

COMPORTAMENTO DO USUÁRIO EM REDE SOCIAL: MINERAÇÃO DE DADOS COM A CLASSIFICAÇÃO DA SEVERIDADE DEPRESSIVA DOS USUÁRIOS COM A UTILIZAÇÃO DA ESCALA DE DEPRESSÃO DE BECK BDI-II PARA A PREDIÇÃO DE TRAÇOS DEPRESSIVOS

Maricy CAREGNATO
cmaricy@ime.usp.br

RESUMO

As redes sociais tornaram-se fontes de pesquisas científicas dado ao grande volume de dados registrados pelos usuários em seus perfis sociais, possibilitando a mineração desses dados na produção de novos conhecimentos que podem contribuir na identificação de traços de comportamento depressivo. Objetivo - Apresentar um Modelo de Regressão Logística Multinomial a partir do comportamento longitudinal do usuário na rede social, para a predição de probabilidades de traços depressivos. Métodos - De natureza aplicada, com uma abordagem quantitativa, exploratória e descritiva quanto aos objetivos, bibliográfica, de levantamento e experimental quanto aos procedimentos. Dados de postagens, curtidas e sintomas depressivos de 692 usuários da rede social Facebook, brasileiros e maiores de 18 anos foram coletados via aplicativo específico. Resultados Os dados relativos aos sintomas depressivos, obtidos via Inventário de Depressão de Beck, e os dados da rede social formaram a base experimental para às análises do modelo de regressão logística multinomial, demonstrado a viabilidade na execução de um modelo capaz de prever as probabilidade de um usuário apresentar sintomas depressivos considerando os traços depressivos e seu comportamento na rede social caracterizado por postagens e curtidas. Conclui-se que as predições dos níveis de traços depressivos em rede social representam a capacidade de as ciências atuarem de maneira interdisciplinar a contribuir para a saúde pública em prol do bem-estar social.

Palavras-chave: Depressão; Rede Social; BDI-II; Mineração de Dados; Regressão Logística.

RESUMEN

Las redes sociales se convirtieron en fuentes de investigación científica, dado el gran volumen de datos registrados por los usuarios en sus perfiles sociales, posibilitando la minería de esos datos en la producción de nuevos conocimientos que pueden contribuir en la identificación de rasgos de comportamiento depresivo. Objetivo - Presentar un Modelo de Regresión Logística Multinomial a partir del comportamiento longitudinal del usuario en la red social, para la predicción de probabilidades de rasgos depresivos. Métodos - De naturaleza aplicada, con un abordaje cuantitativo, exploratorio y descriptivo en cuanto a los objetivos, bibliográfica, de levantamiento y experimental en cuanto a los procedimientos. Los datos de posturas, curtidos y síntomas depresivos de 692 usuarios de la red social Facebook, brasileños y mayores de 18 años fueron recolectados vía aplicación específica. Resultados Los datos relativos a los síntomas depresivos obtenidos a través del Inventario de Depresión de Beck y los datos de la red social formaron la base experimental para los análisis del modelo de regresión logística multinomial, demostrada la viabilidad en la ejecución de un modelo capaz de predecir las probabilidades de un usuario presenta síntomas depresivos considerando los rasgos depresivos y su comportamiento en la red social caracterizado por posturas y curtidas. Se concluye que las predicciones de los niveles de trazos depresivos en red social representan la capacidad de las ciencias para actuar de manera interdisciplinaria a contribuir a la salud pública en favor del bienestar social.

Palabras clave: Depresión; Red social; BDI-II; Minería de datos; Regresión Logística.

1 INTRODUÇÃO

De acordo com a Organização Mundial da Saúde OMS (2017) os dados mundiais destacam que a depressão é responsável por 7,5% da incapacidade humana e a principal causa de mortes por suicídio, com cerca de 800 mil casos por ano. Com a estimativa para 2020 de se tornar a segunda causa de mortes no mundo, ficando somente atrás das doenças cardíacas.

Segundo Baldwin e Birtwistle (2002), a combinação de fatores genéticos, ambientais e psicológicos podem ser a causa do desenvolvi-

mento da depressão. Quando a depressão não é diagnosticada e, por conclusão tratada, ela tende a ser crônica e se associa a um maior grau de incapacidade em realizar atividades diárias, conforme aponta a OMS (2017).

A depressão pode ser entendida como tendo dois componentes: o componente afetivo (humor) e o componente físico (ou somático) por exemplo, perda de apetite. A BDI-II reflete esta abordagem e pode ser dividida em duas subescalas. O propósito destas subescalas é determinar a causa primária da depressão, segundo Beck et al. (1996).

A subescala afetiva contém 8 itens: pessimismo, perdas passadas, sentimento de culpa, sentimentos de punição, auto-desprezo, auto-crítica, pensamentos ou desejos suicidas e pensamentos de desvalor. A subescala somática consiste de outros 13 itens: tristeza, alterações no apetite, perda de prazer, choro, agitação, perda de interesse, cansaço ou fadiga, indecisão, perda de energia, alterações nos padrões de sono, irritabilidade, dificuldades de concentração e diminuição da libido. As duas subescalas são moderadamente correlacionadas a 0,57, o que sugere que os aspectos físicos e psicológicos da depressão são relacionados ao invés de completamente distintos Storch et al. (2004).

Embora Conforme Beck et al. (1996), o desenvolvimento da BDI tenha sido um importante acontecimento em psiquiatria e psicologia, bem como, representado uma mudança na visão da depressão, pois transitou de uma abordagem freudiana e psicodinâmica para uma abordagem guiada pelos pensamentos ou cognições. Assim como outros inventários ou escalas de auto-relato a BDI pode possuir um certo viés, nos quais os resultados podem ser facilmente exacerbados ou reduzidos pelo usuário que as responde.

