

Análise de testes de personalidade e uso de drogas ilícitas com auxílio de técnicas de mineração de dados

Pedro Henrique Ventura Rodrigues de Almeida
PPGCC, dpt. de Ciência da Computação
Universidade Federal de Juiz de Fora
Juiz de Fora, Brasil
pedro.ventura@ice.ufjf.br

Abstract — Computing is increasingly present in various areas of knowledge, mainly as a tool for help and information extraction. However, the lack of knowledge of such tools by areas mainly unrelated to the exact sciences, such as psychology, end up losing great opportunities for exploration and generation of information that may be useful within the area itself. The purpose of this work is through data mining techniques, and with a database of psychological profiles and drug use, be able to discover psychological patterns for users of predetermined drug groups.

Keywords – data mining, personality and drug use, psychological profile analysis

I. RESUMO

A computação encontra-se cada vez mais presente nas mais diversas áreas de conhecimento, principalmente como ferramenta de auxílio e de extração de informação. Contudo, a falta de conhecimento de tais ferramentas por áreas principalmente não relacionadas a ciências exatas, como exemplo a psicologia, acabam perdendo grandes oportunidades de exploração e geração de informações que possam vir a ser utilizadas dentro da própria área. O intuito deste trabalho é através de técnicas de mineração, e uma base de dados sobre perfis psicológicos e uso de drogas, ser capaz e descobrir padrões psicológicos para usuários de grupos pré-determinados de drogas.

Palavras chaves — mineração de dados, personalidade e uso de drogas, análise de perfil psicológico

II. INTRODUÇÃO

A. Contextualização

É de senso comum que, pessoas possuam características físicas e psicológicas distintas dentre as demais. Além disso, há também linhas de pensamento (sociológico) em que “o homem é o conjunto das relações sociais” [1], ou seja, o homem seria o resultado de seus relacionamentos dentro a sociedade em que vive [2].

Logo, é de difícil indicação o que leva o ser humano ao uso de drogas, uma vez que, o mesmo possa ser ocasionado por fatores socioculturais, fuga da realidade, traumas, criação, problemas de saúde, entre outros. Contudo, através de traços psicológicos, medidos através de questionários vindo de pesquisas da área da Psicologia, seja possível talvez um estudo inicial em busca de existência (ou não) padrões psicológicos (mensuráveis como traços de personalidade)

que possam (ou não) estar correlacionados ao uso de certas substâncias.

Objetivo geral deste trabalho é através de técnicas de mineração supervisionadas, classificar perfis psicológicos em relação a uso de determinados tipos de drogas, com o intuito de obter perfis tendenciosos as mesmas. Inclusive, através do processo, também inferir a possibilidade do uso do ferramental de mineração por não especialistas, principalmente a pesquisadores da área de psicologia.

Três pontos valem ser ressaltados em relação a este trabalho:

1. O mesmo foi desenvolvido como resultado de estudos e trabalhos da disciplina de “Mineração de Dados” do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora (2019), não havendo domínio pelo autor na área da Psicologia e relacionadas;
2. A base utilizada para estudo foi obtida de repositório externo, e a mesma tendo um trabalho relacionado já aprovado por um comitê de ética.
3. Entendesse neste trabalho droga como: “produto químico que influencia a função biológica (exceto fornecendo nutrição ou hidratação)”[3]; e ilícitas aquelas as quais são obtidas de maneira ilegal, sem correlação ética ou moral.

