e evolução. Após isso, à seção 5 traz a discussão dos resultados obtidos. Por fim, as conclusões e trabalhos futuros na seção 6.

## 2. TRABALHOS RELACIONADOS

Nesta seção, aborda-se estudos anteriores que têm relação com o estudo proposto, reforçando a necessidade e a importância de pesquisas sobre as causas do divórcio entre casais. Esses estudos são fundamentais para a construção de medidas que auxiliem na compreensão e no aconselhamento familiar. A relevância dessas pesquisas é destacada para fornecer uma base sólida ao trabalho.

No trabalho realizado por [28], foi aplicada uma técnica chamada seleção de características baseada em correlação em conjunto com redes neurais artificiais para prever casos de divórcio. Utilizando o mesmo conjunto de dados utilizado neste trabalho e disponibilizado pelo UCI *Machine Learning*, obtiveram resultados da taxa de acurácia de 98,23% na classificação dos casos de divórcio com a *Radial Basis Function* (RBF). Entretanto eles não utilizaram nenhuma outra métrica de avaliação ou mesmo citaram a utilização de normalização ou de balanceamento da base.

Alguns anos após os autores [22] utilizaram algoritmos de *machine learning*, como *Perceptron*, *Naive Bayes*, K-NN, árvore de decisão, SVM e a regressão logística com diferentes proporções de *folds*, para treinar modelos de classificação capazes de identificar os casos de divórcio com base nas características disponíveis no mesmo conjunto de dados utilizado neste trabalho. Dentre os modelos utilizados, o modelo *perceptron* atingiu a acurácia de 98,5%.

Em [1], os autores propuseram um método baseado na *Divorce Predictors Scale* (DPS) de Gottman para prever divórcios. Utilizaram uma amostra de 200 participantes, sendo 126 casados e 74 divorciados. Aplicaram os algoritmos *Multilayer Perceptron* (MLP), *Naive Bayes* (NB) e *Random Forest* (RF), com técnicas de seleção de características para identificar as mais relevantes. O *Naive Bayes* alcançou 87,14% de precisão inicial, enquanto o MLP atingiu 84,29% após seleção de características. Os resultados indicam que a DPS é eficaz na previsão de divórcios, útil para planejamento de intervenções familiares.

Já [21], propuseram a utilização de um algoritmo de aprendizado de máquina otimizado (PSO) para eliminar recursos redundantes e melhorar a previsão da taxa de divórcio em um conjunto de dados padrão. Alcançando uma precisão de 99,67%, sugerindo que o modelo pode ser uma ferramenta útil para especialistas em aconselhamento familiar, promovendo o bem-estar psicológico dos casais.

A Tabela 1, apresenta uma síntese comparativa entre as taxas de acurácias obtidas nos trabalhos relacionados citados

juntamente com esta proposta. Nessa perspectiva deve-se salientar que os resultados de [28] relatam apenas a taxa de acerto relacionada a 98,23% com o classificador RBF, [22] 98,50% com a rede neural *Perceptron*, [1] 87,14% com o *Naive Bayes* e [21] com 99,672% com PSO. Percebe-se nos trabalhos citados que outras métricas de avaliação como: *recall*, *F1-score*, precisão, índice de *kappa* e curva média de ROC não foram mencionados. Por outro lado, pode-se salientar que na proposta deste trabalho foram utilizadas outras métricas, observando-se que os classificadores RF, MLP e SVM atingiram, respectivamente, 99,98%, 99,98% e 99,92% de precisão, proporcionando assim resultados mais significativos em comparação com os trabalhos dos autores [28], [22], [1], [21].

Table 1: Síntese Comparativa Trabalhos Relacionados

Autor(es)	Técnica Utilizada	Acurácia (%)
[28]	RBF	98,23%
[22]	<i>Perceptron</i>	98,50%
[1]	<i>Naive Bayes</i>	87,14%
[21]	PSO	99,67%
Nossa proposta	<i>Random Forest</i>	98,62%

Fonte: Elaboração própria.

## 3. FUNDAMENTAÇÃO TEÓRICA

Nesta seção, será apresentada uma revisão técnica e teórica dos temas que fundamentam este trabalho, organizando-se da seguinte maneira: a subseção 3.1 explora definições e pontos de vista relacionadas ao divórcio, a subseção 3.2 introduz a definição do *Knowledge Discovery in Databases* (KDD), em seguida apresenta de forma resumida vantagens, desvantagens e aplicações dos algoritmos utilizados no trabalho em questão.

