Aplicação de Técnicas de Aprendizagem de Máquina para Diagnóstico de Depressão, Ansiedade e Estresse

Ana Laura Barros de Melo[1], André Luiz Firmino Alves[1]

Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB) – Picuí, PB

ana.barros@academico.ifpb.edu.br, andre.alves@ifpb.edu.br

Resumo. *O estilo de vida moderno das pessoas está causando diferentes tipos de transtornos psicológicos, onde há resistência na busca de tratamento e as pessoas integram em uma árdua jornada de sofrimento com os sintomas de depressão, estresse e ansiedade. Nesse contexto, considerando que a aprendizagem de máquina (AM) tem sido amplamente eficaz no auxílio de diagnóstico médico, o presente trabalho tem como objetivo testar a eficácia dos modelos de aprendizagem de máquina utilizando a mineração de dados, como ferramentas auxiliares no pré-diagnóstico dessas doenças, fazendo uso da linguagem Python. Foi feita a classificação e previsão da ocorrência dos três problemas psicológicos utilizando quatro algoritmos, medindo cinco diferentes níveis de gravidade entre as doenças, unindo a tecnologia e a psicologia.*

Abstract. *The modern lifestyle of people is causing different types of psychological disorders, where there is resistance in seeking treatment and people are part of an arduous journey of suffering with the symptoms of depression, stress and anxiety. In this context, considering that machine learning (ML) has been widely effective in aiding medical diagnosis, the present work aims to test the effectiveness of machine learning models using data mining as auxiliary tools in pre-diagnosis. of these diseases, making use of the Python language. The classification and prediction of the occurrence of the three psychological problems was made using four algorithms, measuring five different levels of severity between the diseases, combining technology and psychology.*

1. Introdução

Um retrato do mundo contemporâneo nas questões de saúde é o contraste entre o tratamento que a sociedade disponibiliza quando presencia uma situação de dor física e uma situação de dor mental, emblemático cenário de como os transtornos mentais ainda são tratados com ignorância, desprezo e negligência, tanto por aqueles que são portadores, quanto pelos que não vivenciam tal agonia. O que reúne a depressão, ansiedade e estresse num mesmo grupo é o fato dos pacientes portadores de tais patologias estarem em duas situações-limites: a de nunca procurarem tratamento, resistindo a ceder por ser algo "psicológico", não dando assim a atenção à doença que ela precisa e a de quem busca o tratamento "rápido demais". Assim, na medida em que o quadro se agrava, diminuem os recursos psíquicos para que a pessoa procure tratamento [Dunker 2021].

Geralmente, os psicólogos utilizam critérios "simples" para mapeamento desses transtornos: quando o paciente sente dificuldade de ir no trabalho ou na escola, atitude crônica de "desistir antes de tentar", alterações relativas ao sono, libido ou alimentação ou

dificuldade para iniciar o dia ou um ciclo de atividades [Oliveto 2018]. Sendo assim, o questionário DASS-42, utilizado no presente trabalho, é composto por questões baseadas nessas métricas e divididas para cada transtorno, onde o usuário será orientado a definir a aplicabilidade daquele sintoma à sua vida. Esses questionários avaliativos em conjunto com a classificação garante um certo conforto ao usuário, por estarem habituados a uma interface não-humana, definindo um pré-diagnóstico de sua saúde mental. Por outro lado, o uso da tecnologia na área da saúde mental não tem como objetivo substituir os profissionais, e sim complementar o tratamento, com o propósito de atingir os pacientes em pontos que a abordagem tradicional não tem surtido efeito [Oliveto 2018].