B. Personalidade

Comumente utiliza-se de adjetivos para descrever pessoas, no caso, baseado geralmente em características mais aparentes/marcantes. Após estudos, há um acordo entre psicólogos que é possível definir personalidades através de cinco adjetivos e traços em cada um deles[4], ou seja, quanto mais propenso o indivíduo é a ter tal característica (adjetivo) ou não. Há na literatura o Five Factor Model (FFM) [5], que utiliza do modelo citado, composto por cinco características [6]:

- Neuroticismo (Neuroticism – N). Instabilidade emocional, tendência a experimentar emoções negativas como raiva, ansiedade ou depressão;
- Extroversão (Extraversion – E). Tendência a procura por estímulos e companhia de terceiros e tendência a emoções positivas;
- Abertura à Experiência (Openness to experience – O). Tendência a buscar pelo novo e novas

experiências, possuindo maior criatividade e curiosidade;

- Agradabilidade (Agreeableness – A). Tendência a ser compassivo e cooperante. Busca por boas relações sociais e pouca tendência a pensamentos negativos sobre outras pessoas;
- Conscienciosidade (Conscientiousness – C). Tendência a autodisciplina e planejamento para obtenção de objetivos.

III. BASE DE DADOS

C. Compreendendo a base selecionada

A base escolhida para o desenvolvimento deste é a “Drug consumption (quantified)”, obtida em [7], obtida no repositório UCI. A base possui 32 atributos com 1885 instâncias, sendo todas dados reais, coletados entre março de 2011 e março de 2012, em uma pesquisa feita por Elaine Fehrman, após ser aprovada a proposta da pesquisa em 2011 pelo comitê de ética da Psicologia Forense da Universidade de Leicester. A mesma não possui dados faltantes. A base é formada por coleta de dados de falantes da língua inglesa, com perguntas pessoais, testes de personalidade e uso de drogas.

Neste trabalho os 32 atributos foram agrupados em 3 subgrupos; Dados Pessoais, Psicológicos e Drogas:

- Pessoais: ID, Idade, Gênero, Educação, País e Etnia;
- Psicológicos: Nscore, Escore, Oscore, Ascore, Cscore, Impulsividade e Busca de Sensação
- Drogas: Álcool, Anfetamina, Nitrito de amila, Benzodiazepinas, Cafeína, Cannabis, Cetamina (ketamina), Chocolate, Cocaína, Cogumelos, Crack, Drogas Legais, Ecstasy, Heroína, LSD, Metanfetamina, Nicotina, Semer, VSA.

Tanto os dados Pessoais quanto os Psicológicos se encontravam “quantificados”, mas com o auxílio de uma tabela de tradução obtida em [7], os mesmos foram passados para valores reais, sendo eles numéricos ou categóricos. Já as Drogas se encontravam como valores categóricos (esboçados mais a frente)

C. Atributos Psicológicos

Durante a criação da base foi aplicado três testes, sendo um deles o NEO PI-R, um questionário para obter os níveis de personalidade citados em 1.2 (N, E, O, A e C), com valores que podem variar de 12 a 60. Um segundo teste foi aplicado, o “Escala de impulsividade de Barratt” (BIS-11), teste que possui o intuito de compreender a impulsividade do indivíduo, podendo ela ser por falta de planejamento, impulsividade por falta de atenção e impulsividade motora [8], sendo nesta base interpretada e editada para valores entre 0 à 9, com número em uma escala com relação a grau de impulsividade geral (menor para maior) [9]. Por fim, o terceiro teste aplicado foi o ImpSS, teste o qual busca associar impulsividade com busca por sensações, sendo apresentado na base em uma escala de 0 à 10 [9].

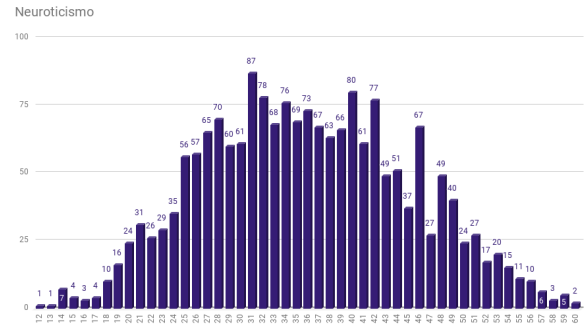


Fig. 1. Gráfico de barras da distribuição de dados referentes a "Neuroticismo (Nscore)" na base

D. Atributos Drogas

Para a base foi feito um levantamento em relação tempo da última vez do uso de 18 drogas, sendo uma delas fictícia, com o intuito de obter falsos respondentes;

- CL0 – Nunca usado;
- CL1 – Usado a mais de uma década;
- CL2 – Usado na última década;
- CL3 – Usado no último ano;
- CL4 – Usado no último mês;
- CL5 – Usado na última semana;
- CL6 – Usado no último dia.