Da mesma maneira que os demais questionários, a forma na qual o instrumento é administrado pode causar um efeito no resultado final. Se um usuário completar o questionário da escala na presença de outras pessoas, em um ambiente clínico, por exemplo, as expectativas sociais podem criar resultados divergentes quando comparados à aplicação através de envio.

Em usuários com manifestações físicas associadas, o peso do BDI em traços físicos como fadiga, por exemplo, pode artificialmente incrementar os resultados devido a sintomas físicos de doenças ao invés de depressivos. Em uma tentativa de resolver este problema, Beck e colegas desenvolveram o Beck Depression Inventory for Primary Care (BDI-PC), uma curta escala de medida que consiste em sete itens da BDI-II considerados independentes da função física que são: tristeza,

pessimismo, fracasso passado, auto-estima, autocrítica, pensamentos ou desejos suicidas e perda de interesse. Diferentemente da BDI, a BDI-PC produz apenas o resultado binário de “sem depressão” ou “com depressão” para usuários que pontuam acima de 4. Embora planejado como um instrumento de investigação e não como ferramenta diagnóstica o BDI-PC pode ser ocasionalmente utilizado para obter um rápido diagnóstico, de acordo com Gorenstein et al. (2011).

Portanto, este trabalho tem como objetivo geral: preparar os dados coletados a partir da mineração de dados com a classificação da severidade depressiva dos usuários da rede social com a utilização da Escala de Depressão de Beck BDI-II para a predição de traços depressivos.

2 DIAGNÓSTICO DA DEPRESSÃO A PARTIR DE ESCALAS DE CLASSIFICAÇÃO: INVENTÁRIO DE DEPRESSÃO DE BECK (BDI)

O diagnóstico da depressão tem como referência a descrição das experiências por parte da pessoa, por meio de entrevista, e a posterior avaliação do estado mental, segundo descreve Baldwin e Birtwistle (2002). Uma avaliação diagnóstica pode ser realizada por um clínico geral devidamente treinado, ou por um psiquiatra ou psicólogo, e inclui uma infinidade de fatores que podem ser levados em consideração para um diagnóstico mais preciso, conforme Patton (2015).

O exame de saúde mental pode incluir o uso de uma escala de classificação, como a Hamilton Rating Scale for Depression, conforme expõe Zimmerman et al. (2004) ou o Suicide Behaviors Questionnaire-Revised, segundo Osman et al. (2001). Outras escalas como a Escala de Avaliação de Depressão de Montgomery-Asberg, Escala de Avaliação de Melancolia de Bech-Rafaelsen, Escala de Auto avaliação de Depressão de Zung, Inventário de Auto-avaliação de Wakedield, a Escala de Avaliação de Depressão de Carrol e Inventário de Depressão de Beck, também podem ser utilizadas na investigação de sintomas depressivos, conforme exibem Calil e Pires (1998).

A pontuação em uma escala de classificação por si só não é suficiente para diagnosticar depressão para satisfação da regulamentação do DSM ou CID, mas fornece uma indicação da gravidade dos sintomas por um período de tempo, de modo que uma pessoa que pontue acima de um determinado ponto de corte pode ser mais detalhadamente avaliada para um diagnóstico de transtorno depressivo, conforme Sharp e Lipsky (1998).

Para ser útil, uma escala investigativa de depressão deve ser capaz de explicar os principais sintomas da depressão em termos de princípios comportamentais estabelecidos empiricamente. O DSM-IV, lista os principais sintomas da depressão, o número e a duração dos sintomas que devem estar presentes para garantir um diagnóstico específico, que incluem feições deprimidas ou tristes, redução do interesse ou prazer em realizar atividades (anedonia), alterações de apetite (tanto ganho como perda de peso), alterações de sono (tanto insônia como excesso de sono), redução geral do nível de atividades (retardo psicomotor), agitação ou ansiedade, fadiga ou perda de energia, sentimentos de inferioridade e/ou culpa contínua acompanhados por autocrítica, recordação seletiva ou atenção para eventos negativos, distorção cognitiva e ideação suicida, conforme Dougher e Hackbert (2003).

Aliando os princípios estabelecidos no DSM destaca-se o Inventário de Depressão de Beck (BDI), compreendo uma escala avaliada e validada no Brasil, a qual foi utilizada nessa pesquisa em sua variação BDI-II, criada por Beck. Nesta seara, cabe delinear que Aaron Temkin Beck é um psiquiatra norte-americano e professor emérito do departamento de psiquiatria na Universidade da Pensilvânia, ele é conhecido como pai da Terapia Cognitiva e inventor das Escalas de Beck, que são vastamente utilizadas.

Beck é famoso por sua pesquisa em psicoterapia, psicopatologia, suicídio e psicométrica, que levou à criação da Terapia Cognitiva, pelo qual recebeu o Prêmio Lasker, também é o criador do BDI, um dos instrumentos mais utilizados como métrica dos sintomas da depressão. Beck acreditava que a depressão manifestava-se por causa das visões negativas não realistas sobre o mundo, onde as pessoas deprimidas teriam uma percepção negativa em três áreas, que são estabelecidas como a tríade depressiva, desenvolvem visões negativas sobre: elas mesmas (o self), o mundo e seu futuro, tendo grande relevância no desenvolvimento do quadro depressivo, um exemplo da tríade, retirado de Brown et al. (1995) é o caso do estudante que obteve maus resultados nas provas:

O estudante apresenta pensamentos negativos sobre o mundo, assim ele passa a acreditar que não gosta das aulas.

O estudante apresenta pensamentos negativos sobre seu futuro, pois pensa que não será aprovado na disciplina.

O estudante apresenta pensamentos negativos sobre o seu self, já que acredita que não merece estar na faculdade.