### 3.1 Divórcio

O casamento, apesar de ainda ser um sonho para muitos, enfrenta uma tendência de declínio conforme as estatísticas recentes mostram. Matrimônios civis são escolhidos por aqueles que desejam construir suas vidas com seus entes queridos, mas nem sempre esse projeto se concretiza, resultando frequentemente em dissoluções desde a legalização do divórcio no Brasil, em 1977 [18].

O divórcio é uma experiência complexa, que envolve rupturas, conflitos e repercussões que afetam todo o sistema familiar. Considerando isso, a busca por alternativas que ajudem a manter relações saudáveis após a dissolução conjugal constitui um desafio para a sociedade atual [29]. Há diversos adjetivos utilizados para definir o divórcio, tais como trauma [13], crise de vida [24] e luto [26]. [3] classificaram o divórcio como uma experiência de vida psicologicamente desafiadora, que provoca confusão na vida dos indivíduos envolvidos e exige reorganização em várias áreas.

Por outro lado, o divórcio é uma forma de dissolução do casamento onde nada é fácil ou simples, mesmo quando consensual, pois envolve sofrimento e dor, ainda que tenha um sentido libertador. Porém o fim da conjugalidade

desembarca em um momento que se depara com o desamparo estrutural do ser humano [20].

Para [11] o divórcio pode ser entendido como uma medida dissolutiva do vínculo matrimonial válido entre pessoas, tendo como consequência a extinção de deveres conjugais.

Ainda conforme [11] a ação de desquite no Brasil fundamenta-se nos seguintes motivos: (i) adultério; (ii) tentativa de morte; (iii) maus-tratos ou injúria grave; e (iv) abandono voluntário do lar conjugal, durante dois anos contínuos.

Vale ressaltar que pessoas divorciadas experienciam menor bem-estar psicológico, piores níveis de felicidade e maiores índices de sintomatologia psicológica, nomeadamente depressão e ansiedade, quando comparadas às pessoas que permanecem casadas [16].

É importante destacar que os motivos pelos quais os casais se divorciam são variados e complexos, abrangendo questões emocionais, sociais, financeiras e culturais. As razões mais comuns incluem falta de comunicação, conflitos constantes, diferenças irreparáveis nos valores, infidelidade, problemas financeiros, dificuldades na criação dos filhos e desgaste do relacionamento. Esses fatores frequentemente criam um ambiente de tensão e insatisfação, tornando a vida conjugal insustentável [8].

Em vista disso, a Tabela 2 mostra a cronologia do divórcio no Brasil evidenciando a evolução ocorrida durante as décadas como forma de compreensão da evolução sobre o processo de separação.

Mediante o exposto pode inferir que a legislação brasileira sobre o divórcio evoluiu de um modelo rígido e conservador para um sistema mais flexível e adaptado às necessidades contemporâneas, refletindo a adaptação às transformações sociais e valorizando a autonomia individual, tornando o processo de dissolução conjugal mais acessível, rápido e justo.

### 3.2 Knowledge Discovery in Databases (KDD)

Mineração de dados, também conhecida como descoberta de conhecimento em bancos de dados é o processo de explorar grandes volumes de dados em busca de padrões, tendências, relações e informações úteis que possam ser transformadas em conhecimento valioso para tomada de decisões e solução de problemas.

Dessa forma, a mineração de dados envolve a aplicação de técnicas e algoritmos de análise de dados em conjuntos de dados complexos, geralmente armazenados em bancos de dados ou sistemas de informação. Esses conjuntos de dados podem ser estruturados, semiestruturados ou não estruturados, e podem incluir uma ampla variedade de informações, como dados transacionais, textos, imagens, áudio, vídeos, entre outros [9].

Vale ressaltar que o principal objetivo da mineração de dados é descobrir padrões e informações ocultas nos dados, que podem não ser aparentes à primeira vista. Isso pode incluir identificar tendências de mercado, prever comportamentos futuros, segmentar clientes, otimizar processos, detectar fraudes, entre outros aspectos relevantes para diferentes áreas de aplicação.

Dessa forma como observado na Figura 1, o KDD possui diversas etapas que tem como objetivo extrair conhecimento útil a partir de grandes volumes de dados, onde cada fase possui um propósito específico.