E. Análise e tratamento da base

O objetivo inicial deste trabalho era buscar toda e qualquer relação possível em relação ao consumo de cada tipo de droga ilícita. Contudo, à medida que discussões acerca da base foram ocorrendo, encaminhou-se para análise somente dos perfis psicológicos e grupos de drogas ilegais. Alguns dos porquês serão descritos nos próximos tópicos referentes ao tratamento da base.

F. Remoção e tratamento de atributos e instâncias

A base não possui valores faltantes, não havendo a necessidade de tratamento de instâncias. Com o objetivo de manter a veracidade dos dados a serem analisados, instâncias as quais possuíam valores diferentes de CL0 (Nunca usado) para a droga fictícia (Semer), foram removidas. No final ficaram um total de 1877 (de 1885) instâncias.

Em relação aos atributos pessoais, exceto pelo atributo ‘sexo’ (gender) que possui aproximadamente 50% para ambos valores (masculino ou feminino), os demais possuem grande desbalanceamento em relação aos seus atributos, principalmente a ‘etnia’ (91,2% branco) e ‘país’ (55,4% Reino Unido). Baseado nisso, optou-se por remover todos os atributos relacionados a Dados Pessoais, para evitar um grande tratamento de dados que poderiam ocasionar uma grande remoção (em uma base já não muito grande) ou grande replicação de certas instâncias na fase de testes. Desta maneira, possibilita-se também o enfoque sobre uso de drogas e perfis para este trabalho.

Para que pudesse haver uma noção de grandeza e melhor compreensão dos algoritmos de mineração e de indivíduos que venham a interpretar a base, os atributos psicológicos foram normalizados entre 0 e 1. Os cinco atributos relacionados ao “big-five” (NEOACscore) foram

normalizados com valores entre o mínimo possível do teste 12 (tornando-se '0' após a normalização), e 60 o máximo possível (tornando-se '1' após a normalização). O BIS-11 teve 0 como menor e 9 maior. ImpSS menor valor sendo 0 e maior 10.

Como o trabalho possui como intuito analisar quem já fez e não fez uso de drogas, foi feito uma binarização dos valores contidos nas classes referentes aos Dados de Drogas. Valores referentes a 'CL0' (nunca usado) foram transformados para '0', e os demais como '1'. Apesar da bibliografia para qual a base foi criada [9] tratar usuários do grupo CL1 (Usado a mais de uma década) como 'não usuário', seguiu-se neste trabalho, o raciocínio de que o "período de tempo" era irrelevante se levar em relação a última vez do uso da droga e idade do indivíduo. Uma vez que o tempo é relativo ao indivíduo, um indivíduo "a" de 20 anos possa dar como CL6 (Usado no último dia) para uma droga, e posteriormente nunca mais utilizar, o que seria o caso de outro indivíduo "b" de 40 anos dado para a mesma droga CL1, sendo que poderia possuir a mesma idade que o indivíduo "a" quando a utilizou. Tendo então a partir deste ponto, valores com suas respectivas interpretações: 1(Já utilizou) e 0 (Nunca utilizou).

Seguindo classificações de [10] e [11], foi feito um agrupamento de todas as drogas, no intuito de encontrar alguma tendência sobre grupos de drogas e seus efeitos e perfis. Como o intuito era analisar drogas ilegais, drogas de obtenção legal que se enquadram em outros grupos foram desconsideradas, como por exemplo o Alcool que era enquadrado como "Depressivos".