O desenvolvimento do BDI reflete estes problemas na sua estrutura com frases como: “eu perdi todo o interesse em outras pessoas”, para refletir o mundo; “eu me sinto desencorajado sobre o futuro”, para refletir o futuro; e “eu me culpo por tudo de ruim que acontece”, para refletir o self. A visão da doença depressiva como mantida por cognições negativas intrusivas tem particular aplicação na Terapia Cognitivo Comportamental, que surgiu com o objetivo de corrigir os pensamentos distorcidos e aliviar os sintomas depressivos, segundo Beck et al. (1996).

O BDI é uma escala de autoavaliação que teve a sua validade extensamente estudada e documentada na literatura especializada, o instrumento original, o Beck Depression Inventory, foi desenvolvido em 1961 por Beck e colaboradores para avaliar a sintomatologia depressiva, apoiado no paradigma teórico de que as cognições depressivas seriam as alterações psicopatológicas mais importantes para mensurar o quadro clínico de depressão, de acordo com Clak e Beck (1999).

A fácil aplicação do BDI e sua positiva aceitabilidade pelos usuários ampliou ainda mais o seu uso em um grande número de estudos, o BDI recebeu várias revisões do próprio autor para aperfeiçoar as necessidades clínicas e de pesquisa. Em 1974, foi desenvolvida uma forma abreviada com 13 itens; em 1978, a escala de 21 itens foi revisada, diferindo da original quanto ao tempo de referência da avaliação, “última semana” em vez de “hoje”, e por pequenas alterações na redação de seus itens, conforme Clak e Beck (1999).

O BDI original foi apoiado nas afirmações descritivas típicas dos sintomas que eram frequentemente relatados por pacientes psiquiátricos com depressão e, somente às vezes, por pacientes psiquiátricos sem depressão, as observações clínicas e as descrições dos pacientes foram sistematicamente consolidadas em 21 itens representativos de sintomas e atitudes depressivos, esses itens foram organizados de acordo com a intensidade do conteúdo das diferentes afirmações, a cada item foi atribuído um valor de acordo com uma escala de 4 pontos, variando de 0 a 3 em intensidade.

A versão original foi construída para ser aplicada por entrevistadores treinados que liam em voz alta as afirmações aos pacientes e em seguida, os pacientes selecionavam as afirmações de cada item que melhor se ajustavam ao seu estado atual de humor. Os entrevistadores geralmente levavam de 10 a 15 minutos para aplicar o instrumento e obter a pontuação total mediante a soma dos escores atribuídos pelos pacientes em cada um dos 21 itens, segundo Beck et al. (1996).

Desde 1971, Beck e seus associados começaram a empregar uma versão modificada do BDI, onde este eliminava expressões alternativas para os mesmos sintomas, como afirma Beck e Rush (1979).

O BDI-IA substituiu o instrumento original (BDI), e a edição de 1993 desse manual incluiu pequenas revisões da faixa de pontuação recomendadas para interpretar o nível de intensidade dos sintomas depressivos, de acordo com Beck et al. (1996).

Os sucessivos lançamentos do DSM-III-R e do DSM-IV, destacaram a necessidade de uma nova medida psicológica de depressão que avaliasse os sintomas conforme os critérios desses sistemas de classificação. Além disso, alguns dos sintomas originais do BDI que foram tipicamente observados, principalmente a mudança da autoimagem, perda de peso e preocupações somáticas, tornaram-se gradualmente menos úteis com o passar dos anos para a avaliação da intensidade da depressão, sendo assim iniciou-se a criação de uma nova versão do DBI, denominada DBI-II.

Beck e sua equipe começaram o trabalho piloto em 1994 com o Inventário de Depressão de Beck segunda edição (BDI-II), sendo esse um instrumento de auto aplicação composto por 21 itens, cujo objetivo é medir a intensidade da depressão em adultos e adolescentes a partir dos 13 anos de idade. O questionário completo e detalhado com os 21 itens encontra-se no H. Essa versão do BDI-II foi desenvolvida para avaliar os sintomas correspondentes aos critérios diagnósticos dos transtornos depressivos descritos no DSM-IV da APA.

Registra-se que após 35 anos de experiência e pesquisa com o BDI foi considerado propício revisar e modernizar a versão modificada do BDI-IA, descrito em Gorenstein et al. (2011). Assim, o BDI-II é uma ferramenta de avaliação consolidada no meio científico internacional, para medir a presença e gravidade de sintomas depressivos, tanto da população clínica como da população geral. Ao longo de sua existência, esse inventário alcançou grande popularidade entre pesquisadores e clínicos, sendo utilizado em numerosos estudos científicos sobre depressão no mundo todo.

No Brasil o BDI-II foi traduzido e adaptado por Cunha (2001), e passou por diversos estudos de validação. Sua fácil aplicação e aceitabilidade pelos usuários ampliou ainda mais o seu uso em diversas pesquisas.

Embora a reformulação do BDI-II tenha claros objetivos para se adequar aos critérios diagnósticos do DSM-IV, este instrumento não serve para fazer diagnóstico psiquiátrico por não envolver avaliação

clínica, entretanto, ele pode ser valioso para documentar a presença de sintomas depressivos e avaliar a sua gravidade, tanto em pessoas da população geral quanto em pacientes deprimidos diagnosticados clinicamente, conforme afirma Gorenstein et al. (2011).

