

Previsor de Parecer para Ética Médica Baseado em Aprendizado de Máquina

Carlos Eduardo Valareti

Mateus Cintra Costa

Carlos Henrique Valério de Moraes (Orientador)

Gerson Hiroshi Yoshinari Junior (Coorientador)

Resumo — Nos últimos anos, houve um aumento exponencial no interesse do uso de técnicas de aprendizado de máquina na área da saúde. Atualmente, os conselhos de medicina oferecem códigos de ética médica que devem ser seguidas pelos médicos no exercício de sua profissão, inclusive nas atividades relativas a ensino, pesquisa e administração de serviços de saúde, bem como em quaisquer outras que utilizem o conhecimento advindo do estudo da medicina. A avaliação ética destas atividades é complexa e deve ser realizada por conselhos regionais de medicina, demandando meses de análise e deliberação de comitês técnicos dessas análises. Desta forma, automatizar a obtenção desse parecer ético melhoraria significativamente a fiscalização desses órgãos e garantiria qualidade de serviços a população. Assim, foi empregada um conjunto de técnicas de aprendizado de máquina associada a métodos de processamento de linguagem natural, para prever o parecer ético e simplificar o processo. Os resultados foram promissores mesmo usando uma base de dados reduzida, obtida do site do Conselho Regional de Medicina do Estado de São Paulo (CREMESP).

Palavras-Chave — Código de Ética Médica, Parecer Ético, Aprendizado de Máquina, Processamento de Linguagem Natural, Classificação de Texto.

Abstract — In recent years, there has been an exponential increase in interest in the use of machine learning techniques in healthcare. Currently, medical councils offer codes of medical ethics that must be followed by physicians in the exercise of their profession, including in activities related to teaching, research and administration of health services, as well as in any other activities that use the knowledge arising from the study of medicine. The ethical assessment of these activities is complex and must be carried out by regional medical councils, requiring months of analysis and deliberation by technical committees of these analyses. In this way, automating the obtaining of this ethical opinion would significantly improve the inspection of these bodies and guarantee the quality of services to the population. Thus, a set of machine learning techniques associated with natural language processing methods was used to predict the ethical opinion and simplify the process. The results were promising even using a reduced database, obtained from the website of the Regional Council of Medicine of the State of São Paulo (CREMESP).

Keywords — Medical Ethics Code, Ethical Opinion, Ma-

chine Learning, Natural Language Processing, Text Classification.

I. INTRODUÇÃO

A ética pode ser entendida como um grupo de regras de conduta ou hábitos julgados válidos na sociedade, retratando o acordo entre a consciência e os preceitos consagrados. Assim, ela traz em seu conteúdo uma ideia que seu conhecimento se faz no processo de enculturação do grupo social, considerando os valores de formação da conduta humana. A ética médica, também chamada de Bioética, se coloca entre os valores de visão mais profundos, sendo oferecidos a grupos selecionados da sociedade moderna [1], com documentação ampla e de fácil acesso aos profissionais da medicina e população em geral.

Os dados contidos nos pareceres de ética médica que são coletados rotineiramente, estão aumentando em escala exponencial, infelizmente a maioria dessas informações ainda não é usada em modelos de previsão estatísticos de forma que os médicos possam utilizar para melhorar suas análises de situações que envolvem a ética médica. Com isso, tais esforços, se bem sucedidos, poderiam proporcionar grandes benefícios não apenas para a segurança e qualidade do atendimento ao paciente, mas também na redução dos custos de saúde [2].

As abordagens de aprendizado de máquina funcionam bem na prática, a manutenção destes recursos digitais por especialista é necessária frequentemente, em especial quando expostos a problemas complexos, como por exemplo a identificação de condições médicas de pacientes a partir da textura de imagens médicas [3].

Atualmente, as aplicações mais bem sucedidas da utilização de aprendizado de máquina na área médica, tem sido na análise de exames por imagem. Algoritmos de aprendizado são capazes de realizar detecção automática e precisa de diagnósticos de retinopatia diabética, câncer de pele e exames de fundo de retina [4]. O potencial do aprendizado de máquina para desvendar padrões complexos e sutis em imagens sugere que essas técnicas podem ser úteis em outras áreas da medicina. Desafios substanciais devem ser abordados, no entanto, antes que o aprendizado profundo possa ser aplicado de forma mais ampla [5].

