

IA para reconhecimento de estado de risco psicológico

Gustavo Zanfelize Dib
Orientadora: Debora Medeiros

July 2019

1 Introdução

Desde o início do Século XX, transtornos de depressão e ansiedade tem se tornado mais comuns e recorrentes indicando uma área de estudo de extrema importância. Segundo a Organização Mundial de Saúde (ORGANIZATION et al., 2017), no Brasil, 5.8% da população sofre de depressão e 9.3% de transtorno de ansiedade. Vale-se lembrar que estes dados referem-se apenas a casos diagnosticados, por se tratar de patologias psicossociais, muitos casos não tem registro e/ou diagnóstico.

Com o avanço da área de ciências de dados, *deep learning*, *data mining*, *machine learning* e Inteligência Artificial se tornaram assuntos recorrentes na computação. Este paradigma oferece recursos para realizar estudos mais aprofundados sobre dados gerais, além de fornecer previsões baseadas em uma análise prévia. Em uma correlação entre ambos os cenários descritos acima, este projeto propõe utilizar de técnicas computacionais de previsão no auxílio de indivíduos em mal-estar psicológico.

Utilizando de técnicas descritas acima, é possível encontrar correlações mais sofisticadas entre algumas características e atributos do indivíduo. Este projeto visa obter, através de dados de perfil de um grupo de indivíduos, um estudo sobre transtornos depressivos e de ansiedade em caráter preditivo, quer possa auxiliar a tomada de uma ação proativa ao invés de reativa à possíveis indicadores.

2 Referencial teórico

Para a construção de modelos computacionais capazes de detectar padrões e realizar previsões comportamentais complexas, é necessário um processo de treinamento onde este modelo irá extrair informações sobre os diferentes atributos.

No livro *Learning from Data* (YASER S. ABU-MOSTAFA MALIK MAGDON-ISMAIL, 2012), este panorama se dispõe em uma função alvo desconhecida, uma amostra

de treino, um conjunto de hipóteses e uma hipótese final descoberta pelo algoritmo de aprendizado. Existem diferentes metodologias para um algoritmo aprender novas informações e correlações baseado em uma amostra. Essas técnicas podem ser agrupadas em três grandes tipos. Supervisionado, não-supervisionado e semi-supervisionado.

2.1 Aprendizado supervisionado

Para alguns problemas, a base de dados de treino, aonde a hipótese é ajustada, já possui a informação sobre a resposta desejada para cada instância de dado, ou seja, para todos os registros de minha amostra de aprendizado, eu já possuo o conhecimento sobre a resposta esperada.

Neste tipo de aprendizado, os problemas se dispõem em classificação, no qual o alvo é atribuir o indivíduo à uma classe pré-definida, e regressão, caso a função alvo seja contínua.

Dentre os métodos de treino supervisionado, existem as *Support Vector Machines* (SVM). As SVM utilizam-se da Teoria do Aprendizado Estocástico (TAE) para encontrar um classificador f dentre os F classificadores possíveis para que encontre o ponto ideal de separação, minimizando o risco esperado de *overfitt* e aprimore a acurácia. Este risco é calculado baseando-se na dimensão Vapnik-Chervonenkis (VC) (VAPNIK, 2013), na complexidade do classificador e tamanho da amostra de aprendizado. Para classificações não-lineares, ou seja, em casos onde um hiperplano não pode separar as diferentes classes apresentadas utilizam-se funções *Kernels* para a transformação dos dados. Estes modelos são ditos como robustos para se trabalhar com dados de grande dimensão, onde outros modelos sofrem dificuldades. (LORENA; CARVALHO, 2007) No conceito de algoritmos de regressão, uma técnica utilizada é a da Regressão por Mínimos Quadrados Parciais (PLS do inglês *partial least squares*). Este método busca encontrar uma função linear $f : X \rightarrow y$ através da computação de fatores de pontuação para os atributos baseado em combinações lineares. Uma vantagem desta metodologia é sua flexibilidade dentre os regressores lineares, podendo ser aplicada em cenários mais amplos e diversificados.

3 Trabalhos relacionados

A ideia de predição de estados mentais de risco é algo já retratado em alguns estudos e pesquisas através de diversas abordagens diferentes. Não é conceitualmente definido quais são todos os sinais de uma patologia psicológica, por isso, nas diferentes tentativas de análise, possuem-se diversas fontes de dados que podem ajudar à estudar a questão. Dentre essas possíveis fontes de dados, temos como exemplo a análise de ressonâncias magnéticas de pacientes, vídeos para análise comportamental através de fatores audiovisuais ou dados demográficos e estáticos sobre os indivíduos em análise.