Nesta versão do BDI-II, os quatro itens: perda de peso, mudança na autoimagem, preocupações somáticas e dificuldade de trabalhar foram retirados e substituídos por outros 4 novos itens: agitação, desvalorização, dificuldade de concentração e falta de energia, com o objetivo de identificar sintomas típicos de depressão grave, foram modificados dois itens para incluir tanto o aumento quanto a diminuição de apetite e sono, também foram reescritas algumas afirmações (ou alternativas) utilizadas para avaliar outros sintomas. Portanto a versão do BDI-II aqui referenciada compõe-se de 21 sintomas, sendo eles a agitação, a alteração no apetite, o choro, a falta de concentração, a crítica, a culpa, a desvalorização, a falta de energia, a falta de auto estima, a fadiga, o fracasso, a indecisão, a falta de interesse por sexo, a falta de interesse, a irritabilidade, o pessimismo, a perda de prazer, a punição, a alteração na rotina de sono, os pensamentos ou ideação suicida e o sentimento de tristeza.

Cabe aqui ressaltar estudos relevantes quanto a métrica da depressão por meio de escalas validadas no contexto científico no mundo e no Brasil.

Conforme Sauer et al. (2013) a depressão é um dos transtornos de humor mais relevantes clinicamente. Por esse motivo, muitos instrumentos de avaliação foram desenvolvidos para medi-la no intuito de auxiliar o diagnóstico. Nesse sentido, o BDI é um dos instrumentos com maior frequência de uso, sendo muito aplicado em estudos no mundo e no Brasil.

Ahrari et al. (2013) ao realizar um estudo para determinar a gravidade dos sintomas da depressão em pacientes vítima de queimaduras e avaliar o efeito sobre esses fatores e a ocorrência da depressão no Irã, usando o inventário BDI-II, com 184 voluntários: 58 com sintomas de depressão leve, 52 moderados e 74 severa, concluiu que o BDI é um instrumento altamente confiável para triagem e validação da depressão. Foi possível com esta pesquisa destacar uma relação significativa entre os sintomas de depressão e idade, sexo e nível educacional dos voluntários, além de ter sido descoberta a necessidade do diagnóstico e tratamento com terapias psicofarmacológicas e comportamentais no hospital, devendo o tratamento ser continuado até a reabilitação.

Em outro momento Manian et al. (2013) analisou as dimensões fa-

toriais do BDI-II, naturalmente e economicamente, em uma amostra de mulheres em situações de pós-parto, avaliando a contribuição relativa. Embora a diferenciação entre os sintomas da depressão e as alterações fisiológicas relativas ao pós-parto possam ser difíceis de serem detectadas, os achados desse estudo afirmam que: mais pesquisas são necessárias para entender melhor a relação das dimensões somáticas e outras dimensões com a sintomatologia, gravidade e funcionamento da depressão.

Já Oliveira et al. (2014) realizou um estudo comparativo para triagem de depressão em pessoas com epilepsia com a utilização dos instrumentos Neurological Disorders Depression Inventory for Epilepsy (NDDI-E), Hospital Anxiety and Depression Scale (HADS-D) e Beck Depression Inventory II (BDII), detectando que os três instrumentos avaliados têm utilidade clínica na triagem da depressão em pessoas com epilepsia. Sendo que, o NDDI-E e o HADS-D são instrumentos mais ágeis neste caso, e, o BDI-II se mostrou mais robusto, mais por exigir maior tempo para aplicação dificulta a sua utilização por clínicos sem tanta disponibilidade de tempo.

Com o interesse em validar uma Escala de Depressão Pós-Natal de Edimburgo e analisar sua estrutura fatorial, Loscalzo et al. (2015) realizou seu estudo utilizando duas amostras: a primeira composta de 334 pais onde 39 eram deprimidos; e, a segunda composta por 102 pais, dos quais 22 eram deprimidos. A partir da aplicação da versão italiana do EPDS, do IBDI-II e do Centro de Estudos Epidemiológicos Depression Scale não foi possível detectar a depressão, mas foi identificado um estado de angústia, incluindo sintomas depressivos, ansiedade e infelicidade.

Para Schutt et al. (2016) a triagem de sintomas depressivos é importante quando se avaliam candidatos à cirurgia bariátrica. O BDI-II e o Patient Health Questionnaire-9 (PHQ-9) são dois instrumentos de triagem de depressão amplamente utilizados. E, ao avaliar a semelhança no desempenho desses dois instrumentos descobriu que os escores do PHQ-9 e do BDI-II nos pacientes que pretendem realizar a cirurgia bariátrica estão estreitamente correlacionados. Concluindo que estes resultados apoiam o uso do BDI-II como uma alternativa viável para o PHQ-9 para triagem de pacientes que procuram a cirurgia bariátrica.

Diante da observação dos estudos descritos, consegue-se inferir que o BDI-II é utilizado nas mais diversas áreas de pesquisa, para os mais distintos objetivos específicos quando se trata da investigação de sintomas depressivos considerando o nível de depressão que cada paciente apresenta, sendo portanto o instrumento selecionado e utili-

zado para auxiliar no processo investigativo dos sintomas depressivos em usuários na rede social Facebook, como parte investigativa dessa pesquisa.

2.1 Mineração de dados: uma ferramenta para o conhecimento

Segundo Kumar et al. (2009), a mineração de dados é o processo de descoberta de informações úteis em grandes bases de dados com o intuito de descobrir padrões úteis que, ao contrário, poderiam permanecer ocultos. Elas também provêm a capacidade de previsão do resultado de uma observação futura.

Para Kotu e Deshpande (2015), a mineração de dados também é conhecida como descoberta de conhecimento, aprendizado de máquina e análise preditiva, no entanto, cada termo tem uma conotação ligeiramente diferente, dependendo do contexto. Witten et al. (2017), acrescenta que mineração de dados é um tópico prático e envolve aprender de forma prática, não de forma teórica, e que o processo pode ser automático ou semi-automático, sendo que os padrões descobertos devem ser significativos.