Segundo [19] a seleção de dados refere-se à etapa em que são escolhidos os dados relevantes e adequados para a análise e construção de modelos de aprendizado de máquina. É uma etapa fundamental no processo de KDD, pois a qualidade dos dados selecionados influencia diretamente nos resultados e na eficácia das análises.

Consequentemente essa seleção pode ser feita de diferentes maneiras, como por critérios de inclusão/exclusão de atributos, amostragem aleatória, seleção de instâncias baseada em critérios específicos, entre outros [19].



Figure 1: Etapas do Processo KDD Fonte: Adaptado de [10]

O pré-processamento de dados na mineração de dados é uma etapa crucial que envolve a preparação e transformação dos dados brutos antes de aplicar técnicas de mineração de dados, sendo assim, o objetivo do pré-processamento de dados é melhorar a qualidade dos dados, torná-los adequados para análise e garantir que os algoritmos de mineração de dados possam extrair padrões e informações relevantes de maneira eficaz [10].

Já a etapa de transformação de dados envolve uma série de atividades que visam modificar os dados de entrada de forma a prepará-los para a aplicação dos algoritmos para o processo de mineração de dados. Sendo assim, nesta etapa é possível, dentre outras atividades normalizar e padronizar os dados, tratar dados inconsistentes, criar novos atributos derivados dos atributos originais e outras [19].

Durante a etapa de mineração de dados há a aplicação de algoritmos e técnicas para explorar os dados em busca de padrões, relacionamentos e informações úteis que possam ser utilizadas para tomada de decisões ou até mesmo a geração de conhecimento. Nesta etapa são realizados procedimentos, como por exemplo, a seleção da atividade, a seleção das técnicas de mineração, a configuração dos parâmetros, a aplicação das técnicas e outras.

Além disso, para a mineração de dados realizada neste trabalho foram utilizadas as Redes Neurais Artificiais (RNAs): *Multilayer Perceptron* (MLP), *Extreme Learning Machine* (ELM) e os classificadores estatísticos: *Support Vector Machine* (SVM), *Arvore de Decisão* (AD), *K-Nearest Neighbors* (K-NN) e *Random Forest* (RF). Estes foram selecionados de acordo com as características da base de dados e resultados gerados.

Nessa perspectiva sabe-se que as RNAs são modelos computacionais inspirados no funcionamento do cérebro humano, sendo compostas por um conjunto de unidades de processamento interconectadas, chamadas de neurônios artificiais, que trabalham em conjunto para resolver

problemas complexos de aprendizado e reconhecimento de padrões [15].

Os classificadores são algoritmos utilizados em mineração de dados para classificar ou categorizar dados com base em padrões estatísticos identificados, em outras palavras, são funções que utilizam como entrada os padrões desconhecidos e como saída as classes que estes padrões provavelmente pertencem. São ferramentas que realizam o reconhecimento de padrões em um aprendizado supervisionado. Assim, percebe-se que a diferença entre os classificadores está no método em que se criam as fronteiras de decisões, porém as diferentes técnicas para estimar as fronteiras de decisões têm em comum os seguintes objetivos:

Table 2: Cronologia do Divórcio no Brasil

1977	A Emenda Constitucional nº 9, de 28 de junho de 1977, instituiu o divórcio no Brasil, permitindo pela primeira vez a dissolução do vínculo conjugal por meio desta lei.
1988	A Constituição de 1988 reduziu os prazos para concessão do divórcio, tanto para separação judicial quanto para separação de fato.
1989	A Lei nº 7.841, de 17 de outubro de 1989, eliminou a proibição de divórcios sucessivos, permitindo que as pessoas se divorciem quantas vezes quiserem.
2007	A Lei nº 11.441, de 4 de janeiro de 2007, permitiu divórcios e separações consensuais em cartório, excluindo casos com filhos menores, incapazes ou mulheres grávidas.
2010	A Emenda Constitucional nº 66/2010 eliminou a necessidade de prévia separação judicial ou de fato para concessão do divórcio no Brasil.
2020	O Provimento nº 100/2020 autoriza divórcios on-line no Brasil com consentimento mútuo do casal e presença de advogado, excluindo casos com filhos menores, incapazes ou gestação da mulher.

Fonte: Adaptado de [18]

(i) minimizar o erro de classificação; e (ii) permitir que a classificação seja eficiente computacionalmente otimizando o processo fazendo com que leve o menor tempo [17], [4].