A. Objetivos

Esta pesquisa visa desenvolver um classificador inteligente para documentos, visando identificar procedimentos médicos

que podem ferir alguma norma do código de ética médica. Este classificador será obtido através de metodologias de aprendizado de máquina (*Machine Learning* - ML) associadas a processamento de linguagem natural para os textos (*Natural Language Processing* - NLP) para extração de padrões, permitindo automatizar o processo de parecer ético de forma simples e rápida.

II. FUNDAMENTAÇÃO TEÓRICA

Para o desenvolvimento deste trabalho alguns focos de estudo foram levantados. A obtenção de modelos matemáticos para a classificação de texto nos leva a extração de informação de linguagem natural, em seguida do uso de aprendizado de máquina para a identificação de padrões nesses textos decompostos em conjuntos numéricos. Desta forma, são levantados os estudos dos próximos tópicos.

A. Processamento de Linguagem Natural (NLP)

O NLP é a abordagem computadorizada para analisar texto que se baseia tanto em um conjunto de teorias quanto em um conjunto de tecnologias [6]. Este método emprega técnicas computacionais com o objetivo de aprender, entender e produzir conteúdo de linguagem humana. As primeiras abordagens computacionais para a pesquisa de linguagem se concentraram na automatização da análise da estrutura linguística da linguagem e no desenvolvimento de tecnologias básicas, como tradução automática, reconhecimento de fala e síntese de fala. Os esforços de pesquisadores hoje em dia se concentram em refinar e fazer uso dessas ferramentas em aplicativos do mundo real, criando sistemas de diálogo falado e mecanismos de tradução de fala para fala, minerando mídias sociais para obter informações sobre saúde ou finanças e identificando sentimentos e emoções em relação a produtos e serviços [7].

B. Aprendizado de Máquina

Nos últimos anos, as redes neurais avançaram em um ritmo notável e encontraram aplicações práticas em várias indústrias [4]. As redes neurais definem entradas para saídas por meio de um complexo sistema de composição de camadas que apresentam blocos de construção, incluindo transformações não-lineares e funções [8]. Hoje, algoritmos de aprendizado de máquina são ferramentas que podem resolver problemas que dificilmente são solucionáveis com os métodos tradicionais. O aprendizado de máquina pode utilizar informações não rotuladas durante o treinamento, portanto, temos um método adequado para lidar com informações e dados heterogêneos [5].

Apesar do alto potencial dos dados disponíveis, escalar o desenvolvimento de modelos preditivos pode ser um grande desafio porque, técnicas tradicionais de modelagem preditiva requerem a cada resultado, a criação de um conjunto de dados personalizado com variáveis [9]. No entanto, é um consenso na literatura que 80% do esforço em uma análise preditiva é dividida em: pré-processamento, *merging*, customização e limpeza do conjunto de dados para obter os *insights* desejados

[10]. Contudo, este procedimento limita profundamente a escalabilidade dos modelos preditivos e o aprendizado de máquina surge como uma alternativa para este cenário [11].

O aprendizado de máquina permite que modelos computacionais de várias camadas de processamento, aprendam representações de dados com vários níveis de abstração. Este método encontra estruturas complexas em conjuntos massivos de dados, usando o algoritmo de retro-propagação para indicar como uma máquina deve alterar seus parâmetros internos, que são utilizados no cálculo de representação em cada camada a partir da representação na camada anterior. Redes convolucionais profundas trouxeram avanços no processamento de imagens, vídeo, fala e áudio, enquanto as redes recorrentes iluminaram dados sequenciais, como texto e fala [8].

III. DESENVOLVIMENTO

As metodologias de aprendizado de máquina supervisionado classificador necessitam para sua utilização uma base de dados pré-processada com tutor para identificar a classe de cada caso existente. Além disso, tendo uma ampla variedade de técnicas classificadoras atuais requer procedimentos de validação dos modelos matemáticos obtidos nos treinamentos, garantindo qualidade nos padrões selecionados da base de dados e capacidade de generalizar corretamente resultados para novos casos. Assim, serão apresentados todos esses procedimentos realizados nesta pesquisa.

A base de dados utilizada nesta análise, como citado previamente, foi extraída do banco de dados do Conselho Regional de Medicina do Estado de São Paulo (CREMESP) via *Web-scraping*. O principal objetivo desta técnica é extrair informações de um ou vários sites e processá-las em estruturas simples como planilhas, banco de dados ou arquivo CSV. Realizar a extração de dados de maneira manual, além de ser uma tarefa bastante complicada, consome recursos e tempo, no entanto, esta técnica mostrou-se bastante eficaz na montagem do banco de dados.