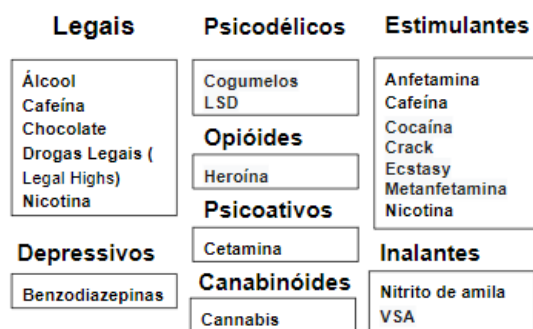


Fig. 2. Agrupamento das drogas por classe de droga

As drogas legais foram removidas, e as demais agrupadas conforme a figura X, definindo o valor '1' (já utilizou) caso já tenha utilizado alguma vez qualquer uma das drogas do grupo, e '0' (nunca utilizou) caso não tenha utilizado nenhuma das drogas do grupo. Por fim foram geradas sete bases. Todas contendo as sete classes de Dados Psicológicos (Nscore, Escore, Oscore, Ascore, Cscore, ImpSS e BIS-11) e cada uma com uma classe sendo um dos grupos de drogas (Psicodélicos, Estimulantes, Opióides, Psicoativos, Canabinóides, Inalantes ou Depressivos) como a classe alvo (target) da base e trabalho.

IV. MINERAÇÃO DE DADOS

Essa etapa consiste na aplicação de algoritmos já existentes na literatura para análise e descoberta de padrões, que devem ser analisados, sendo capazes de gerar informação, classificações ou previsões[12][13]. Os métodos devem ser escolhidos de acordo com o objetivo predeterminado e levando em consideração o estado da base, havendo talvez a necessidade de uma adaptação da base.

A. Algoritmos

Como o trabalho tem como foco compreender quais classes mais influenciam no momento da tentativa de predeterminar um usuário, dois métodos supervisionados foram selecionados devido a saída comumente geradas pelos mesmos.

A Regressão Logística trata-se de um modelo capaz de a partir de um conjunto, produzir um modelo capaz de prever valores de uma variável categórica (classe alvo)[14], como no caso deste o conjunto de resultados dos testes dos indivíduos, e a categoria de "Usuário" e "Não Usuário". Sendo utilizado então para através de coeficientes gerados, buscar quais classes mais importantes na determinação de um "Usuário (1)".

Já a segunda foi selecionada o algoritmo de Árvore de Decisão, para uma melhor visualização e compreensão de valores utilizados para definição de um "Usuário(1)" ou "Não usuário(0)". Uma árvore de decisão segue a ideia de uma árvore com "Nós" que possuam "Nós Filhos". Há também um nó raiz inicial. Cada nó "Pai" e nó "Filho" possuem regras para derivar (descer) de pai a filho. Quando se tenta prever algum dado, é feita uma busca na árvore gerada a partir de uma base de treinamento. A partir da percussão da árvore é tomada uma decisão, de que se pertence ou não a uma classe[15]. Esses nós e suas derivadas são aqui analisadas então para visualizar padrões de usuários.

Árvore de Decisão para Jogar Tênis

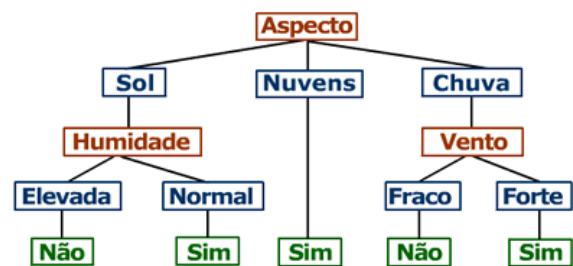


Fig.

3. Exemplo de árvore de decisão sobre Jogar Tênis. [15]
B. Treinamento e teste

Para poder realizar a predição/classificação, a base foi dividida em Treinamento, 70% e 30% para Teste. Como se tratava de uma base desbalanceada, em relação a uso ou não de drogas, após a separação, foi feito um "over-sample" com a técnica de SMOTE [17] sobre a parte de Treinamento (70%) para replicar instâncias da classe de menor incidência até possuir 50% de "Usuários" e "Não usuários".