No trabalho *Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition*. (KOUT-

SOULERIS, 2009), foi realizada uma pesquisa através do uso de um classificador supervisionado SVM para analisar ressonâncias magnéticas de pacientes selecionados para o reconhecimento de pessoas com risco em seu estado mental. A proposta da pesquisa era diferenciar pacientes entre os diferentes tipos de estado de risco aos não-diagnosticados. Os modelos gerados demonstraram boa acurácia (superior a 80%) em sua análise para os cenários dispostos. O que indica um possível diagnóstico prévio para os transtornos. Um revés desta tratativa é o difícil acesso à ressonâncias magnéticas em larga escala, tornando a reprodutibilidade baixa para cenários amplos.

Em um trabalho publicado em parcerias das universidades *Brunel University* e *Beihang University* (MENG, 2013), vídeos de pessoas interagindo com uma simulação foram utilizados como entradas para um modelo de regressão. Os vídeos foram processados em busca de correlações entre si, através da técnica *Partial least squares* (PLS), utilizada para encontrar a covariância entre matrizes, nesse caso, os dados extraídos do vídeo após um tratamento. A pesquisa avaliou o desempenho de diferentes tratativas para os vídeos comparando duas métricas de avaliação, a média do erro absoluto e a média do erro quadrático dos modelos. Por se tratar de uma regressão, a proposta é se analisar uma escala para a depressão do indivíduo, diferente da proposta deste projeto, porém o estudo realizado aponta uma correlação forte entre a fala e expressões faciais com a saúde mental.

Na análise feita por Mortier (MORTIER, 2017), a pesquisa WHO World Mental Health International College Student (WMH-ICS) foi utilizada como triagem para estabelecer um espectro de mal-estar mental. Dentre os dados analisados, foram obtidas variáveis demográficas dos estudantes, indicadores de pensamentos suicidas, patologias psicológicas dos pais e experiências traumáticas na adolescência e infância, risco para transtorno mental em 12 meses e eventos de alto nível de stress nos últimos 12 meses. Todos esses dados foram baseados em estudos prévios de centros de psicologia.

Para a modelagem, o estudo utilizou o *multivariate prediction model*, um modelo de predição baseado em séries temporais para analisar os eventos tratados. Com isso, a análise baseada no *Nagelkerke pseudo- R^2* , uma medida de erro quadrático, resultou em valores relativamente baixos, enquanto a AUC foi alta para ambos os cenários, ideação e plano suicida. O artigo conclui que a predição foi moderada, porém alguns dados como desistência poderiam ser adicionados, além de que alguns dados foram auto-diagnosticados, o que pode prejudicar a qualidade da base.

4 Proposta

Este trabalho tem como objetivo investigar a aplicação de técnicas de aprendizado semi-supervisionado para predição de ocorrência de problemas emocionais em estudantes do ensino superior.

Serão consideradas abordagens de classificação e de agrupamento capazes de incorporar tanto dados rotulados e não-rotulados. Como a tarefa final con-

siste em classificação (predição), no caso de abordagens de agrupamento, os modelos ajustados serão adaptados para realizarem a tarefa de classificação. Mais especificamente, os grupos de estudantes encontrados pelas abordagens de agrupamento serão submetidos a um pós-processamento para serem atribuídos a uma das classes de estudantes. A predição de estudantes novos será dada pelo grupo ao qual eles são mais aderentes de acordo com suas características.

5 Materiais e métodos

Este trabalho avalia o desempenho de algoritmos de aprendizado de máquina implementados em *Python* para a predição de dois alvos dentro da base de dados.

5.1 Base de Dados

Para o trabalho desenvolvido foi utilizada para estudo a base de dados de acompanhamento dos discentes da Universidade Federal do ABC (UFABC), contendo dados sociais (renda familiar, tipo de moradia, etc.) e dados estudantis (número de reprovações, coeficiente de rendimento, etc.)(UFABC, 2018).

A base de dados crua é constituída pela informação de 9929 discentes distribuídas em 323 colunas sem nenhuma identificação individual. Para o aprendizado dos modelos computacionais, foram necessárias as transformações dos dados descritas abaixo.

5.1.1 Tratamento dos dados

As informações presentes foram divididas em três tipos:

- Atributos nominais: Valores que apenas podem ser comparados em relação de igualdade, ou seja, não possuem relação de superioridade ou inferioridade, como "Tipo de moradia", "Estado Civil", etc. Essas variáveis foram transformadas através da técnica *One-Hot Encoding*, onde para cada possível valor, é criado uma coluna (binária) própria.
- Atributos Ordinais: Valores que podem ser ordenados porém não pertencem ao domínio contínuo, por exemplo, respostas como "Nunca / Esporadicamente / Diariamente", ou casos similares. Para essas variáveis foram atribuídos números inteiros representativos da ordem de grandeza da classe. Para degraus de valores como "1 a 3 vezes durante a semana", os números empregados são os limites superiores do degrau, sendo o valor para degraus superiormente ilimitados a média da diferença dos demais somada ao limite inferior do maior.
- Atributos contínuos: Valores pertencentes ao domínio contínuo, ou seja, respostas dentro do conjunto dos números inteiros ou números reais, como "Ano de ingresso" ou "Tempo de viagem até a faculdade". Esses valores