A. Base de Dados

A base de dados desta pesquisa foi extraída do banco de dados do site do Conselho Regional de Medicina do Estado de São Paulo (CREMESP) [12] via uma rotina de *Web-Scraping*. O *Web-Scraping* realiza uma varredura em um endereço eletrônico, minerando informações que atendem um conjunto de filtros pré-estabelecidos, exemplificado na Figura 1. Esse procedimento é complexo, pois deve ser otimizado por tipo de informação e site analisado, com o risco de sofrer banimento pelo próprio servidor de páginas Web devido ao acesso intenso e anormal dessas informações.

A mineração dos pareceres éticos da ferramenta de busca do CREMESP obteve 5.555 documentos. Em seguida, realizou-se a limpeza dos campos com pareceres inconclusivos. Como as informações contidas no banco foram inseridas manualmente e coletadas de uma gama bastante heterogênea de localidades do Estado de São Paulo, muitos pareceres estavam com dados incompletos. A base final de documentos foi reduzida para 502 ofícios após a filtragem. Ainda, mesmo com documentos com explicações amplas, muitos tópicos destes textos não

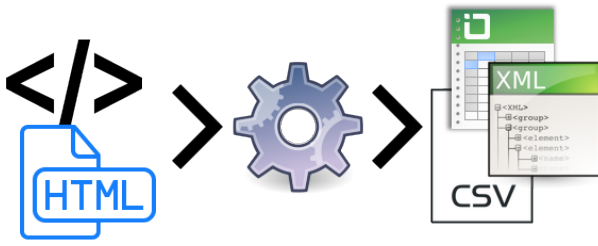


Figura 1. Procedimento simples de *Web-Scraping*.

tinham clareza da informação conclusiva do parecer ético, sendo que alguns documentos eram avaliados por itens, e não em uma conclusão geral. A base foi finalmente condensada a documentos com parecer claro e global por documento, isso reduziu mais ainda a base, chegando a apenas 49 documentos com pareceres claros e objetivos que foram classificados como éticos ou não éticos de forma manual, tendo como base o código de ética médica.

Portanto, a base de dados útil resultante era de apenas 0,9% da base original obtida pelo *Web-Scraping* no site da CREMESP. Desta forma, foi utilizada uma segunda base de dados para aglutinar outras informações ao treinamento do aprendizado de máquina: o código de ética médica. O código de ética médica também foi obtido do site da CREMESP [12] de forma direta, com cópia do seu conteúdo em página para um arquivo texto. Esse arquivo texto foi incorporado à base de dados como um simples ofício onde seu parecer era classificado ficticiamente como favorável. Esse procedimento de expansão da base de dados permitiu equilibrar as classes de pareceres éticos e não éticos.

Os treinamentos do aprendizado de máquina classificador deverão ser avaliados com as seguintes bases de dados (Figura 2):

- Simples: pareceres éticos;
- Expandida: mais código de ética médica.

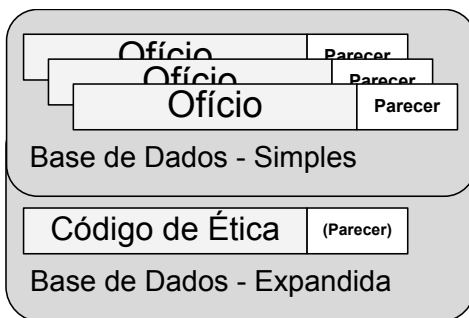


Figura 2. Base de dados simples e expandida.

B. Pré-Processamento do Texto

Como a base final ficou com apenas 49 ofícios classificados como éticos ou não, o número de casos não era suficiente para a obtenção de um modelo de aprendizado mínimo. Assim, foi aplicada uma sobreamostragem nas linhas dos documentos,

com o intuito de ampliar o conjunto de dados disponíveis para a análise.

No procedimento inicial, cada elemento dos pareceres médicos foram divididos em linhas, mantendo a rotulação original, em seguida as linhas foram combinadas para gerar um conjunto de 39.421 casos classificados, sendo 9.464 textos com parecer favorável e 29.957 contrários. Esse números de textos por classe ética demonstra o desequilíbrio dos pareceres sobre os ofícios obtidos do site.

Com o uso do código de ética, o número de casos com a combinação de linhas, salta para 64.621 elementos classificados. Gerando 34.664 pareceres favoráveis e 29.957 contrários. A influência do código de ética médica permitiu equalizar as classes e ajustar melhor a base de dados.