Todo o processo foi feito individualmente em cada uma das sete bases geradas após a derivação da base original. Posteriormente foi submetido aos algoritmos de classificação, Regressão Logística e Árvore de Decisão. Os seguintes dados de acurácia foram obtidos em suas respectivas bases nos testes:

	Canabinoide s	Depressi vos	Estimulant es	Inalantes	Opióides	Psicoativos	Psicodélic os
AD	72%	52%	59%	55%	74%	65%	59%
RL	74%	65%	70%	62%	66%	65%	69%

Table 1. Acuraria dos algoritmos de Árvore de Decisão (AD) e Regressão Logística (RL) em cada base dos grupos de drogas

A matriz de confusão gerada por cada algoritmo de classificação, além da curva ROC podem ser visualizadas em [18].

C. Regressão Logística

Como o interesse é relacionado a descoberta de principais variáveis relacionadas o uso de drogas, através do coeficiente gerado pela regressão logística é possível ter uma visualização de quais atributos mais impactam na decisão do algoritmo no momento da classificação. No quadro a seguir se encontram os valores gerados. Sob o fundo vermelho se encontram a de maior impacto em cada Droga, e em azul o segundo de maior impacto.

Teste/ Drogas	Canabinoi des	Depressiv os	Estimulan tes	Inalantes	Opioides	Psicoativo s	Psicodélic os
Nscore	0.29	-0.08	0.35	0.589	1.33	0.13	-0.17
Escore	-2.57	-1.35	-1.12	-0.18	-2.45	-0.23	-1.51
Oscore	3.93	2.39	3.51	0.586	2.43	2.63	3.83
Ascore	-1.39	-1.16	-1.86	-0.93	-2.08	-1.43	-1.18
Cscore	-2.99	-1.27	-2.40	-1.53	-0.47	-1.13	-1.16
BIS-11	3.15	0.71	2.63	1.83	1.57	1.98	2.31
ImpSS	-0.15	0.67	-0.27	-0.11	0.82	-0.25	0.26

Table 2. Coeficiente da Regressão Logística das classes Psicológicas

D. Árvore de Decisão

Foi gerada uma árvore de decisão binária, sem podas. As imagens geradas podem ser obtidas em [18]. Tais imagens foram utilizadas para analisar as ramificações e buscar compreender quais comportamentos mais valiosos para cada tipo de droga.

Na geração da árvore foram utilizados dos seguintes termos para cara teste com o intuito de facilitar a leitura e compreensão de não especialistas:

Nscore → Instabilidade Emocional;

Escore → Extroversão;

Oscore → Abertura a experiências;

Ascore → Agradabilidade;

Cscore → Conscienciosidade;

BIS-11 → Busca por Experiência;

ImpSS → Impulsividade.

Segue em sequência o resultado das análises:

Canabinoides

Indivíduos identificados como “Já Utilizou” desta classe de drogas, em sua maioria obtiveram valores superior a 0.6 no teste de busca por experiência (BIS-11), e instabilidade emocional (Nscore) superior a 0.64. Valores referentes a um maior nível de conscienciosidade (Cscore, autodisciplina) se mostraram significativos.

Depressivos

Baixo valores de impulsividade (ImpSS, 0,22), aparentam decisivos para o “não uso” de depressivos. Contudo, indivíduos com instabilidade emocional (Nscore) superiores a 0.7 mostraram relevantes em relação a classificação de “Já

utilizou”, com uma forte ligação aparente com alta conscienciosidade (Cscore) e média/alta abertura a experiências (Oscore).

Estimulantes

Valores acima de 0.59 em busca por experiências (BIS-11) mostram-se relevantes para classificação de “Já utilizou”, seguidos de valores maiores para Abertura de experiências (Oscore), e media/baixa extroversão (Escore).

Inalantes

Busca por experiência (BIS-11) pouco acima da média demonstram grande impacto para classificação de “Uso”, seguido de variações entre impulsividade (ImpSS) e agradabilidade (Ascore), mas com grande impacto de valores mais significativos em abertura a experiências (Oscore).