A Tabela 3 apresenta um comparativo entre as RNAs e os classificadores estatísticos utilizados, apontando as principais vantagens, desvantagens e as principais indicações de aplicações de acordo com os autores: [23], [15], [12], [19], [17] e [6].

Logo após a mineração de dados, há a interpretação e evolução que é a etapa de interpretação e evolução que possibilita retroalimentar o processo de KDD, contribuindo para o refinamento das etapas anteriores, como a seleção de dados, o pré-processamento e a escolha das técnicas que serão utilizadas na mineração dos dados. A análise dos resultados obtidos e a compreensão de seu impacto no domínio de aplicação podem direcionar ajustes e melhorias em todas as etapas do processo.

## 4. MATERIAIS E MÉTODOS

Este trabalho se caracteriza, em termos de natureza, como uma pesquisa aplicada, pois busca aplicar os conhecimentos e técnicas de mineração de dados e classificação em um contexto específico, no caso, a análise dos divórcios. Entretanto, em relação aos objetivos, a pesquisa é exploratória, pois busca explorar e descrever os padrões e fatores relacionados aos divórcios, bem como identificar as características e variáveis que estão associadas a esse fenômeno.

Em relação à abordagem, o estudo adota uma abordagem quantitativa, uma vez que utiliza métodos estatísticos e algoritmos de classificação, como SVM, K-NN, AD, MLP, RF e ELM, para analisar e interpretar os dados dos divórcios. Quanto ao procedimento, o trabalho é classificado como do tipo

experimental, pois envolve a aplicação dos diferentes classificadores estatísticos e redes neurais artificiais em um conjunto de dados de divórcios para observar e analisar seus desempenhos e resultados. Portanto, o trabalho pode ser caracterizado como uma pesquisa aplicada, exploratória, de abordagem quantitativa e com procedimento do tipo experimental.

### 4.1 Ferramentas utilizadas

No desenvolvimento deste projeto, foi empregada a linguagem de programação *Python*, versão 3.11.4 com a *Integrated Development Environment (IDE) Google Colaboratory*, abreviada

para Colab. Essa plataforma foi escolhida pela facilidade de acesso ao ambiente, sendo necessário somente um navegador Web, aproveitando o alto poder de processamento, além do acesso fácil a diversas bibliotecas que foram necessárias, como *pandas*, *NumPy* e *scikit-learn*.

Além disso, há possibilidade de utilização de *Graphics Processing Unit (GPUs)* e *Tensor Processing Units (TPUs)*, que podem acelerar o treinamento de modelos de aprendizado de máquina. A linguagem foi definida pela simplicidade, legibilidade, flexibilidade e a grande quantidade de bibliotecas disponíveis para aplicações de mineração de dados.

### 4.2 Base de dados

Para esta pesquisa, foi utilizado uma base de dados pública disponível no repositório UCI, [27]. Esse conjunto de dados contém informações essenciais para a previsão de casos de divórcio e é composto por respostas de 170 casais a um conjunto de 54 perguntas. As perguntas seguem a teoria do casal de Gottman com base na DPS, as respostas foram classificadas em uma escala de 0 a 4. Com isso, as respostas a essas perguntas são utilizadas como características pelo algoritmo de previsão, que é treinado para fazer previsões com base nelas.

Sendo assim, a saída é fornecida na forma de 0 e 1, onde o valor 1 representa divorciado e o valor 0 representa não divorciado. E as respostas foram coletadas em uma escala de 5 pontos (0 = Nunca, 1 = Raramente, 2 = Em média, 3 = Frequentemente, 4 = Sempre). As perguntas são do tipo: O tempo que passo com minha esposa é especial para nós? Ou mesmo, não temos tempo em casa como parceiros?

### 4.3 Pré-processamento e Transformação

Na etapa relacionada ao pré-processamento fez-se necessária a exclusão da coluna referente ao índice, já que ele não contribui e pode ser considerado um ruído para os modelos utilizados. Considerando que a base de dados possui 84 sáidas referentes ao valor 1 e 86 referentes ao valor 0, fez-se necessário a aplicação do *Synthetic Minority Over-sampling Technique* (SMOTE). Este método consiste em lidar com o desbalanceamento de classes em conjuntos de dados, gerando amostras sintéticas da classe minoritária para equilibrar as classes.

