

Trabalho - Etapa 2

Resumo

O objetivo é desenvolver uma aplicação para avaliar e prever o risco de envolvimento de um usuário com drogas ilícitas a partir de seus dados de perfil, traços de personalidade obtidos através de testes psicológicos (NEO-FFI-R¹, BIS-11² e ImpSS³) e histórico de consumo de drogas lícitas e ilícitas. O resultado da avaliação de risco será um valor em porcentagem da chance estimada de o usuário se envolver com qualquer droga ilícita, assim como uma porcentagem específica para cada droga ilícita de um conjunto pré-determinado. Para tal, utiliza-se técnicas de Machine Learning e uma base de dados de 1885 instâncias resultados de uma pesquisa de consumo de drogas.

Requisitos do Modelo Conceitual

- Armazenamento de 32 atributos para cada entrada (usuário), sendo eles:
 - 5 atributos de dados pessoais;
 - 7 atributos que representam traços de personalidade;
 - 20 atributos relacionados ao consumo de drogas e substância pelo usuário.
- Inserção de novas entradas (usuários) com seus respectivos dados. Todos os atributos de dados pessoais e de personalidade são obrigatórios para inserção.
- Remoção de entradas (usuários) com seus respectivos dados.
- Avaliação de dados para determinação de relações entre o consumo de drogas e demais características do usuário para cálculo de risco.

Fonte de Dados

Título: Drug consumption (quantified) Data Set

Download: [Data Folder](#)

¹ Este exame avalia os seguintes traços de personalidade:

- O Neuroticismo (N) é uma tendência de longo prazo de experimentar emoções negativas, como nervosismo, tensão, ansiedade e depressão;
- A Extroversão (E) manifesta-se em extrovertida, calorosa, ativa, assertiva, falante, alegre e em busca de características de estimulação;
- A abertura à experiência (O) é uma apreciação geral da arte, de ideias incomuns e de interesses imaginativos, criativos, não convencionais e amplos;
- A Amabilidade (A) é uma dimensão das relações interpessoais, caracterizada pelo altruísmo, confiança, modéstia, bondade, compaixão e cooperação;
- A Consciência (C) é uma tendência da pessoa ser organizada e confiável, obstinada, persistente, confiável e eficiente.

² Este exame avalia impulsivity.

³ Este exame avalia sensation seeking.

Resumo: Classificar o tipo de consumidor de drogas por meio de dados sobre sua personalidade