Ambas tecnologias estão se tornando cada vez mais populares em muitas áreas do conhecimento como mecanismos auxiliares para a resolução de vários problemas. Atualmente, a MD (mineração de dados) e a AM têm aplicações em uma grande variedade de campos do conhecimento e, em particular, elas se tornaram ferramentas poderosas nos campos da medicina, saúde e biologia. Como exemplo desse uso, em Sapkal *et al.* [Sapkal; Mehta; Nimgaonkar; Devasthale; Phansalkar 2021], foi utilizada a aplicação da técnica de aprendizado de máquina no apoio de diagnósticos de transtornos mentais, usando um sistema para reconhecer os transtornos com mineração de dados, feito uma coleta de dados utilizando questionários sobre os sintomas.

O presente trabalho tem por objetivo testar a eficácia de algumas técnicas de AM e MD como ferramentas auxiliares para o diagnóstico de depressão, ansiedade e estresse, por meio da classificação das doenças por gravidade do diagnóstico. Esse objetivo pode auxiliar profissionais da saúde com um pré-diagnóstico, já que, por meio das perguntas feitas, alguns sintomas são possíveis portas de entrada para um diagnóstico feito com terapia.

2. Materiais e Metodologia

Esta seção objetiva descrever a metodologia utilizada no trabalho, conforme Figura 1:

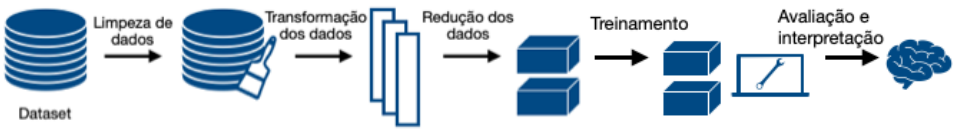

Figura 1. Fluxo de processamento.

Os dados do conjunto, coletados por [Peleg 2021], foram baseados no *DASS-42* (*Depression, Anxiety and Stress Scale*) [Lovibond 1995] e no *TIPI* (*Ten Iten Personality Inventory*). O *DASS 42* contém 42 perguntas, que foram utilizadas para avaliar os níveis de gravidade no diagnóstico referente a sintomas de depressão, onde cada sintoma possuía 14 perguntas correspondentes. O usuário tinha a opção de responder de 4 formas possíveis, desde "Não se aplica a mim de forma alguma", até "Aplica-se muito a mim, na maioria das vezes". O *TIPI* correspondia a aspectos de personalidade do usuário.

A pontuação de Depressão, Ansiedade e Estresse foram calculados de forma prática, somando os valores associados às respostas de cada questão e de cada classe. A partir desses

valores, cada pessoa do conjunto de dados foi avaliada de acordo com a gravidade do diagnóstico, ou seja, Normal, Médio, Moderado, Severo e Extremamente Severo.

O pré-processamento foi aplicado para melhorar a qualidade dos dados, removendo linhas com respostas erradas, colunas do *check-list* que não eram úteis ou qualquer outro fator que pudesse comprometer a eficiência das etapas da mineração. A primeira etapa diz respeito à Limpeza de dados, onde foi feita a eliminação de respostas que não atendiam o que as perguntas do *DASS-42* pediam ou qualquer outra informação incompatível e dados faltantes ou duplicados. Além disso, colunas desnecessárias para a mineração foram eliminadas.

Após a limpeza de dados, na segunda etapa, foi feita a seleção de atributos, onde novos atributos foram gerados a partir de um outro conjunto de atributos fornecidos para ajudar no processo de classificação, além da representação reduzida do *dataset* por meio da eliminação de atributos irrelevantes ou redundantes: variáveis altamente correlacionadas que não agregam informação para a construção do modelo. Também foi utilizado o método de discretização de atributos, que consiste em transformar valores contínuos de atributos em valores simbólicos (discretos).

Na terceira etapa, o processo de redução dos dados proferidos foi de extrema importância para que os dados se adequassem para análise. Foi utilizado as técnicas de redução de dimensionalidade, numerosidade e *PCA* (*Principal Components Analysis*).

Na quarta etapa, foi necessário separar o conjunto de dados em dois conjuntos, um de treinamento e outro de teste, obtendo, respectivamente, 25.267 e 13.606 registros.

