

Processo Seletivo 2024

IEEE Computational Intelligence Society, UnB

Instruções Gerais

1. Esse desafio tem caráter classificatório e eliminatório no Processo Seletivo para o CIS.
2. Sua solução deve ser entregue, via e-mail para pscisunb2024@gmail.com, até às **23h59 de quinta, 11/07**. Com o título: Desafio_CIS_01_2024_PrimeiroNome_SegundoNome.
Exemplo: Desafio_CIS_01_2024_João_Alencar
3. Essa etapa visa avaliar as habilidades dos candidatos em manipular, visualizar e interpretar grandes volumes de dados, de forma em que a classificação para a segunda etapa não depende obrigatoriamente de responder todas as perguntas de forma correta.
4. O envio é recomendado ainda que o candidato tenha desenvolvido apenas uma solução parcial, visto que **todo o desenvolvimento será considerado**.
5. É esperado que o candidato apresente soluções originais e de própria autoria.
 - a. Caso seja constatada a influência de terceiros na solução apresentada, o candidato será sumariamente desclassificado do presente Processo Seletivo.
6. Soluções enviadas após o prazo limite serão desconsideradas.
7. Em caso de dúvidas, não hesite em nos contatar por meio do e-mail pscisunb2024@gmail.com.

Introdução

É cada vez mais evidente a relevância da análise de dados em diversas esferas, sendo essa uma ferramenta fundamental na tomada de decisões. Especificamente, coletar e examinar informações sobre o consumo de drogas nos diferentes países possibilita uma comparação mais exata e equilibrada entre as regiões, considerando suas singularidades. Isso facilita a formulação de estratégias mais eficazes e bem direcionadas para combater o problema do uso de substâncias ilícitas.

Dataset e análise de dados

Neste contexto, o Desafio proposto utilizará um dataset contendo dados sobre drogas, ([Link do Dataset](#)). Para cada respondente, são conhecidos 12 atributos: medidas de personalidade, que incluem NEO-FFI-R (neuroticismo, extroversão, abertura à experiência, amabilidade e conscienciosidade), BIS-11 (impulso) e ImpSS (procura por sensação), nível de educação, idade, gênero, país de residência, etnia e renda.

Dicionário das variáveis:

1. **ID:** é um número de registros em um banco de dados original. Não pode ser relacionado ao participante. Pode ser usado apenas como referência.
2. **Age (Real):** é a idade do participante.
3. **Gender:** Masculino ou Feminino.
4. **Education:** nível de educação do participante.
5. **Country:** país de origem do participante.
6. **Ethnicity:** etnia do participante.
7. **Income (USD):** renda mensal em dólares.
8. **Nscore (Real):** é o Neuroticismo da NEO-FFI-R.
9. **Escore (Real):** é a Extraversão da NEO-FFI-R.
10. **Oscore (Real):** é a Abertura à Experiência da NEO-FFI-R.
11. **AScore (Real):** é a Amabilidade da NEO-FFI-R.
12. **CScore (Real):** é a Conscienciosidade da NEO-FFI-R.

13. **Impulsive (Real):** é impulsividade medida pelo BIS-11.
14. **SS (Real):** é percepção sensorial medida pela ImpSS.
15. **Alcohol:** consumo de álcool.
16. **Amphet:** consumo de anfetaminas.
17. **Amyl:** consumo de nitritos (poppers).
18. **Benzos:** consumo de benzodiazepínicos.
19. **Caff:** consumo de cafeína.
20. **Cannabis:** consumo de maconha.
21. **Choc:** consumo de chocolate.
22. **Coke:** consumo de cocaína.
23. **Crack:** consumo de crack.
24. **Ecstasy:** consumo de êxtase.
25. **Heroin:** consumo de heroína.
26. **Ketamine:** consumo de quetamina.
27. **Legalh:** consumo de altos legais.
28. **LSD:** consumo de LSD.
29. **Meth:** consumo de metadona.
30. **Mushrooms:** consumo de cogumelos mágicos.
31. **Nicotine:** consumo de nicotina.
32. **Semer:** classe de consumo fictício de droga Semeron (ou seja, controle).
33. **VSA:** classe de consumo de abuso de substâncias voláteis.