Embora o SMOTE seja mais utilizado em bases de dados em que o desbalanceamento de classes nos conjuntos de dados é significativamente grande. O seu uso foi justificado, pois foi realizado o teste com e sem o SMOTE e a acurácia aumentou em uma média de 0,2% considerando todas as técnicas de seleção e treinamento utilizadas [5].

todas as características em uma escala comum, o que facilita a comparação e o processamento pelos algoritmos de mineração de dados.

### 4.4 Mineração, Interpretação e Evolução

Para o processo de treinamento e avaliação foi utilizada a estratégia de validação cruzada *k-fold*, com 10 *folds* ( $k=10$ ), outras configurações foram testadas, entretanto essa obteve os melhores resultados. Vale ressaltar que a estratégia de validação cruzada foi aplicada por ser fundamental e permitir estimar o desempenho do modelo em dados não vistos e reduzir o viés na estimativa de desempenho. A validação cruzada é uma técnica que permite avaliar o desempenho do modelo usando todo o conjunto de dados disponível, de forma mais robusta e confiável.

Em vez de simplesmente dividir o conjunto de dados em conjuntos de treinamento e teste, a validação cruzada divide o conjunto de dados em várias partes (*folds*) e realiza várias iterações de treinamento e teste. Cada parte dos dados é utilizada

Table 3: Comparativo Entre Algoritmos e Modelos Utilizados

Algoritmo	Vantagens	Desvantagens	Aplicações
SVM	Bom desempenho em conjuntos de dados pequenos a médios, lida bem com alta dimensionalidade, pode lidar com problemas lineares e não-lineares.	Sensível à escolha do kernel, requer ajuste de parâmetros, não é adequado para grandes conjuntos de dados.	Classificação, regressão e detecção de anomalias.
Arvore de Decisão	Lida com atributos categóricos e numéricos, não requer pré-processamento intenso dos dados.	Tendência ao <i>overfitting</i> , sensível a pequenas variações nos dados de treinamento.	Classificação, regressão e análise de decisões.
K-NN	Simplicidade de implementação, adapta-se a padrões complexos, não requer treinamento prévio.	Sensível à escala dos atributos, custoso em termos de memória e tempo de busca.	Classificação, regressão, recomendação e detecção de anomalias.
MLP	Capaz de lidar com problemas não-lineares complexos, aprendizado com base em gradientes, capacidade de generalização.	Requer ajuste de hiperparâmetros, sensível à inicialização dos pesos, pode ser propenso a <i>overfitting</i> com conjuntos de dados pequenos.	Classificação, regressão, reconhecimento de padrões e processamento de sinais.
ELM	Treinamento rápido, eficiente em termos computacionais, boa generalização, não requer ajuste de parâmetros iterativo.	Menor interpretabilidade comparado a outros algoritmos, pode ter desempenho inferior em problemas com dados desbalanceados.	Classificação, regressão, reconhecimento de padrões e análise de dados.
Random Forest	Bom desempenho em conjuntos de dados grandes, lida com atributos numéricos e categóricos, reduz <i>overfitting</i> através de bagging e aleatorização.	Pode ser lento em tempo de treinamento e classificação, menor interpretabilidade comparado a uma única árvore de decisão.	Classificação, regressão, detecção de anomalias e seleção de atributos.

Fonte: Elaboração própria.

Com os dados pré-processados, foi utilizada a técnica *StanderScaler* que é extremamente útil em situações, como essa, onde os atributos do conjunto de dados têm diferentes escalas e variabilidades. A padronização dos dados é uma etapa importante do pré-processamento, pois ajuda a colocar

tanto para treinamento quanto para teste em diferentes iterações, garantindo que todas as amostras sejam utilizadas para avaliação [25].

Visando obter os melhores resultados possíveis, os hiperparâmetros foram definidos de acordo com os testes

realizados e os resultados gerados, sendo os melhores resultados obtidos estão representados na Tabela 4.

Além disso, fez-se necessário a utilização de algumas métricas de avaliação dos modelos, foram calculadas:

- **Acurácia:** medir a proporção de previsões corretas em relação ao total de previsões feitas pelo modelo.
- **Precisão:** métrica que mede a proporção de previsões corretas de uma classe específica em relação ao total de previsões positivas feitas para essa classe.
- **Recall:** também conhecido como sensibilidade ou taxa de verdadeiros positivos, mede a proporção de amostras positivas corretamente identificadas em relação ao total de amostras positivas.
- **F1-score:** métrica que combina precisão e *recall* em uma única medida resumida.
- **Curva de ROC:** fornece uma medida da capacidade discriminativa do modelo em diferentes pontos de corte de probabilidade

taxa de acerto. Em relação a métrica de precisão o RF (99,98%) e a MLP (99,98%) apresentaram os melhores resultados, indicando uma alta precisão nas previsões positivas.