Características da base de dados: Multivariado

Número de instâncias: 1885

Área: Social

Características atributo: Real

Número de atributos: 32

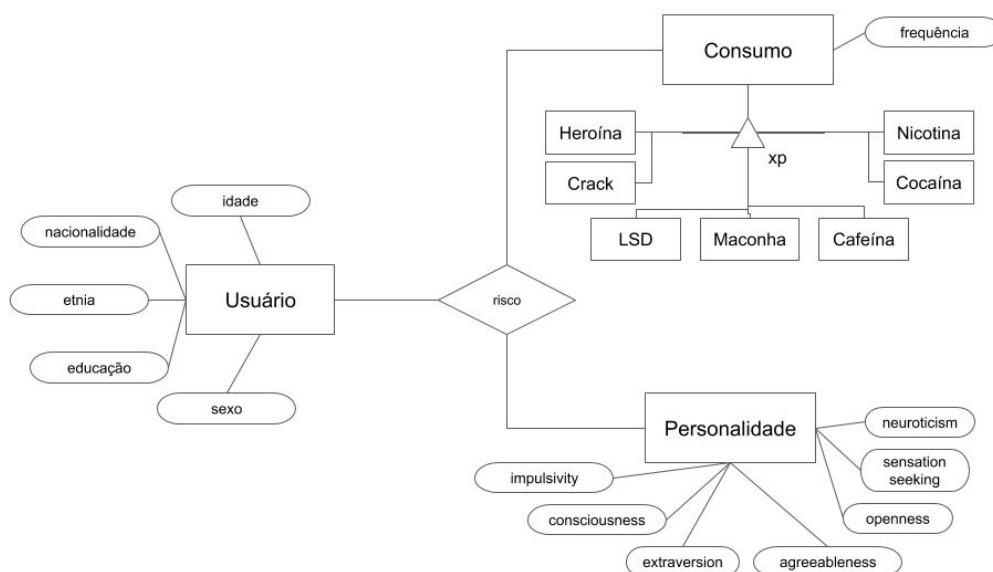
Data de doação: 17/10/2016

Link: <http://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29#>

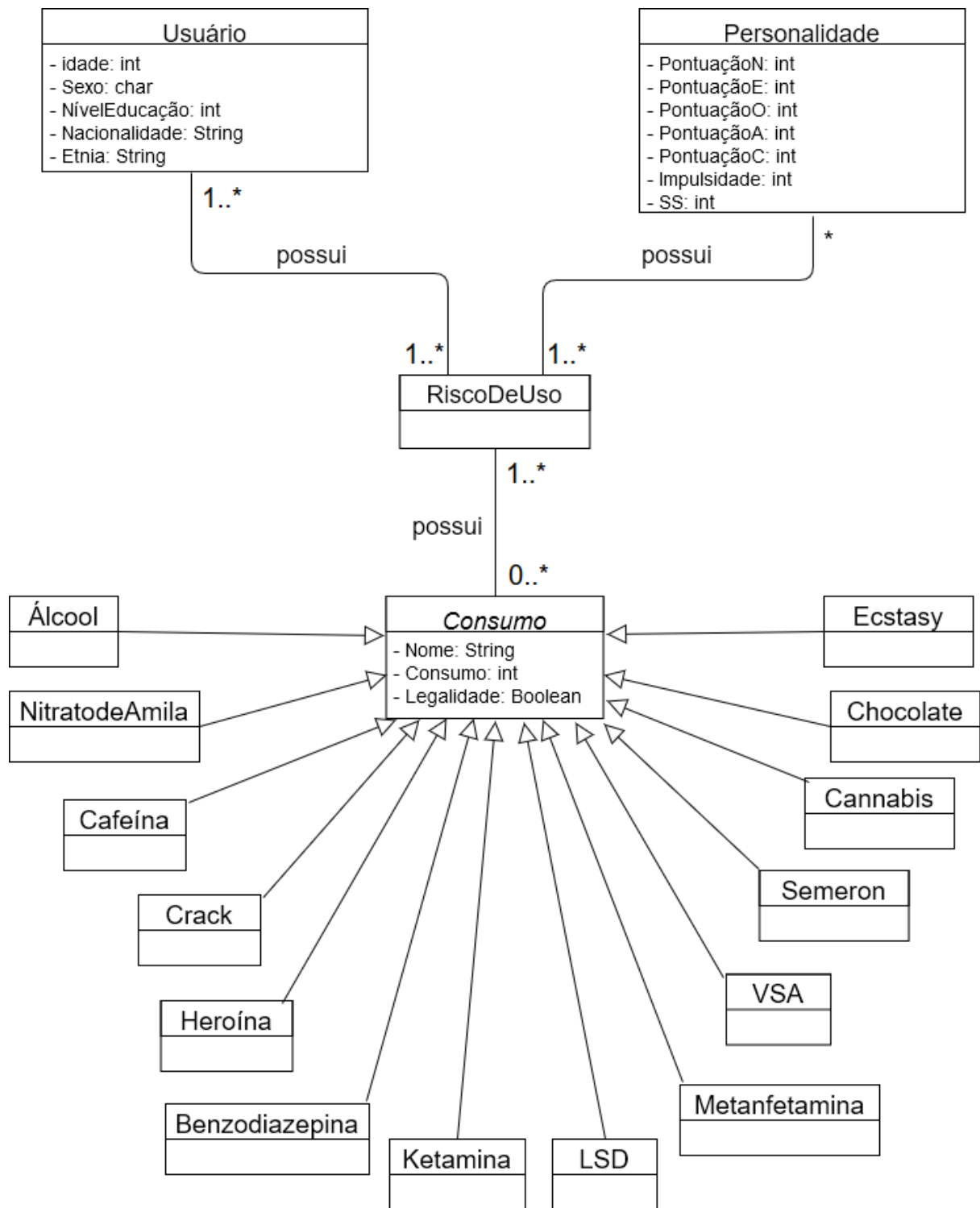
Descrição: base de dados contendo registros para 1885 respondentes. Para cada entrevistado, 12 atributos são conhecidos: Medições de personalidade que incluem NEO-FFI-R (neuroticismo, extroversão, abertura à experiência, amabilidade e conscienciosidade), BIS-11 (impulsividade) e ImpSS (busca de sensações), nível de educação, idade, sexo, país de residência e etnia. Todos os atributos de entrada são originalmente categóricos e são quantificados. Após a quantificação, os valores de todos os recursos de entrada podem ser considerados de valor real. Além disso, os participantes foram questionados sobre o uso de 18 drogas legais e ilegais (álcool, anfetaminas, amil nitrito, benzodiazepina, cannabis, chocolate, cocaína, cafeína, crack, ecstasy, heroína, ketamina, bebidas alcoólicas, LSD, metadona, cogumelos, nicotina e abuso de substância volátil e uma droga fictícia (Semeron) que foi introduzida para identificar os reclamantes em excesso. Para cada droga eles têm que selecionar uma das respostas: nunca usaram a droga, a usaram mais de uma década atrás, ou na última década, ano, mês, semana ou dia. Base de dados contém 18 problemas de classificação. Cada uma das variáveis de rótulo independente contém sete classes: "Nunca usado", "Usado durante uma década", "Utilizado na última década", "Utilizado no ano passado", "Utilizado no último mês", "Utilizado na última semana" e "usado no último dia".