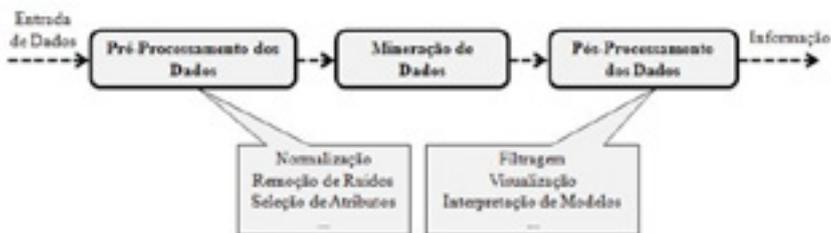
Witten et al. (2017), acrescenta que mineração de dados é um tópico prático e envolve aprender de forma prática, não de forma teórica, e que o processo pode ser automático ou semi-automático, sendo que os padrões descobertos devem ser significativos. Para o autor o processo de mineração de dados segue o modelo de referência CRISPDM, que é o acrônimo de Cross Industry Standard Process for Data Mining, e envolve conhecer o que se quer alcançar, seguindo as etapas relativas ao entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e desenvolvimento.

Kotu e Deshpande (2015), apresentam, além do processo CRISP-DM, os processos SEMMA, SAS, DMAIC, Sigma e KDD. Pelo fato de todas essas estruturas exibirem características comuns, além do fato de que um processo de mineração de dados deveria seguir um determinado conjunto de tarefas para alcançar seus objetivos, na presente pesquisa seguiremos o processo exposto em Kumar et al. (2009).

Kumar et al. (2009), em concordância com Kotu e Deshpande (2015), afirmam que a mineração de dados é uma parte integral da descoberta de conhecimento em banco de dados (KDD Knowledge Discovery in Databases), que é o processo geral de conversão de dados brutos em informações relevantes, conforme mostra a figura 2.2. Este processo consiste de uma série de passos de transformação, do pré-processa-

mento dos dados até o pós-processamento dos resultados da mineração de dados.

Figura 2.1: O processo de descoberta de conhecimento em banco de dados - KDD



Fonte: Introdução ao Data Mining, Kumar et al. (2009)

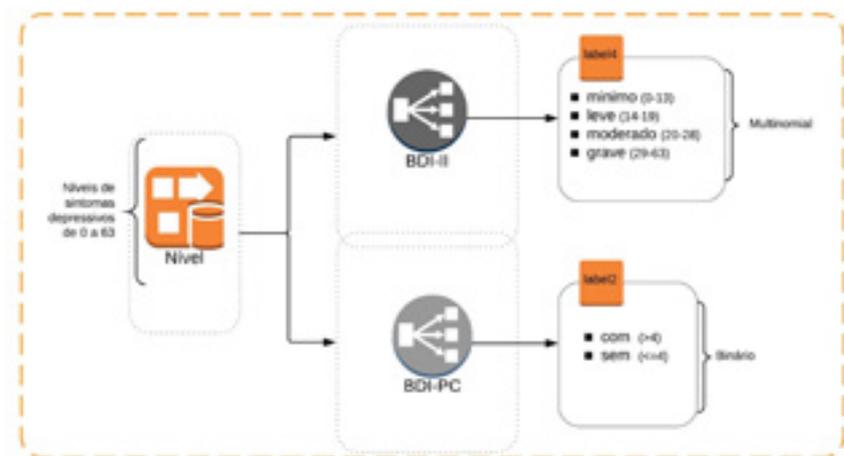
Ainda para Kumar et al. (2009), os dados de entrada podem ser armazenados em diversos formatos (arquivos simples, planilhas ou tabelas relacionais) e podem ficar em um repositório central de dados ou serem distribuídos em múltiplos locais. O propósito do pré-processamento é transformar os dados de entrada que estão na forma bruta, em um formato apropriado para análises subsequentes. As etapas envolvidas no pré-processamento de dados, incluem a fusão de dados de múltiplas fontes, a limpeza dos dados, observações duplicadas, seleção de registros e características que sejam relevantes à tarefa de mineração de dados. Por causa das muitas formas através das quais os dados podem ser coletados e armazenados, o pré-processamento de dados talvez seja o passo mais trabalhoso e demorado no processo geral de descoberta do conhecimento. Já a etapa de pós-processamento de dados deve assegurar que apenas resultados válidos e úteis sejam incorporados nas análises, nessa fase demonstram-se os padrões de filtragem, a visualização e a interpretação das análises.

Segundo Han et al. (2012), como um domínio altamente orientado a aplicativos, a mineração de dados incorporou muitas técnicas de outros domínios, como estatísticas, aprendizado de máquina, reconhecimento de padrões, banco de dados e sistemas de armazenamento de dados, recuperação de informação, visualização, algoritmos, computação de alto desempenho, entre outros. A natureza interdisciplinar da pesquisa e o desenvolvimento da mineração de dados contribui significativamente para a sucesso da mineração de dados e suas extensas aplicações.

Kumar et al. (2009) afirma que o trabalho que culminou na área de

mineração de dados, foi construído sobre a metodologia e algoritmos que os pesquisadores havia usado anteriormente, sendo a mineração de dados é constituída por áreas, como a (1) amostragem, estimativa e teste de hipóteses à partir de estatísticas e (2) algoritmos de busca, técnicas de modelagem e teoria de aprendizagem da inteligência artificial, reconhecimento de padrões e aprendizagem de máquina. Uma quantidade de outras áreas também desempenha papéis chave, em especial, os sistemas de banco de dados são necessários para fornecer eficiente suporte ao armazenamento, indexação e processamento de consultas. A figura 2.2 mostra o relacionamento da mineração de dados com outras áreas.

Figura 2.2: Mineração de dados com uma união em outras disciplinas tese-fluxo



Fonte: Introdução ao Data Mining, Kumar et al. (2009)

Observa-se na figura 2.2, que as áreas da estatística e da inteligência artificial ajudam a compor a área de mineração de dados.