A seguir, os textos devem ser convertidos para números, pois os aprendizados de máquina são metodologias numéricas. Assim é realizada a extração das palavras do texto (tokenização) contando sua aparição no documento. Outra opção é utilizando a frequência dos termos de texto com inverso da frequência no documento (*Term frequency -Inverse Document Frequency* TF-IDF). O objetivo de calcular a idf ao invés simplesmente da frequência de ocorrência, é para diminuir o impacto de tokens que ocorram com elevada frequência no documento e são empiricamente menos informativos do que recursos que são obtidos apenas em uma pequena parcela do documento.

Ambos os procedimentos, tokenização e TF-IDF são avaliados pela decomposição das bases em duas componentes principais, utilizando o método *Single Value Decomposition* (SVD) e plotados em gráficos 2D.

Na Figura 3 não é possível visualizar uma região de separação entre as classes favorável (1) e contrário (0) dos pareceres analisados, com as duas classes estando sobrepostas no gráfico. Na Figura 4 as regiões favoráveis (1) e contrárias (0) são mais pronunciadas, indicando melhor os resultados para obtenções de padrões pelas técnicas de aprendizado de máquina. As bases de dados com apenas pareceres e a base com código de ética geraram gráficos muito similares para ambos os pré-processamentos.



Figura 3. Localização das classes via Tokenização.

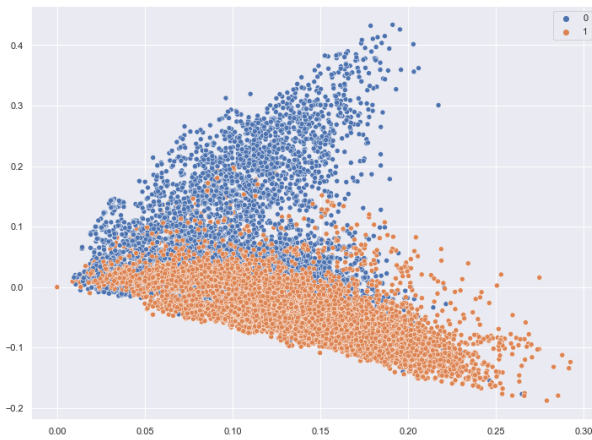


Figura 4. Localização das classes via TF-IDF.

C. Treinamento

O procedimento de treinamento foi realizado com um conjunto de metodologias de aprendizado de máquina, a fim de determinar por meio da análise de modelos matemáticos gerados, qual era a técnica com melhores métricas sobre a classificação da atividade médica como ética ou não.

Como são técnicas supervisionadas, um cuidado especial deve ser tomado durante o treinamento para evitar o sobre-treino (*Overfitting*), ou seja, evitar que o modelo seja tão ajustado ao conjunto de dados, que se torna incapaz de prever novos resultados. Para isso foi utilizado o método de treinamento por validação cruzada. A validação cruzada é um procedimento de re-amostragem usado para avaliar modelos de aprendizado de máquina em uma amostra de dados limitada. O procedimento tem um único parâmetro chamado k que se refere ao número de grupos em que uma determinada amostra de dados deve ser dividida. Como tal, o procedimento é frequentemente chamado de validação cruzada k -Fold. Quando um valor específico para k é escolhido, ele pode ser usado no lugar de k na referência ao modelo, como $k=5$ tornando-se validação cruzada de 5 vezes. Sendo aplicada no treinamento para estimar a habilidade de um modelo de aprendizado de máquina em dados não vistos. Ou seja, usar uma amostra limitada para estimar o desempenho do modelo em geral quando usado para fazer previsões sobre dados não usados durante o treinamento do modelo. É um método popular porque é simples de entender e porque geralmente resulta em uma estimativa menos tendenciosa ou menos otimista da habilidade do modelo do que outros métodos, como uma simples divisão de treinamento e teste.

As metodologias de aprendizado de máquina supervisionado classificador avaliadas são pertencentes ao pacote em Python *Scikit-Learn* (*SkLearn*) [13]. Foram utilizadas 16 metodologias funcionais com as bases de dados utilizadas nesta pesquisa, o classificador *Dummy* teve finalidade apenas de uma referência estatística às demais técnicas, pois cria um modelo sorteado de classificação que obtém o resultado baseado em uma seleção aleatória. A seguir são listadas os aprendizados classificadores utilizados:

- *DummyClassifier*
- *KNeighborsClassifier* [14];
- *SGDClassifier* [15];
- *PassiveAggressiveClassifier* [16];
- *NuSVC* [17];
- *LinearSVC* [18];
- *RidgeClassifier* [19];
- *ExtraTreeClassifier* [20];
- *AdaBoostClassifier* [21];
- *DecisionTreeClassifier* [20];
- *GradientBoostingClassifier* [22];
- *RandomForestClassifier* [23];
- *ExtraTreesClassifier* [23];
- *BaggingClassifier* [24];
- *SVC* [18];
- *MLPClassifier* [25];

Com a obtenção do modelo matemático inteligente pelo treinamento, são empregadas as métricas de avaliação dos resultados sobre os dados de teste obtidos da validação cruzada. Essas métricas foram obtidas do pacote *Scikit-Learn*.