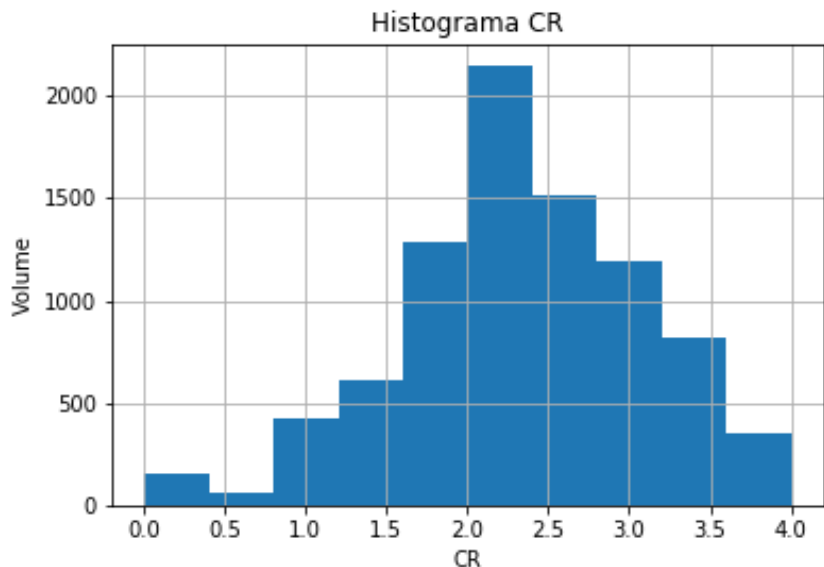
não foram transformados por já serem naturalmente interpretáveis pelos modelos implementados.

- Coeficiente de Rendimento: Média ponderada das notas recebidas em todas as disciplinas cursadas pelo discente.
- Trancamento de Quadrimestre por justificativa psicológica: Atributo binário sinalizando se foi efetuado o trancamento total de um período letivo devido a problemas psicológicos.

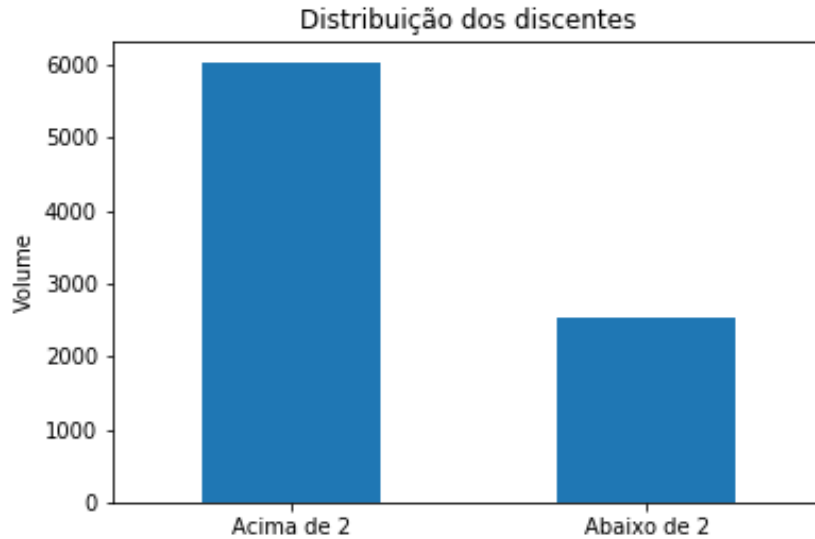
5.2 Coeficiente de Rendimento (CR)

A base de dados utilizada possui um total de 1360 indivíduos com coeficiente não identificados, estes registros foram excluídos da base para evitar erros de compreensão estatística. Dos 8569 restantes foram analisadas as seguintes informações básicas: distribuição das amostras, média e desvio padrão.

- Média: 2.3
- Desvio Padrão: 0.76



Para a implementação de um classificador, os coeficientes foram discretizados em nas classes acima de 2 e abaixo de 2. Essa divisão foi feita baseada no critério de graduação da Universidade, que estabelece que para um discente concluir o curso, deve-se obter um coeficiente acima de 2. Com essa transformação, a nova distribuição se torna:



Como pode-se ver no gráfico acima, existe um desbalanceio entre as classes, para que isso não influencie o resultado do modelo, foi realizado o *UpSampling* de dados da classe minoritária, sendo assim, foram adicionados novos indivíduos para a classe "Abaixo de 2" através do algoritmo *SMOTE* (CHAWLA, 2002), que aleatoriamente modifica os indivíduos da classe minoritária e calcula os K vizinhos mais próximos (*K-Nearest Neighbours*) para adicionar na base.