Modelo Conceitual

- ER:



- UML:



Modelo Lógico

Não fizemos alterações no modelo conceitual da etapa 1 para esta. Para o modelo lógico, deixamos separadas as informações de risco de uso de drogas ilícitas (que serão calculadas através de machine learning) das informações de consumo. Assim, a tabela RiscoUso_ilícitas será a saída da execução do algoritmo de machine learning, ou seja, será preenchida com dados resultantes do machine learning. Todos os demais dados das demais tabelas são provenientes da fonte de dados escolhida na etapa 1. As informações de consumo foram distribuídas em várias tabelas, uma para cada produto de consumo.

- Usuario(id, idade, sexo, nívelEducação, nacionalidade, etnia);
- Personalidade(user_id, pontuaçãoN, pontuaçãoE, pontuaçãoO, pontuaçãoA, pontuaçãoC, impulsividade, pontuaçãoSS);
 - FK: user_id -> id em Usuario;
- RiscoUso_ilícitas(user_id, geral, amphetamines, amyl_nitrite, benzodiazepine, cannabis, cocaine, crack, ecstasy, heroin, ketamine, legal_highs, LSD, methadone, mushrooms, VSA);
 - FK: user_id -> id em Usuario;
- Alcohol(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Amphetamines(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Amyl_nitrite(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Benzodiazepine(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Cannabis(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Chocolate(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Cocaine(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Caffeine(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Crack(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Ecstasy(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Heroin(user_id, consumo);
 - FK: user_id -> id em Usuario;
- LSD(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Methadone(user_id, consumo);
 - FK: user_id -> id em Usuario;
- Mushrooms(user_id, consumo);

- FK: user_id -> id em Usuario;
- Nicotine(user_id, consumo);
 - FK: user_id -> id em Usuario;
- VSA(user_id, consumo);
 - FK: user_id -> id em Usuario;