Todos os passos descritos anteriormente, como a aquisição dos ofícios com pareceres éticos, conversão texto, validação cruzada, treinamento e avaliação são apresentados no fluxograma da Figura 5.

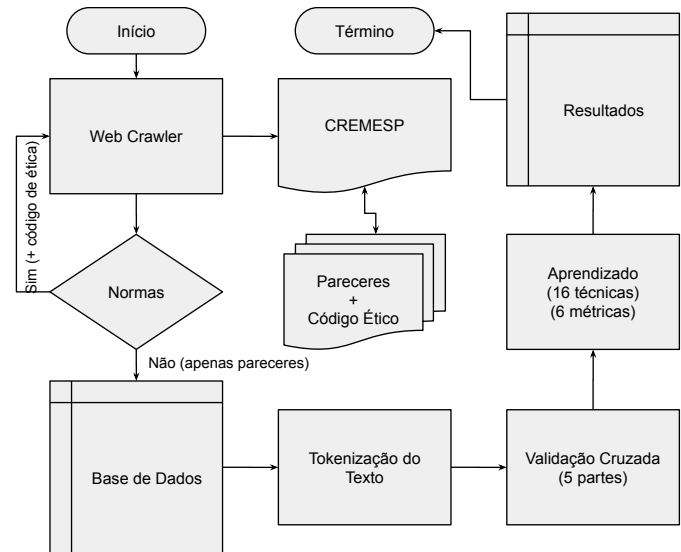


Figura 5. Etapa de treinamento com avaliação das técnicas.

D. Parametrização

Na implementação de cada método, foi utilizada uma técnica de busca aleatória (*RandomizedSearchCV*, em inglês) para a otimização dos parâmetros de ajuste utilizados a cada iteração. Os parâmetros do estimador utilizados para aplicar esses métodos são otimizados por uma busca com validação cruzada nas configurações dos parâmetros, o fluxograma do do algoritmo está demonstrado na Figura 6. O número de iterações definidas interfere diretamente na qualidade da solução obtida, um número maior de iterações gera um melhor

resultado, porém o custo computacional também é elevado. Foram definidas inicialmente 10 iterações para cada método.

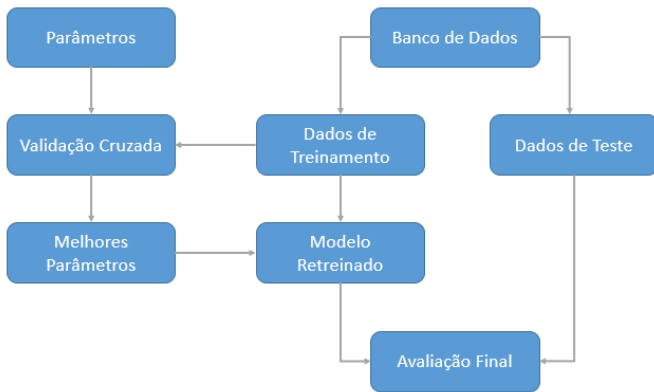


Figura 6. Fluxograma do algoritmo de otimização dos parâmetros.

E. Métricas Utilizadas

A análise dos desempenhos obtidos para cada método serão medidas por seis métricas tipicamente utilizadas em aplicações de aprendizado de máquina [26]. São elas:

- 1) *Tempo*: Se refere ao tempo gasto pelo computador para processar a aplicação em cada método;
- 2) *Acurácia*: Indica dentre todas as classificações, quantas o modelo classificou de forma correta;
- 3) *Acurácia Balanceada*: Também indica um valor em relação aos acertos do modelo, mas não é influenciado pelo desbalanceamento entre as classes de positivos e negativos;
- 4) *Precisão*: Mostra a assertividade do modelo em acertar a detecção de pareceres verdadeiros;
- 5) *Recall*: É a métrica que avalia quanto o modelo consegue reconhecer determinada classe, analisando os falsos negativos;
- 6) *F1*: Faz a média harmônica entre a precisão e o recall estabelecendo a performance geral do modelo.

