

CURSO: SISTEMAS DE INFORMAÇÃO
DISCIPLINA: SISTEMAS DE APOIO A TOMADA DE DECISÃO
TRABALHO AV1

Previsão de Desempenho no ENEM usando k-Nearest Neighbors (k-NN)

Contexto:

O Exame Nacional do Ensino Médio (ENEM) é uma das principais avaliações educacionais no Brasil, sendo utilizado como critério de acesso ao ensino superior. Os microdados do ENEM fornecem informações detalhadas sobre os participantes, incluindo suas notas nas diferentes áreas do conhecimento, informações socioeconômicas, e dados sobre a escola de origem. Com esses dados, é possível prever o desempenho de novos participantes no exame.

Situação Problema:

Você foi contratado por uma instituição de ensino que deseja prever o desempenho de futuros candidatos com base em suas características pessoais e socioeconômicas, bem como seu histórico escolar. O objetivo é identificar alunos que possam precisar de apoio adicional para melhorar seu desempenho no ENEM.

Objetivo da Tarefa:

Desenvolver uma aplicação em Python que implemente o algoritmo k-Nearest Neighbors (k-NN) para prever a nota de um novo aluno no ENEM. A previsão será baseada nas notas e nas características dos alunos que participaram do exame em anos anteriores.

Base de Dados Pública:

Você usará os **Microdados do ENEM**, que estão disponíveis para download e incluem uma vasta gama de informações sobre os participantes do exame.

- Base de Dados: Microdados do ENEM
- Local de Acesso:
<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>

Atributos Relevantes:

Os seguintes atributos dos microdados do ENEM serão utilizados para a previsão:

- Notas nas Áreas do Conhecimento:
 - Nota em Matemática e suas Tecnologias (NU_NOTA_MT)
 - Nota em Ciências da Natureza e suas Tecnologias (NU_NOTA_CN)
 - Nota em Linguagens, Códigos e suas Tecnologias (NU_NOTA_LC)
 - Nota em Ciências Humanas e suas Tecnologias (NU_NOTA_CH)
 - Nota da Redação (NU_NOTA_RED)

- **Informações Socioeconômicas:**
 - Renda Familiar (Q006)
 - Escolaridade da Mãe (Q002)
 - Tipo de Escola (Pública ou Privada) (TP_ESCOLA)
 - Raça/Cor (TP_COR_RACA)
- **Localidade:**
 - Estado (SG_UF_RESIDENCIA)
 - Município (CO_MUNICIPIO_RESIDENCIA)

Tarefa a Ser Desenvolvida:

1. Preparação dos Dados:

- Acesse os microdados do ENEM no site do INEP e faça o download dos dados.
- Filtre e organize os dados relevantes, concentrando-se nos atributos mencionados acima.

2. Implementação do Algoritmo k-NN:

- Implemente o algoritmo k-Nearest Neighbors (k-NN) em Python usando a biblioteca scikit-learn ou outra pertinente.
- O algoritmo deve receber como entrada os dados de um novo aluno e prever suas notas nas diferentes áreas do conhecimento com base nos dados históricos dos alunos anteriores.

3. Desenvolvimento da Interface de Entrada de Dados:

- Crie uma interface simples onde o usuário possa inserir os dados de um novo aluno (como notas anteriores em simulados, perfil socioeconômico, etc.).
- A interface deve enviar esses dados para o algoritmo k-NN para realizar a previsão das notas do aluno no ENEM.

4. Análise e Comparação:

- Após a entrada dos dados do novo aluno, o sistema deve comparar com os registros existentes e fornecer uma previsão das notas em cada uma das áreas do conhecimento.
- O resultado deve ser apresentado ao usuário de forma clara, indicando as notas esperadas e como elas se comparam com os "vizinhos" mais próximos identificados pelo algoritmo.

Solução Esperada:

1. **Script em Python:** entregar um script que implemente o algoritmo k-NN para prever as notas do ENEM. O script deve ser bem documentado e estruturado.
2. **Interface de Entrada de Dados:** A interface deve permitir a inserção de dados de forma intuitiva e ser capaz de processar a entrada para realizar a previsão.
3. **Previsão das Notas:** A solução deve prever as notas do novo aluno com base nas características dos alunos anteriores e apresentar uma justificativa para as previsões com base nos vizinhos mais próximos.
4. **Relatório Explicativo:** O aluno deve incluir um relatório explicativo detalhando o processo de desenvolvimento, incluindo como os dados foram manipulados, como o algoritmo foi implementado, e a justificativa para as previsões realizadas.