Após separar os dados de treino e de teste, na quinta etapa, quatro modelos de classificação foram gerados: *K-nearest Neighbours*, *Random Forest*, *Support Vector Machines* e *Naive Bayes*. Após a criação e treinamento dos modelos, os dados de teste foram aplicados à eles, obtendo assim os resultados das métricas de avaliação.

4. Resultados e Discussões

Para cada modelo foi gerada a matriz de confusão de cada transtorno, e as métricas para validação foram extraídas para comparação, conforme a Tabela 1.

Tabela 1. Métricas de avaliação

Classificador	Doença	Acurácia	Precisão	Recall	Medida F1
K-nearest neighbour	Ansiedade	80.1%	77.4%	80.1%	77.6%
	Depressão	86.8%	87%	86.8%	85.8%
	Estresse	84.3%	83.9%	84.3%	83.7%
Random Forest	Ansiedade	85.3%	84.7%	85.3%	83.1%
	Depressão	93%	93%	93%	92.8%
	Estresse	88.3%	88.3%	88.3%	87.8%
SVM	Ansiedade	**98.7%**	**98.7%**	**98.7%**	**98.7%**
	Depressão	**99%**	**99%**	**99%**	**99%**
	Estresse	**98.5%**	**98.5%**	**98.5%**	**98.5%**
Naive Bayes	Ansiedade	81%	83.8%	81%	81.8%
	Depressão	87%	89.2%	87%	87.5%
	Estresse	85.3%	87%	85.3%	85.8%

Observando a Tabela 1, é possível perceber que as melhores avaliações para as três escalas foram obtidas pelo algoritmo SVM.

5. Conclusão

O presente trabalho se propôs a construir modelos de predição capazes de, por meio da aprendizagem de máquina e mineração de dados, identificar pessoas com índices de transtornos psicológicos através de análise de questionários de avaliação baseados em métricas semelhantes às que profissionais da saúde mental utilizam. A pesquisa serve como um apoio e incentivo no processo de identificação dos transtornos, sem pretenção de substituir o diagnóstico humano, discutindo a importância dessas tecnologias tão necessárias numa sociedade onde a tecnologia se mostra tão corriqueira.

Técnicas de MD e AM foram utilizadas para comparar resultados de classificadores e garantir uma boa performance dos modelos, resultando em um sistema de classificação eficaz. O *Support Vector Machine* obteve os melhores resultados, com uma **taxa de acurácia acima de 98% para os três sintomas**, como demonstra as métricas de avaliação do desempenho dos classificadores.

Pela eficiência apresentada no trabalho, outras pessoas podem continuar respondendo ao questionário para obter uma previsão do seu estado de saúde mental, aperfeiçoando ainda mais os modelos de AM com a inserção de novos dados. Por fim, a contribuição do presente trabalho para a área de previsão de casos psicológicos irá auxiliar outras pesquisas que possuem objetivos similares na área de ciência de dados em conjunto com psicologia. Como trabalho futuro, poderia-se analisar a aplicação de técnicas de *deep learning* para comparar os resultados com técnicas tradicionais de aprendizagem de máquina.

Referências Bibliográficas

Peleg, Yam. (2021). "Predicting Depression, Anxiety and Stress". Kaggle.

Lovibond, S.H. and Lovibond, P.F. (1995). Manual for the Depression Anxiety Stress Scales (2nd. Ed.). Sydney: Psychology Foundation.

Dun, C. (2021). Uma biografia da depressão. São Paulo: Planeta.

Oliveto, Paloma. (2018). Aplicativos de celular são as novas armas no combate à depressão. Correio Braziliense.

Sapkal, D.; Mehta, C.; Nimgaonkar, M.; Devasthale, R.; Phansalkar, S. (2021). Prediction of mental disorder using artificial neural network and psychometric analysis. In: Data Management, Analytics and Innovation. [S.l.]: Springer. p. 369–377.