

Projeto do Nanodegree – 2025/1

Objetivo geral:

O objetivo desta atividade é avaliar as habilidades que foram aprendidas no nanodegree de Machine Learning & Inteligência Artificial em um caso real de aplicação prática. Sua missão será construir um modelo preditivo da evasão de estudantes em um curso online síncrono.

Contexto:

A PensComp, uma organização educacional, promove anualmente cursos online síncronos e gratuitos com o objetivo de desenvolver habilidades de pensamento computacional em estudantes matriculados entre o 5º e o 9º ano do Ensino Fundamental.

A estrutura do curso compreende aulas online semanais e o acesso a uma plataforma de aprendizagem baseada em Moodle. Por meio desta plataforma, os estudantes visualizam sua trajetória de aprendizagem e são instruídos a completar, semanalmente, tarefas que incluem:

- Resolução de questionários;
- Contribuições em fóruns de discussão específicos da plataforma;
- Submissão de códigos relativos a atividades de programação.

A maior parte das atividades propostas não é formalmente avaliada por nota, mas algumas contribuem para uma pontuação final.

Diante dessas características, observa-se uma variação no nível de engajamento dos estudantes com a plataforma. Postula-se que os alunos mais engajados, isto é, que mais interagem com a plataforma, demonstrem:

- Um aprendizado mais amplo e efetivo;
- Desempenho acadêmico superior, refletido em notas mais altas;
- Redução na taxa de abandono do curso.

Anualmente, a organização enfrenta o problema persistente da evasão escolar, resultando na perda de um número substancial de estudantes a cada ano. Embora a evasão completa seja de difícil prevenção, dadas suas causas multifatoriais e, em parte, externas ao alcance da instituição, há estratégias para reduzi-la.

A capacidade de identificar precocemente os estudantes em risco de abandono permite à organização implementar intervenções direcionadas, visando promover o reengajamento e assegurar a permanência destes alunos no curso.

Por isso, a PensComp decidiu contratar sua equipe de Ciência de Dados para construir um modelo preditivo da evasão dos estudantes com os dados gerados pela plataforma educacional.



Dados:

A equipe da PensComp extraiu um conjunto de dados da plataforma educacional. Os dados se referem à interação dos alunos com a plataforma do curso. Esse conjunto de dados foi dividido em conjuntos de treinamento e teste:

- Treinamento: 400 objetos, 36 atributos
- Teste: 100 objetos, 36 atributos

Uma lista dos atributos (variáveis) consta no arquivo de Excel suplementar. Os atributos foram nomeados com iniciais que indicam o tipo de informação contida:

Inicial	Tipo de informação
ds	Descritiva: contém strings.
ts	Timestamp: contém registros de data.
nr	Número: trata-se de uma contagem.
vl	Valor: trata-se de uma métrica ou escore.

Os dados são reais e foram apenas descaracterizados. Nessa descaracterização, os nomes das cidades reais foram substituídos por outras e uma reamostragem por *bootstrap* foi realizada. Esses procedimentos foram adotados para evitar a identificação da organização promotora do curso e dos alunos participantes.

Projeto:

O projeto consistirá em uma entrega e uma apresentação, ambas no dia **25/06**. O projeto será realizado em grupos de 5 alunos, definidos em aula e registrados no AVA (seção “Projeto do Nano”). Não serão aceitos trabalhos entregues individualmente.

Entrega:

Para este projeto, deverá ser entregue um notebook Jupyter com os códigos e explicações de todos os passos e etapas seguidos no desenvolvimento do projeto. Neste notebook, deverá constar as seções:

- **Introdução:** explica a fundamenta a ideia geral do projeto;
- **Manipulação e exploração dos dados:** explora (estatísticas descritivas e visualizações) e organiza os dados de acordo com as necessidades de análise;
- **Modelagem:** treinamento e avaliação dos modelos de aprendizado de máquina escolhidos. Um dos modelos deverá ser, necessariamente, uma rede neural treinada no Keras;
- **Teste do modelo final:** avaliação do modelo final no conjunto de teste;
- **Explicabilidade do modelo final:** uso de técnicas de XAI para interpretar o funcionamento geral e as previsões do modelo final;
- **Exportação do modelo para produção:** exportação do modelo para produção e documentação.

Todas as etapas deverão ser devidamente explicadas e justificadas. Seu projeto será avaliado em termos de:

- Poder preditivo do modelo (de acordo com a métrica escolhida)
- Explicabilidade do modelo (usando técnicas de XAI)
- Clareza e concisão do código: aderência a diretrizes de estilo, uso de funções e classes etc
- Justificativas dos procedimentos realizados: comentários de código e notas markdown
- Design do projeto e do notebook
- Correção e criatividade das soluções propostas

Apresentação:

Cada grupo terá 15 minutos para apresentação do seu projeto no dia **25/06**. As apresentações serão avaliadas em termos de:

- Clareza
- Criatividade
- Storytelling
- Cumprimento do tempo limite
- Organização na participação de todos os membros do grupo

Nota:

O projeto terá peso 5,0, compondo a nota do bloco 3 do nanodegree.