De acordo com Kumar et al. (2009), as tarefas de mineração de dados são geralmente divididas em duas categorias principais, a primeira categoria está relacionada às tarefas de previsão, onde o objetivo dessas tarefas é prever o valor de um determinado atributo respaldado nos valores de outros atributos. O atributo a serem previsto é geralmente conhecido como variável dependente ou alvo, enquanto que os atributos usados para fazer a previsão são conhecidos como variáveis independentes ou explicativas. A outra categoria refere-se

às tarefas descritivas e seu objetivo é derivar padrões (correlações, tendências, grupos, trajetórias e anomalias), que possam resumir os relacionamentos subjacentes nos dados. As tarefas descritivas da mineração de dados são muitas vezes exploratórias em sua natureza e frequentemente requerem técnicas de pós-processamento para validar e explicar os resultados.

Em conformidade com Kotu e Deshpande (2015), de uma perspectiva histórica, existem duas classes principais das técnicas de análise preditiva: aquelas que evoluíram a partir de estatísticas como a regressão, e aquelas que emergiram de uma mistura de estatísticas, ciências e matemática. O interesse maior nessa pesquisa está voltado às tarefas referentes a categoria preditiva, onde, segundo Kumar et al. (2009), a regressão é uma das suas técnicas, sendo a sua variável alvo ou dependente, aquela que está sendo avaliada, de natureza contínua.

3 PROCEDIMENTOS METODOLÓGICO

Quanto à sua abordagem considera-se quantitativa, que segundo Marconi e Lakatos (1982) a define como a descrição objetiva sistemática e quantitativa do conteúdo manifesto da comunicação.

Para Marconi e Lakatos (1982), a análise quantitativa se efetua com toda a informação numérica resultante da investigação”, que se “apresenta como um conjunto de quadros, tabelas e medidas”. Essa quantificação deve ocorrer de forma sistemática e objetiva de acordo com, Marconi e Lakatos (1982): “Sistemática: ser ordenada e metódica; Objetiva: proceder de forma rigorosa e reaplicável, ou seja, evidenciando a objetivação dos fenômenos e a hierarquização das ações, devendo descrever, compreender e explicar os fatos”.

Corresponde à pesquisa de levantamento na medida em que se utiliza de questionários semiestruturados, como explica Silva (2005): “Levantamento: quando a pesquisa envolve a interrogação direta das pessoas cujo comportamento se deseja conhecer.”

E por fim, caracteriza-se como pesquisa experimental, que segundo Wazlawick (2010) caracteriza-se pela manipulação de um aspecto da realidade pelo pesquisador, implicando em se ter uma ou mais variáveis experimentais que podem ser controladas pelo pesquisador, e uma ou mais variáveis observadas, cuja mensuração poderá levar, à conclusão de que existe algum tipo de dependência com a variável experimental.

Os dados foram transformados em uma nova base em um formato adequado a ser consumido por uma ferramenta de mineração de

dados. A última etapa que precisou ser aplicada aos dados foi uma agregação temporal, tornando algumas variáveis dummies, para que os dados pudessem estar bem adequados à aplicação dos modelos de teste na fase da análise (processamento ou mineração dos dados).

4 A MINERAÇÃO DOS DADOS E NÍVEIS DE SINTOMAS DEPRESSIVOS: ANÁLISE E DISCUSSÕES

Tanto para Witten et al. (2017) quanto para Kumar et al. (2009), um processo de mineração de dados segue um determinado conjunto de tarefas para alcançar seus objetivos. Seguindo essa afirmação, na análise dos dados referentes ao segundo objetivo específico, que é preparar os dados coletados pela Escala BDI-II, foram utilizadas técnicas de mineração de dados, especificamente a tarefa de preparação ou pré-processamento dos dados, para categorizar a variável label4, conforme expõe a figura 4.2.

Figura 4.2: Figura que representa a categorização e binarização da variável nível



Fonte: Elaboração própria

Para tornar possível a utilização de um modelo de regressão logística multinomial, bem como seguir a denominação de cortes para o BDI-II, definida por Beck et al. (1996) e descrita em Gorenstein et al. (2011), a variável que representa o fenômeno de interesse, mantenedora nos 4 níveis de sintomas depressivos dos usuários, denominada por label4, precisou ser categorizada.

A escolha na utilização da categorização multinomial ao invés da binária, decorreu do fato de que a variável pôde ser classificada em 4 categorias oriundas de 21 sintomas, o que a tornou mais completa

e robusta, seria admissível reduzir as categorias para duas, mas isso poderia ocasionar alguma perda de informação ou causar imprecisão do que se está tentando averiguar, Allison (2014) afirma essa decisão.

Com relação ao modelo multinomial, essa categorização possibilitou a transformação da variável label4 em 4 variáveis dummy, decorrente da quantidade de possibilidades de categorias, isso foi necessário pois as técnicas de regressão logística são utilizadas quando o fenômeno a ser estudado apresenta-se de forma qualitativa, conforme define Fávero e Belfiore (2017).

Essas transformações foram necessárias e fundamentais para que a pesquisa pudesse continuar alinhada aos critérios dos modelos de regressão logística multinomial, transformações semelhantes podem ser destacadas nos trabalhos de Kuramoto et al. (2013) que categoriza a variável dependente em 3 categorias de ideação suicida: no ano passado, há mais de um ano, e nunca, Randall et al. (2014), categorizou em: sem ideação, com ideação e com ideação e planejamento, Madhu et al. (2014), referente a ocorrência de câncer de mama categorizou em alta, média e baixa, Guo et al. (2018), referente a causa de trauma na infância, classificou em: sintomas afetivos, psicóticos, ansiosos, maníacos isolados ou com todo o conjunto, Nierop e Germeys (2016) e Kingsbury et al. (2018), categorizaram as variáveis pela quantidade de sintomas depressivos presentes: dois, três ou quatro sintomas.