IV. RESULTADOS

O processo de classificação dos pareceres éticos resultou em duas bases de dados diferentes para sua análise, a base simples com os pareceres e a expandida adicionando o código de ética médica. Desta forma, foram realizadas duas baterias de testes com as 16 metodologias de aprendizado de máquina supervisionado classificador e as 6 métricas de avaliação dos testes.

A. Base de Dados Simples

Empregando a base de dados apenas com os ofícios extraídos dos pareceres do CREMESP foi possível obter os resultados da Tabela I. Nela existe um destaque para o *KNeighborsClassifier* que obteve os melhores resultados para a métrica F (F1) e recuperação. Um segundo ponto de destaque fica para a acurácia balanceada, que é inferior da acurácia original devido ao desbalanceamento de classes (favorável e contra) dessa base de dados. A metodologia de referência do

DummyClassifier apresentou problemas em sua separação de classes, que pelos valores de precisão e recuperação, utilizou apenas uma única classe para todos os testes realizados, indicando-os como contrários ao código de ética médica. O classificador *NuSVC* com *Support Vector Machine (SVM)* não conseguiu gerar um modelo matemático com seu treinamento.

Portanto, com a base de dados de pareceres simples, foi possível obter um acerto de 58,5% em métrica F para o modelo obtido pelo *KNeighborsClassifier*.

B. Base de Dados Expandida

Na Tabela II é possível identificar uma melhoria geral das avaliações sobre as metodologias de aprendizado de máquina sobre a base de dados, quando incorporado o código de ética médica em seus ofícios. Diferente da outra base, o modelo com *DummyClassifier* conseguiu utilizar ambas as classes (favorável e contrário) para seus teste e nenhuma técnica pereceu sem gerar modelos classificadores.

Portanto, o melhor modelo matemático classificador para os testes foi obtido da metodologia *GradientBoostingClassifier* com métrica F de 77,54%, uma melhoria de 19,03% em relação às métricas com a base de dados simples. Ainda o *MLPClassifier* mesmo com métricas próximas (0,02% abaixo) ao *GradientBoostingClassifier* tem um tempo de treinamento 18 vezes maior para este volume de dados, sendo desconsiderada como segunda melhor opção.

C. Parametrizando Melhores Técnicas

Na parametrização automatizada pelo *RandomizedSearchCV* foram ajustadas as metodologias:

- *GradientBoostingClassifier*;
- *MLPClassifier*;
- *KNeighborsClassifier*.

Nesta análise, foi utilizada a base de dados expandida pelos seus melhores resultados.

Ao fim do processo de parametrização, nenhuma das metodologias apresentaram melhorias em suas métricas, principalmente em F1. Desta forma, para nesta base de dados específica, ajustar os parâmetros de treinamento base das técnicas não gerou resultados significativos para justificar essas alterações.

V. CONCLUSÃO

Conclui-se que é possível obter modelos matemáticos inteligentes para a classificação automatizada de atividades em saúde em relação ao código de ética médica. O procedimento torna-se simples e rápido, facilitando a fiscalização das atividades médicas, garantindo o bem estar dos pacientes e profissionais da saúde.

Como principais contribuições desta pesquisa, destaca-se a geração da base de dados expandida via o uso do próprio código de ética médica como fonte adicional de informação ao treinamento, e ao uso do TF-IDF como principal pré-processamento dos documentos em sua conversão numérica para as metodologias de aprendizado de máquina.

Técnica	Tempo(s)	Acurácia	Ac.Balanceada	Precisão	Recall	F1
KNeighborsClassifier	0,0208	0,7679	0,7157	0,5834	0,6126	0,5851
MLPClassifier	186,1	0,7983	0,7087	0,6202	0,5318	0,5626
PassiveAggressiveClassifier	0,0402	0,8023	0,6973	0,6275	0,4899	0,5382
RandomForestClassifier	11,87	0,7999	0,6855	0,6420	0,4593	0,5247
SGDClassifier	0,0314	0,8075	0,6901	0,6596	0,4581	0,5225
ExtraTreesClassifier	18,53	0,8044	0,6905	0,6337	0,4655	0,5204
LinearSVC	0,0848	0,8106	0,6886	0,6534	0,4476	0,5138
DecisionTreeClassifier	3,364	0,7523	0,6684	0,5342	0,5028	0,5128
RidgeClassifier	0,0742	0,8124	0,6846	0,6614	0,4320	0,5034
BaggingClassifier	18,42	0,7768	0,6651	0,5914	0,4442	0,4982
ExtraTreeClassifier	0,1909	0,7504	0,6622	0,5113	0,4880	0,4895
AdaBoostClassifier	3,048	0,7564	0,6463	0,5640	0,4286	0,4760
SVC	13,05	0,8237	0,6807	0,7168	0,3983	0,4742
GradientBoostingClassifier	13,05	0,8028	0,6362	0,6526	0,3068	0,4121
DummyClassifier	0,0152	0,7529	0,5000	0,0000	0,0000	0,0000
NuSVC	0,0438	-	-	-	-	-