Classificações de Uso de Drogas:

- CL0 Nunca Usou
- CL1 Usou Mais de Uma Década Atrás
- CL2 Usou nos Últimos Dez Anos
- CL3 Usou no Último Ano (59 vezes)
- CL4 Usou nos Últimos Meses
- CL5 Usou na Última Semana
- CL6 Usou Hoje



Brasília, 27 de junho de 2024

Universidade de Brasília Faculdade de
Tecnologia, Departamento de Engenharia
Elétrica

Sendo assim, realize os tratamentos que considerar necessários e responda, da forma mais completa possível, às seguintes questões:

1. No dataset existem alguns valores faltantes. Antes de começar a manipular os dados, trate essas informações e descreva sucintamente as alterações feitas;
2. Qual é a distribuição da idade dos indivíduos na amostra? Existem diferenças significativas nas faixas etárias predominantes de consumo entre os grupos de usuários de diferentes substâncias?
3. Há uma relação entre o nível educacional e o consumo de substâncias?
4. Como o gênero influencia no consumo de drogas alucinógenas (LSD, Ecstasy, Ketamine, Cannabis e Mushrooms)? Explique.
5. Qual é a proporção de participantes que se auto-classificam como impulsivos (score superior a zero)? Existe uma correlação entre a impulsividade e o consumo de substâncias?
6. Classifique as variáveis entre qualitativas (ordinal ou nominal), ou quantitativas (discreta, contínuas).
7. Qual é a proporção de consumo de substâncias legais versus ilícitas na amostra (considere a definição de legalidade segundo a legislação brasileira)?
8. Quais fatores predizem a probabilidade de um indivíduo consumir crack (Crack)?
9. Qual é a média das pontuações Nscore, Escore, Oscore, AScore, Cscore? Calcule a correlação entre elas.
10. Analise a relação entre o nível de educação (Education) e o consumo de diferentes substâncias ilícitas (como LSD, Amphet, Cannabis, etc.). Identifique se há uma correlação significativa entre essas variáveis e, em

caso afirmativo, explore a natureza dessa correlação (positiva/negativa).

11. Treine uma árvore de decisão para prever se um indivíduo consome uma determinada substância (por exemplo, álcool, anfetaminas, cannabis) com base em suas características demográficas e pontuações de personalidade. Utilize a acurácia para avaliar os seus resultados.
12. Explore a correlação entre a idade (variável Age) e a experimentação de diferentes substâncias ilícitas. Verifique se há uma tendência de aumento ou diminuição do consumo conforme a idade avança.
13. Quais são as 3 drogas mais utilizadas para cada país presente na amostra? E quais são as 3 menos utilizadas?



Brasília, 27 de junho de 2024

Universidade de Brasília Faculdade de
Tecnologia, Departamento de Engenharia
Elétrica

Entrega das submissões

Os dados podem ser analisados utilizando quaisquer ferramentas de sua preferência. Recomenda-se o uso das elencadas abaixo, **mas não estão limitadas a:**

- Python;
 - Pandas;
 - Matplotlib;
 - Numpy;
 - Sklearn.
- Matlab;
- R.

Todos os gráficos utilizados para responder os questionamentos devem ser entregues, bem como todas as linhas de código produzidas. No intuito de facilitar o processo e agilizar a etapa de correção, pedimos que um relatório seja preenchido conforme o fornecido [neste link](#) ("Template de Resposta") e entregue por e-mail para pscisunb2024@gmail.com até a **data limite de 11/07, às 23h59**. Cada dia de antecedência no envio gerará um acréscimo diário de 0,5% à nota final do candidato. Tratamento dos dados, boa organização, coerência dos códigos e qualidade das visualizações gráficas também serão avaliadas. Boa sorte!

Ps: procure deixar clara a linha de raciocínio adotada.