A necessidade de a pesquisa estar alinhada às especificações da regressão logística, está relacionada com seu objetivo principal que é estimar a probabilidade de ocorrência de um certo nível de comportamento depressivo, considerando o comportamento dos usuários na rede social.

CONSIDERAÇÕES FINAIS

Por meio de técnicas de mineração de dados este conjunto de informações, coerentes ao objeto desta pesquisa, foram preparadas resultando na classificação da severidade depressiva dos usuários com a utilização da Escala de Depressão de Beck BDI-II, foram obtidos os seguintes resultados médios: para a categoria mínimo a idade foi de 37,6 anos, a quantidade total de postagens foi de 1.347,9 e a quantidade total de curtidas foi de 389,6, a quantidade total de amigos foi de 917,8 e a quantidade de dias de interação ficou em 2.108,6.

Relativo aos sintomas depressivos de maneira geral, dos 21 sintomas, os que mais pontuaram foram, sono, energia, concentração, fadiga e crítica; para a categoria leve a idade foi de 33,3 anos, a quantidade

total de postagens foi de 1.384,5 e a quantidade total de curtidas foi de 533,4, a quantidade total de amigos foi de 913,4 e a quantidade de dias de interação ficou em 2.220,2.

Com relação aos sintomas depressivos, os que mais pontuaram foram, energia, sono, fadiga, concentração e indecisão/choro; para a categoria moderado a idade foi de 29,3 anos, a quantidade total de postagens foi de 1.661,1 e a quantidade total de curtidas foi de 688,1, a quantidade total de amigos pontuou de 888,8 e a quantidade de dias de interação ficou em 1.993,5.

No que tange os sintomas depressivos, as pontuações mais elevadas foram, energia/fadiga, sono, concentração, choro e indecisão; para a categoria grave a idade foi de 23,5 anos, a quantidade total de postagens foi de 1.784,06 e a quantidade total de curtidas foi de 880,8, a quantidade total de amigos pontuou de 1.111,1 e a quantidade de dias de interação ficou em 1.701,8, no que tange os sintomas depressivos, as pontuações mais elevadas foram, fadiga, estima, choro, crítica e desvalorização.

Diante dessas informações pode-se afirmar que as médias variam entre os 4 níveis de sintomas depressivos, havendo distinção entre elas, porém as possibilidades de exploração ainda não se esgotaram, e puderam ser verificadas de forma longitudinal.

REFERENCIAS

OMS(2017) OMS. <http://www.who.int/eportuguese/publications/pt/>.

Baldwin e Birtwistle(2002) D.S. Baldwin e J. Birtwistle. An Atlas of Depression. 2d ed. Philadelphia: Saunders. ISBN 978853625039-7.

Beck e Alford(2009) Aaron T. Beck e Brad A. Alford. Depression: Causes and Treatment. University of Pennsylvania Press.

Beck e Alford(2011) Aaron T. Beck e Brad A. Alford. Depressão: Causas e Tratamento. Artmed.

Beck e Rush(1979) Aaron T Beck e A J Rush. Cognitive therapy of depression. Guilford Press.

Beck et al.(1996) Aaron T Beck, RA Steer e GK Brown. Manual for the Beck Depression Inventory-II. Psychological Corporation. ISBN 978853625039-7.

Storch et al.(2004) EA Storch, JW Roberti e DA Roth. Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-Second Edition in a sample of college students. doi: 10.1002/da.20002.

Gorenstein et al.(2011) C Gorenstein, Wang Y-P, Argimon IL e Werlang BSG. Manual do Inventário de depressão de Beck - BDI-II. Editora Casa do Psicólogo.

Patton(2015) Lauren L. Patton. DEPRESSION. John Wiley and Sons. ISBN 9781118929285.

Zimmerman et al.(2004) Mark Zimmerman, Iwona Chelminski e Michael Posternak. A Review of Studies of the Hamilton Depression Rating Scale in Healthy Controls: Implications for the Definition of Remission in Treatment Studies of Depression . The Journal of Nervous and Mental Disease. ISSN 0022-3018. doi: 10.1097/01.nmd.0000138226.22761.39.

Osman et al.(2001) A. Osman, C. L. Bagge, P. M. Gutierrez, L. C. Konick, B. A. Kopper e F. X. Barrios. The Suicidal Behaviors Questionnaire-Revised (SBQ-R):Validation with Clinical and Nonclinical Samples. doi: 10.1177/107319110100800409.

Calil e Pires(1998) H. M. Calil e M. L. N. Pires. Aspectos gerais das escalas de avaliação de depressão.

Sharp e Lipsky(1998) L K Sharp e M S Lipsky. Screening for depression across the lifespan: a review of measures for use in primary care settings.

Dougher e Hackbert(2003) Michael Dougher e Lucianne Hackbert. Uma explicação analítico – comportamental da depressão e o relato de um caso utilizando procedimentos baseados na aceitação.

Brown et al.(1995) G P Brown, C L Hammen, M G Craske e T D Wickens. Dimensions of dysfunctional attitudes as vulnerabilities to depressive symptoms.

Cunha(2001) Jurema Alcides Cunha. Manual da versão em português das Escalas Beck. Casa do Psicólogo.

Sauer et al.(2013) Sebastian Sauer, Matthias Ziegler e Manfred Schmitt. Rasch analysis of a simplified Beck Depression Inventory. Personality and Individual Differences, 54(4):530–535. ISSN 01918869. doi: 10.1016/j.paid.2012.10.025. URL <http://dx.doi.org/10.1016/j.paid.2012.10.025>.