Tabela I
RESULTADOS PARA A BASE DE DADOS SIMPLES.

Técnica	Tempo(s)	Acurácia	Ac.Balanceada	Precisão	Recall	F1
GradientBoostingClassifier	19,15	0,8018	0,8020	0,7917	0,7989	0,7754
MLPClassifier	355,4	0,7926	0,7915	0,7902	0,8074	0,7752
KNeighborsClassifier	0,0228	0,7537	0,7491	0,7397	0,8192	0,7715
RandomForestClassifier	15,75	0,7855	0,7851	0,7739	0,7925	0,7687
LinearSVC	0,1201	0,7895	0,7884	0,7845	0,8042	0,7673
ExtraTreesClassifier	22,46	0,7850	0,7843	0,7756	0,7948	0,7663
BaggingClassifier	27,82	0,7829	0,7838	0,7914	0,7696	0,7662
SGDClassifier	0,0980	0,7880	0,7870	0,7772	0,8016	0,7662
AdaBoostClassifier	5,281	0,7738	0,7722	0,7558	0,7965	0,7654
PassiveAggressiveClassifier	0,0577	0,7818	0,7806	0,7768	0,7988	0,7639
RidgeClassifier	0,0945	0,7815	0,7801	0,7724	0,8020	0,7625
DecisionTreeClassifier	3,544	0,7649	0,7648	0,7698	0,7664	0,7563
NuSVC	88,82	0,7883	0,7886	0,7758	0,7830	0,7535
SVC	32,23	0,7798	0,7793	0,7717	0,7865	0,7479
ExtraTreeClassifier	0,2629	0,7439	0,7432	0,7343	0,7527	0,7359
DummyClassifier	0,0194	0,5329	0,5000	0,5329	1,0000	0,6953

Tabela II
RESULTADOS PARA A BASE DE DADOS EXPANDIDA.

A obtenção de um resultado de 77,54% de métrica F (F1) para ambas as classes estudadas (favorável ou contrário) na ética médica ainda indica possíveis melhorias a serem realizadas na área, mas é um resultado robusto para emprego em fiscalizações básicas em atividades de saúde em qualquer tipo de unidade médica.

A. Trabalhos Futuros

Neste estudo ficou claro a necessidade de uma base de dados robusta e extensa para garantir melhores resultados nos aprendizados de máquina. Desta forma, a obtenção de novos documentos com pareceres éticos é essencial para a melhoria de seus resultados.

Além disso, empregar redes neurais com aprendizado de máquina profundo para texto pode ser interessante na geração de novas abordagens com os dados e garantir modelos matemáticos mais eficazes na classificação das atividades de saúde no caráter ético.

AGRADECIMENTOS

Deixamos nossos sinceros agradecimentos ao professor Carlos Henrique Valério de Moraes pela orientação e colaboração a este projeto desenvolvido, além de todos professores e amigos que nos acompanharam durante toda a graduação.

BIOGRAFIA



Carlos Eduardo Valareti Natural de Itajubá (MG), ingressou na Universidade Federal de Itajubá em 2017, no curso de Engenharia de Controle e Automação. Durante a graduação realizou intercâmbio na Technische Universität Berlin, Alemanha e na Universidad Carlos III de Madrid, Espanha. Atuou como Trainee na Voith Paper entre 2021 e 2022.



Mateus Cintra Costa Nascido em Piranguinho (MG), em 1998. Ingressou na Universidade Federal de Itajubá em 2016, no curso de Engenharia de Controle e Automação. Foi membro do projeto Equipe Uai!rrior de Robótica. Atuou como estagiário de engenharia na empresa Jonfra Automação Ltda no ano de 2022.