Entretanto na métrica *recall*, o algoritmo K-NN obteve o maior resultado (98,97%), indicando uma boa capacidade de identificar corretamente as instâncias positivas. Já no *F1-score*, novamente, o algoritmo K-NN apresentou o maior resultado (97,26%), indicando um bom equilíbrio entre precisão e *recall*. Em relação a área sob a curva ROC média, o algoritmo K-NN obteve a maior área sob a curva ROC média (0,97), indicando um bom poder de discriminação. Na medida *kappa* o algoritmo K-NN obteve o maior valor de *kappa* (0,98%), indicando uma concordância substancial.

## 6. CONCLUSÕES TRABALHOS FUTUROS

No geral, os resultados são bastante positivos, com desempenho satisfatório em todas as métricas avaliadas. O algoritmo *Random Forest* se destaca em várias métricas, como acurácia, precisão e *F1-score*. O algoritmo K-NN obteve um bom desempenho em várias métricas, incluindo *recall*, *F1score*, área

Table 4: Hiperparâmetros Utilizados Após a Interpretação e Evolução

RNAs e Classificadores	Hiperparâmetros
ELM	- Função de ativação: Sigmoidal; - Método de inicialização dos pesos: aleatórios entre -1 e 1; - Regularização: L1 regularization (Lasso); - Tamanho da camada oculta: 20
RF	- Quantidade de árvores: 100; - Profundidade máxima: 10; - Número mínimo de amostras para nó interno: 2; - Número mínimo de amostras para ser um nó folha: 1;
AD	- Profundidade máxima da árvore: 10 níveis; - Número mínimo de amostras para dividir um nó interno: 2; - Número mínimo de amostras para ser um nó folha: 1; - Semente: Aleatória;
MLP	- Função de ativação: Logística; - Otimização de pesos: SGD; - Taxa de aprendizado: Adaptativa; - Otimização de pesos: SGD; - Máximo de interações: 1000;
K-NN	- Número de vizinhos: 5; - Atribuição de pesos aos vizinhos: auto; - Métrica para distância: Euclidiana;
SVM	- Kernel: Linear; - Parâmetro de regularização: 1.0;

Fonte: Elaboração própria.

- **Kappa:** medida estatística que avalia a concordância entre dois observadores ou a concordância entre o modelo de previsão e os dados observados.

## 5. RESULTADOS E DISCUSSÕES

A Tabela 5 apresenta os resultados obtidos a partir das RNAs e dos classificadores estatísticos utilizados com seus respectivos hiperparâmetros expostos na Tabela 2. É importante destacar que os melhores desempenhos foram alcançados através dos algoritmos RF (98,62%) e MLP (97,80%), indicando uma alta

sob a curva ROC média e *kappa*. No entanto, é importante analisar os resultados em conjunto com o contexto e os requisitos específicos do problema para selecionar o algoritmo mais adequado. Como trabalhos futuros sugere-se utilizar, além dos classificadores estatísticos e redes neurais artificiais mencionados, explorar o uso de outros algoritmos de aprendizado de máquina, como redes Bayesianas, redes neurais convolucionais (CNN) ou redes neurais recorrentes (RNN), para analisar e classificar os divórcios.