Opioides

Baixa impulsividade demonstram de grande peso para classificação de “Não uso”. Valores altos para abertura a experiência (Oscore) e busca por experiências (BIS-11) marcam como determinante para classificação de “uso”.

Psicoativos

Baixa instabilidade emocional (Nscore) ganha destaque para classificação de “Nunca utilizou”. Valores altos de BIS-11 e Oscore demonstram peso para classificação de “Já utilizou”.

Psicodélicos

Busca por experiência e Abertura a experiência mais uma vez demonstram ser os atributos mais significativos quanto a classificação de “Já utilizou”. Quanto maior o valor, maior a quantidade de “usuários”. Baixos valores em instabilidade emocional (Nscore) aparenta também influenciar para “Já utilizou”.

V. RESULTADOS

Este trabalho buscou em seu primeiro momento buscar por padrões que pudessem ser encontrados em relação a perfis psicológicos e certos usos de drogas. Apesar de uma acurácia não muito significativa, tanto a Árvore de Decisão quanto a Regressão Logística apontam em uma mesma direção para classificar “Usuários”.

Em ambos os algoritmos de classificação, há uma grande relevância nas classes de Oscore e BIS-11. Ou seja, o quanto o indivíduo se encontra disposto ou em busca de experiências, influência diretamente sobre as chances do mesmo ser classificado como “Usuário”. Em outra perspectiva, quanto maiores valores nos testes de busca e abertura a experiência, maior a tendência ao “Uso” de drogas.

Vale ressaltar que tais resultados sobre “Experiência” fora encontrado em todos os tipos de drogas. Sendo uma característica dominante ao fato de se utilizar de substâncias químicas, e não a somente uma específica.

VI. CONCLUSÃO

Apesar de a acurácia encontrada desencorajar a “validade” dos algoritmos e resultados, vale ressaltar que a base trata-se de pesquisas feitas com humanos. Seres que possuem grande variedade de personalidade e sucessíveis a erro, inclusive a erros de preenchimento de questionários, método o qual foi utilizado para alimentação da base.

Compreende-se aqui a complexidade de obter tais dados referentes a pesquisas deste nicho, contudo, o desbalanceamento com referência a “Uso” ou não de certas drogas

Não foi possível encontrar alguma característica muito específica que possa influenciar o uso de cada tipo de droga como era desejado, mas resultados apontam para uma grande característica dominante, a de Abertura e Busca por Experiência. Sendo quanto maior o valor do resultado do teste, maior a tendência ao uso. Assim sendo esta uma das maiores contribuições deste.

Apesar de a Regressão Logística apontar diretamente qual classe mais impactante do modelo, a Árvore de Decisão possibilitou uma melhor interpretação humana do ocorrido, sendo possível inferir inclusive alguns traços de outras classes menos significativas (e seus valores) que impactam para o “uso” ou não de drogas específicas.

Por fim, mostrou-se viável a utilização de técnicas de mineração para compreensão de uma parte da psiquê humana obtida através de testes. Esperasse que este então possa servir de motivação para mais pesquisas futuras correlacionadas a psicologia e técnicas de mineração.

VII. TRABALHOS FUTUROS

Para trabalhos futuros sugere-se um estudo com abrangência maior dos dados da base, com inclusão de atributos pessoais, como o de grau escolar, que, em fase de primeira análise dos dados, demonstrou possuir grande correlação com uso de grupos específicos de drogas,

Em debates com profissionais clínicos que realizam atendimento, levantou-se a sugestão de em caso de pesquisas futuras que envolvam a criação de uma nova base, dar mais enfoque em informações Pessoais, principalmente em questões financeiras, socioculturais, existência de traumas e grau de formação.

Atenção também deve ser dada a questão do “tempo”, em relação a idade de indivíduos, como períodos de uso de substâncias e compreensão de se tratar ou não de um indivíduo com vícios, ou já recuperado.