REFERÊNCIAS

- [1] A. R. Meira, M. M. Cunha, *et al.*, “O ensino da ética médica, em nível de graduação nas faculdades de medicina do Brasil,” *Revista brasileira de educação médica*, vol. 18, pp. 07–10, 2021.
- [2] I. Kononenko, “Machine learning for medical diagnosis: history, state of the art and perspective,” *Artificial Intelligence in medicine*, vol. 23, no. 1, pp. 89–109, 2001.
- [3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [4] F. Wang, L. P. Casalino, and D. Khullar, “Deep Learning in Medicine—Promise, Progress, and Challenges,” *JAMA Internal Medicine*, vol. 179, pp. 293–294, 03 2019.
- [5] X.-W. Chen and X. Lin, “Big data deep learning: challenges and perspectives,” *IEEE access*, vol. 2, pp. 514–525, 2014.
- [6] J. Hirschberg and C. D. Manning, “Advances in natural language processing,” *Science*, vol. 349, no. 6245, pp. 261–266, 2015.
- [7] J. Hirschberg and C. D. Manning, “Advances in natural language processing,” *Science*, vol. 349, no. 6245, pp. 261–266, 2015.
- [8] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [9] B. A. Goldstein, A. M. Navar, M. J. Pencina, and J. Ioannidis, “Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review,” *Journal of the American Medical Informatics Association*, vol. 24, no. 1, pp. 198–208, 2017.
- [10] G. Press, “Cleaning big data: Most time-consuming, least enjoyable data science task, survey says,” *Forbes*, March, vol. 23, p. 15, 2016.
- [11] S. Lohr, “For big-data scientists, ‘janitor work’ is key hurdle to insights,” *New York Times*, vol. 17, p. B4, 2014.
- [12] “Cremesp,” <http://www.cremesp.org.br>. Acessado: 2022-07-10.
- [13] O. Kramer, “Scikit-learn,” in *Machine learning for evolution strategies*, pp. 45–53, Springer, 2016.
- [14] J. L. Bentley, “Multidimensional binary search trees used for associative searching,” *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.
- [15] B. Zadrozny and C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 694–699, 2002.
- [16] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, “Online passive aggressive algorithms,” 2006.
- [17] C.-C. Chang and C.-J. Lin, “Libsvm: a library for support vector machines,” *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.
- [18] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” *The Journal of machine Learning research*, vol. 9, pp. 1871–1874, 2008.
- [19] R. M. Rifkin and R. A. Lippert, “Notes on regularized least squares,” 2007.
- [20] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, “Classification and regression trees. belmont, ca: Wadsworth,” *International Group*, vol. 432, no. 151-166, p. 9, 1984.
- [21] T. Hastie, S. Rosset, J. Zhu, and H. Zou, “Multi-class adaboost,” *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.
- [22] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, pp. 1189–1232, 2001.
- [23] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [24] L. Breiman, “Bagging predictors,” *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [25] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [26] Z. C. Lipton, C. Elkan, and B. Narayanaswamy, “Thresholding classifiers to maximize f1 score,” *arXiv preprint arXiv:1402.1892*, 2014.
- [27] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, “API design for machine learning software: experiences from the scikit-learn project,” in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122, 2013.
- [28] Y. Bengio, Y. LeCun, *et al.*, “Scaling learning algorithms towards ai,” *Large-scale kernel machines*, vol. 34, no. 5, pp. 1–41, 2007.
- [29] A. Lone and A. Khan, “Recent trends in programming languages a shift in learning and intelligence acquisition: Technologies and advancements,” 01 2018.
- [30] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, *et al.*, “Scalable and accurate deep learning with electronic health records,” *NPJ digital medicine*, vol. 1, no. 1, pp. 1–10, 2018.
- [31] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep learning for computer vision: A brief review,” *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [32] Y. Yang and X. Liu, “A re-examination of text categorization methods,” in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 42–49, 1999.
- [33] H. Zhang, “The optimality of naive bayes,” *Aa*, vol. 1, no. 2, p. 3, 2004.
- [34] M. Pal, “Random forest classifier for remote sensing classification,” *International journal of remote sensing*, vol. 26, no. 1, pp. 217–222, 2005.
- [35] A. Lins and T. B. Ludermitz, “Hybrid optimization algorithm for the definition of mlp neural network architectures and weights,” in *Fifth International Conference on Hybrid Intelligent Systems (HIS’05)*, pp. 6–pp, IEEE, 2005.
- [36] E. Eğrioglu, Ç. H. Aladağ, and S. Günay, “A new model selection strategy in artificial neural networks,” *Applied Mathematics and Computation*, vol. 195, no. 2, pp. 591–597, 2008.