## 7. REFERENCES

- [1] M. Ashrafujjaman Tutul, M. M. Hasan, S. K. Mondol, M. Hossain, and A. Al Marouf. Divorce prediction using machine learning methods-bangladesh perspective. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–5, 2021.
- [2] C. A. d. S. Bueno. As relações amorosas atuais sob ótica da terapia cognitivo comportamental de casais. 2021.
- [3] B. O. Cabilar and A. E. Yilmaz. Divorce and post-divorce adjustment: Definitions, models and assessment of adjustment. *Psikiyatride Guncel Yaklasimlar*, 14(1):1–11, 2022.
- [4] P. H. R. Cerqueira. *Um estudo sobre reconhecimento de padrões: um aprendizado supervisionado com classificador bayesiano*. PhD thesis, Universidade de São Paulo, 2010.
- [5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
- [6] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.
- [7] F. M. Dattilio. *Manual de terapia cognitivo-comportamental para casais e famílias*. Artmed Editora, 2011.
- [8] M. d. D. S. de Loreto and K. R. P. S. Raposo.
- [9] M. Farias, R. Gusmão, and C. Gusmão. Mineração de dados aplicada à saúde mental de estudantes universitários: Uma revisão sistemática. *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 49–59, 2020.
- [10] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37, 1996.
- [11] P. S. Gagliano and R. Pamplona Filho. *O novo divórcio*. Editora Saraiva, 2016.
- [12] J. Han, J. Pei, and H. Tong. *Data mining: concepts and techniques*. Morgan kaufmann, 2022.
- [13] S. J. Herman. Divorce: A grief process. *Perspectives in Psychiatric Care*, 1(2):108–112, 1974.
- [14] Instituto Brasileiro de Geografia e Estatística. IBGE Registro civil - 2020 e 2021. <https://sidra.ibge.gov.br/pesquisa/registrocivil/quadros/brasil/2021>, 2023. Acesso em 23 de junho de 2023.
- [15] Z. L. Kovács. *Redes neurais artificiais*. Editora Livraria da Física, 2002.
- [16] D. Lamela, B. Figueiredo, and A. Bastos. Adaptação ao divórcio e relações coparentais: contributos da teoria da vinculação. *Psicologia: Reflexão e Crítica*, 23:562–574, 2010.
- [17] S. d. Nascimento Junior et al. Análise das correlações entre as métricas de avaliação de desempenho de classificadores multirrótulo. 2020.
- [18] T. Nigri. *Divórcio*. Editora Blucher, 2022.
- [19] A. A. d. Paula et al. Mineração de dados para análise de desempenho de alunos do ensino fundamental. 2022.
- [20] R. d. C. Pereira. *Divórcio-teoria e prática*. Saraiva Educação SA, 2017.
- [21] P. Ranjitha and A. Prabhu. Improved divorce prediction using machine learning- particle swarm optimization (pso). In *2020 International Conference for Emerging Technology (INCET)*, pages 1–5, 2020.
- [22] A. Sharma, A. S. Chudhey, and M. Singh. Divorce case prediction using machine learning algorithms. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pages 214–219. IEEE, 2021.
- [23] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni. Support vector machine versus random forest for

Classificadores	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)	Area sob a Curva Média de ROC	Kappa (%)
ELM	96,47	98,88	93,97	96,19	0.98	92,75
Random Forest	98,62	99,98	97,25	98,56	0.98	96,33
Arvore de Decisão	97,09	96,88	98,25	97,38	0.97	94,02
MLP	97,80	99,98	94,11	96,96	0.97	94,27
KNN	97,12	98,97	94,82	97,26	0.97	98,92
SVM	97,12	99,92	94,82	97,26	0.97	88,11

Table 5: RNAs e Classificadores Utilizados e Seus Resultados

Fonte: Elaboração própria.

Divórcio, conjugalidade e parentalidade: revisão sistemática e metanálise com foco nos instrumentos de apoio à resolução dos conflitos familiares. *Caderno Eletrônico de Ciências Sociais*, 11(2):52–77, 2023.

- [24] R. S. Wiseman. Crisis theory and the process of divorce. *Social Casework*, 56(4):205–212, 1975.
- [25] T.-T. Wong and P.-Y. Yeh. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8):1586–1594, 2019.
- [26] S. Yárnoz-Yaben. Grief due to divorce: Relationship with attachment style and effects on subjective well-being and co-parenting/el duelo ante el divorcio: Relación con el estilo de apego y efectos en el bienestar subjetivo y el ejercicio de la co-parentalidad. *Studies in Psychology*, 38(3):667–688, 2017.
- [27] A. K. Yntem, Mustafa and S. Klarlan. Divorce Predictors data set. UCI Machine Learning Repository, 2019. DOI: <https://doi.org/10.24432/C52W27>.
- [28] M. K. Yontem, A. Kemal, T. Ilhan, and S. Kiliçarslan. Divorce prediction using correlation-based feature selection and artificial neural networks. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 9(1):259–273, 2019.

- [29] E. P. Zordan, A. Wagner, and C. Mosmann. O perfil de casais que vivenciam divórcios consensuais e litigiosos: uma análise das demandas judiciais. *Psico-usf*, 17:185–194, 2012.