REFERÊNCIAS

[1] GRAMSCI, A. Concepção dialética da História. 3.ed. Rio de Janeiro: Civilização Brasileira, 1978.

[2] OLIVEIRA, Betty. As relações entre conhecimento e valoração no trabalho educativo. Perspectiva, Florianópolis, v. 11, n. 19, p. 29-46, jan. 1993. ISSN 2175-795X. Disponível em: <<https://periodicos.ufsc.br/index.php/perspectiva/article/view/9167/8506>>. Acesso em: 03 set. 2019.

[3] Kleiman, M.A., Caulkins, J.P., Hawken, A.: Drugs and Drug Policy: What Everyone Needs to Know. Oxford University Press, New York (2011)

[4] Fehrman, Elaine & Egan, Vincent & Gorban, Alexander & Levesley, Jeremy & Mirkes, Evgeny & Muhammad, Awaz. (2019). Personality Traits and Drug Consumption: A Story Told by Data. 10.1007/978-3-030-10442-9.

[5] Costa, P.T., MacCrae, R.R.: Revised NEO-personality inventory (NEO PI-R) and NEO five-factor inventory (NEO

FFI): Professional manual. Psychological Assessment Resources, Odessa, FL (1992)

[6] BIG FIVE (PSICOLOGIA). In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2019. Disponível em: <[https://pt.wikipedia.org/w/index.php?title=Big_Five_\(psicologia\)&oldid=55749880](https://pt.wikipedia.org/w/index.php?title=Big_Five_(psicologia)&oldid=55749880)>. Acesso em: 16 jul. 2019.

[7] Fehrman, Elaine & Egan, Vincent & Gorban, Alexander & Levesley, Jeremy & Mirkes, Evgeny & Muhammad, Awaz. Drug consumption (quantified) Data Set. Disponível em: <<https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29>>. Acesso em: 03 set. 2019.

[8] Vasconcelos, A.G., 2012. Adaptação cultural e investigação das propriedades psicométricas da Barratt Impulsiveness Scale (BIS-11).

[9] Fehrman, E., Muhammad, A.K., Mirkes, E.M., Egan, V. and Gorban, A.N., 2017. The five factor model of personality and evaluation of drug consumption risk. In Data Science (pp. 231-242). Springer, Cham. <https://leicester.figshare.com/articles/Drug_consumption_database_original_csv/7588415>. Acesso em: 03 set. 2019.

[10] Editorial Staff. Drug Classifications. <luxury.rehabs.com/drug-abuse/classifications/>. Acesso em: 03 set. 2019.

[11] adf.org. Drug Wheel. <<https://adf.org.au/insights/drug-wheel/>>. Acesso em: 03 set. 2019.

[12] Camilo, C.O. and Silva, J.C.D., 2009. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG), pp.1-29.

[13] Frederico de Oliveira Sales, Yan Mendes Ferreira, Francisco H C Ferreira, Bruno José Dembogurski, Gustavo Silva Semaan, Edelberto Franco Silva. Evasão no Ensino Básico da Rede Pública Municipal de Juiz de Fora: uma Análise com Mineração de Dados. (2018).

[14] Logistic Regression. <<http://userwww.sfsu.edu/efc/classes/biol710/logistic/logisticreg.htm>>. Acesso em 04 set. 2019

[15] <https://medium.com/machine-learning-beyond-deep-learning/%C3%A1rvores-de-decis%C3%A3o-3f52f6420b69>

[16] Árvore de Decisão <<http://web.tecnico.ulisboa.pt/ana.freitas/bioinformatics.ath.cx/bioinformatics.ath.cx/indexf23d.html?id=199>>. Acesso em: 05 set. 2019.

[17] Fernández, Alberto & Garcia, Salvador & Herrera, Francisco & Chawla, Nitesh. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Journal of Artificial Intelligence Research. 61. 863-905. 10.1613/jair.1.11192.

[18] Conteúdo em anexo do autor. <<https://github.com/ArtFicer/Minera-odeDados-DrugsBigFive/>>. Acesso em 05 set. 2019.