Com os dados de entrada tratados, foi utilizado o algoritmo *Random Forest* (HO, 1995). O algoritmo se baseia na criação de n árvores de decisão medindo-se a entropia do sub-grupo de cada nó (para o nó raiz é avaliado todo o conjunto de dados) e maximizando o ganho de informação através de *bootstrap aggregating*, ou seja, utiliza-se amostras aleatórias extraídas com reposição do conjunto original de dados, assim gerando uma floresta. Para a predição é considerada a decisão majoritária da floresta. Este algoritmo foi selecionado devido os seguintes pontos:

- Variáveis categóricas: Algoritmos baseados em árvores tiram grande proveito de atributos divididos em classes, diferente de discriminantes e afins;
- *Bagging*: Como a base final de entrada possui 480 colunas, a técnica de seleção aleatória adiciona maior robustez ao modelo contra possíveis *features* e indivíduos prejudiciais ao modelo.
- Interpretabilidade: Como forte vantagem, algoritmos baseados em árvores ou florestas podem ser bem interpretados através da análise de seus nós.

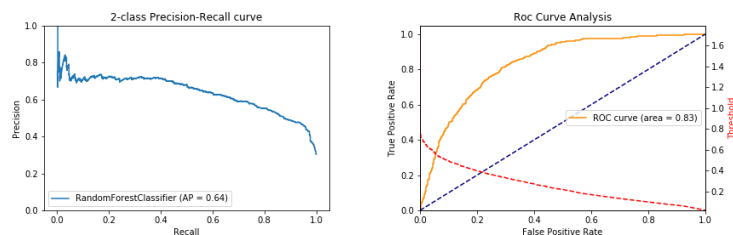
5.3 Avaliação dos resultados

5.3.1 Coeficiente de Rendimento

Para treino e avaliação dos modelos os dados foram divididos em treino e validação na proporção 7:3 de forma aleatória. Vale-se destacar que o *Over-Sampling* citado anteriormente só foi realizado na porção de treino do modelo para não impactar a análise de qualidade do algoritmo para uma base real.

Inicialmente foi avaliada a acurácia geral do modelo que convergiu para 77.36%, porém devido a definição do problema, apenas o índice de predições corretas não retorna o real desempenho desejado. Como o intuito desejado é avaliar e alertar sobre possíveis riscos de mal-estar dos alunos, deseja-se predições que maximizem a revocação (Predições positivas dividido pelo total de verdadeiros positivos) para que possa-se gerar confiança de que a maioria de alunos em situação inadequada seja identificado.

Junto a revocação, foi utilizado outros dois indicadores, a precisão, que é a taxa de predições corretas positivas sobre a taxa de predições corretas somada a taxa de predições incorretas positivas, e a área sob a curva ROC (HANLEY; MCNEIL, 1982) que é calculada a partir das taxas de falsos positivos sobre a taxa de verdadeiros positivos. Os resultados são exibidos nos gráficos abaixo:



Os resultados para ambas as métricas foram positivos para a base de validação e como dito anteriormente, para priorizar o *Recall* ou revocação, foi implementado um *threshold* ao modelo para que os resultados sejam:

- Recall: 80.00%
- Precision: 55.14%
- Acurácia: 74.24%
- Área sob a curva ROC: 75.95%

References

CHAWLA, N. V. et al. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, v. 16, p. 321–357, 2002.

- HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, v. 143, n. 1, p. 29–36, 1982.
- HO, T. K. Random decision forests. In: IEEE. *Proceedings of 3rd international conference on document analysis and recognition*. [S.l.], 1995. v. 1, p. 278–282.
- KOUTSOULERIS, N. et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Archives of general psychiatry*, v. 66, p. 700–12, 08 2009.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007.
- MENG, H. et al. Depression recognition based on dynamic facial and vocal expression features using partial least square regression. *AVEC 2013 - Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, p. 21–30, 10 2013.
- MORTIER, P. et al. First onset of suicidal thoughts and behaviours in college. *Journal of affective disorders*, Elsevier, v. 207, p. 291–299, 2017.
- ORGANIZATION, W. H. et al. *Depression and other common mental disorders: global health estimates*. [S.l.], 2017.
- UFABC. *Perfil do Discente de Graduação*. 2018. Dados obtidos da Universidade Federal do ABC disponível em: <http://propladi.ufabc.edu.br/informacoes/perfil>.
- VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer science & business media, 2013.
- YASER S. ABU-MOSTAFA MALIK MAGDON-ISMAIL, H.-T. L. *Learning From Data. A short course*. AML-Book, 2012. ISBN 1600490069, 978-1600490064. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=71ADDE9248AC4054B474A956D826033B>>.