Ahrari et al.(2013) Farideh Ahrari, Seyyed Hamid Salehi, Mohammad Javad Fatemi, Madjid Soltani, Shahrzad Taghavi e Roghayeh Samimi. Severity of symptoms of depression among burned patients one week after injury, using Beck Depression Inventory-II (BDI-II). Burns, 39(2):285–290. ISSN 03054179. doi: 10.1016/j.burns.2012.07.012. URL <http://dx.doi.org/10.1016/j.burns.2012.07.012>.

Manian et al.(2013) Nanmathi Manian, Elizabeth Schmidt, Marc H. Bornstein e Pedro Martinez. Factor structure and clinical utility of BDI-II factor scores in postpartum women. Journal of Affective Disorders, 149(1-3):259–268. ISSN 01650327. doi: 10.1016/j.jad.2013.01.039. URL <http://dx.doi.org/10.1016/j.jad.2013.01.039>.

Oliveira et al.(2014) Guilherme Nogueira Oliveira, João Marcelo K. Lessa, Ana Paula Gonçalves, Eduardo Jardel Portela, Josemir W. Sander e Antonio Lucio Teixeira. Screening for depression in people with epilepsy: Comparative study among Neurological Disorders Depression Inventory for Epilepsy (NDDI-E), Hospital Anxiety and Depression Scale Depression Subscale (HADS-D) and Beck Depression Inventory (BDI). Epilepsy and Behavior, 34:50–54. ISSN 15255069. doi: 10.1016/j.yebeh.2014.03.003. URL <http://dx.doi.org/10.1016/j.yebeh.2014.03.003>.

Loscalzo et al.(2015) Yura Loscalzo, Marco Giannini, Bastianina Contena, Alessio Gori e Paola Benvenuti. The Edinburgh Postnatal Depression Scale for Fathers: A contribution to the validation for an Italian sample. General Hospital Psychiatry, 37(3):251–256. ISSN 18737714.

doi: 10.1016/j.genhosppsy.2015.02.002. URL <http://dx.doi.org/10.1016/j.genhosppsy.2015.02.002>.

Schutt et al.(2016) PE Schutt, S Kung, MM Clark, AM Koball e KB Grothe. Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) Depression Measures in an Outpatient Bariatric Clinic. General Hospital Psychiatry, 37(3):251–256. ISSN 18737714. doi: 10.1007/s11695-015-1877-2.

Kotu e Deshpande(2015) Vijay Kotu e Bala Deshpande. Predictive Analytics and

Data Mining - Chapter3: Data Exploration. 23138097. ISBN 9780128014608. doi: 10.1016/B978-0-12-801460-8.00010-0. URL <http://www.sciencedirect.com/science/article/pii/B9780128014608000100>.

Kumar et al.(2009) Vipin Kumar, Michael Steinbach e Pang ning Tan. Introdução ao Data Mining - Mineração de Dados. Ciência Moderna. ISBN 9788573937619.

Witten et al.(2017) Ian H Witten, Eibe Frank e Mark a Hall. Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). ISBN 0080890369. doi: 0120884070,9780120884070. URL <http://books.google.com/books?id=bDtLM8CODsQC&pgis=1>.

Han et al.(2012) Jiawei Han, Micheline Kamber e Jian Pei. Data Mining: Concepts and Techniques. 24520147. ISBN 978-0-12-381479-1. doi: 10.1016/B978-0-12-381479-1.00001-0.

Marconi e Lakatos(1982) Marina A Marconi e Eva M Lakatos. Técnicas de pesquisa: planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados. Atlas. ISBN 9788538600718.

Silva(2005) Edna Lucia Silva. Metodologia da pesquisa e elaboração de dissertação. rev. atual, 4 edição. ISBN 9788538600718.

Wazlawick(2010) Raul Sidnei Wazlawick. Uma Reflexão sobre a Pesquisa em Ciência da Computação à Luz da Classificação das Ciências e do Método Científico. Revista de Sistemas de Informação da FSMA, no 6:3-10.

Allison(2014) Paul D. Allison. Logistic Regression Using SAS: Theory and Application. SAS Institute, 2nd edition edição.

Fávero e Belfiore(2017) Luiz Paulo Fávero e Patrícia Belfiore. Manual De Análise De Dados. Elsevier. ISBN 8535270876, 9788535270877.

Kuramoto et al.(2013) Janet Kuramoto, Wilcox Holly e Latkin Carl. Social integration ansuicide-related ideation from a social network perspective: A longitudinal study among inner-city African Americans. Suicide and Life-Threatening Behavior, 43(4):366-378. ISSN 03630234. doi: 10.1111/sltb.12023.

Randall et al.(2014) Jason R. Randall, David Doku, Michael L. Wilson e Karl Peltzer. Suicidal behaviour and related risk factors among school-aged youth in the republic of Benin. PLoS ONE, 9(2):1-9. ISSN 19326203. doi: 10.1371/journal.pone.0088233.

Guo et al.(2018) Nan Guo, Thalia Robakis, Claire Miller e Alexander Butwick. Prevalence of Depression Among Women of Reproductive Age in the United States. Obstetrics & Gynecology, 131(4):671-679. ISSN 0029-7844. doi: 10.1097/AOG.0000000000002535.

Nierop e Germeys(2016) M Van Nierop e I Myin Germeys. Clinic risk associated with comorbidity of (subclinical) psychosis, anxiety and depressive symptoms: A case for stratified medicine in psychiatry. European Psychiatry, 33(SUPPL.):S49. ISSN 1778-3585. doi: <http://dx.doi.org/10.1016/j.eurpsy.2016.01.913>.

Kingsbury et al.(2018) A. M. Kingsbury, M. Plotnikova e J. M. Najman. Commonly occurring adverse birth outcomes and maternal depression: a longitudinal study. Public Health, 155:43-54. ISSN 14765616. doi: 10.1016/j.puhe.2017.11.001. URL <https://doi.org/10.1016/j.puhe.2017.11.001>.