Machine Learning – Predict Students Dropout and Academic Succes

Autores:

Luiz Fernando Rinaldi Riato

Matheus Prusch

Samuel Porcer Pregnolatto

Pietro Adrian Ribeiro

Maycon Sanches Basilio de Moura

Base escolhida: Predict Students’ Dropout and Academic Success (UC Irvine ML Repository – ID 697)

# 1. Quantidade de Dados

Instâncias (linhas): ~44.000 estudantes (dependendo da versão carregada no ucimlrepo).

Atributos (features): 36 variáveis (colunas) que descrevem perfil, histórico escolar e contexto socioeconômico.

# 2. Descrição do Dataset

## 2.1 Perfil do aluno

* Marital Status → estado civil
* Gender → gênero
* Age at enrollment → idade na matrícula
* Nationality → nacionalidade
* International → se é estrangeiro

## 2.2 Informações de admissão

* Application mode → modo de candidatura
* Application order → ordem da candidatura
* Course → curso escolhido
* Daytime/evening attendance → se estuda de dia ou à noite
* Previous qualification → qualificação anterior
* Previous qualification (grade) → nota anterior
* Admission grade → nota de admissão

## 2.3 Situação familiar

* Mother's qualification
* Father's qualification
* Mother's occupation
* Father's occupation

## 2.4 Aspectos socioeconômicos

* Scholarship holder → bolsista ou não
* Displaced → se mora fora de casa para estudar
* Educational special needs → necessidades especiais
* Debtor → devedor de mensalidade
* Tuition fees up to date → se está em dia com mensalidades

## 2.5 Desempenho acadêmico (1º e 2º Semestre)

Cada semestre possui:

* Curricular units (credited)
* Curricular units (enrolled)
* Curricular units (evaluations)
* Curricular units (approved)
* Curricular units (grade)
* Curricular units (without evaluations)

## 2.6 Indicadores macroeconômicos (Portugal)

* Unemployment rate → taxa de desemprego
* Inflation rate → taxa de inflação
* GDP → Produto Interno Bruto

# 3. Variável Target

A coluna Target contém a situação final do aluno, com três possíveis classes:

* Dropout → aluno desistiu/abandonou o curso
* Enrolled → aluno ainda está matriculado
* Graduate → aluno concluiu o curso

A classificação será feita em função de todos os atributos/features listados acima.

# 4. Resumo do Estudo

## 4.1 Código: studens\_dropout.py

Treinamento da rede neural MLPClassifier em várias arquiteturas de camadas ocultas e avaliação do desempenho.  
Arquiteturas testadas:  
- 1 camada: 20, 50, 100 neurônios  
- 2 camadas: 20-20, 50-20, 100-20, 100-50  
Divisão treino/teste: 70% treino, 30% teste  
Pré-processamento:  
- Variáveis categóricas → LabelEncoder  
- Features numéricas → StandardScaler

## 4.2 Saídas do Programa (acurácia e matriz de confusão)

Arquitetura: (20,)

Acurácia: 0.7161

Matriz de Confusão:  
[[298 66 63]  
 [ 64 90 84]  
 [ 30 62 571]]

--------------------------------------------------

Arquitetura: (50,)

Acurácia: 0.7003

Matriz de Confusão:  
[[296 70 61]  
 [ 62 95 81]  
 [ 33 65 565]]

--------------------------------------------------

Arquitetura: (100,)

Acurácia: 0.7018

Matriz de Confusão:  
[[307 65 55]  
 [ 58 100 80]  
 [ 37 74 552]]

--------------------------------------------------

Arquitetura: (20,20)

Acurácia: 0.7048

Matriz de Confusão:  
[[300 72 55]  
 [ 63 100 75]  
 [ 36 79 548]]

--------------------------------------------------

Arquitetura: (50,20)

Acurácia: 0.7011

Matriz de Confusão:  
[[314 53 60]  
 [ 70 94 74]  
 [ 48 92 523]]

--------------------------------------------------

Arquitetura: (100,20)

Acurácia: 0.692

Matriz de Confusão:  
[[304 68 55]  
 [ 55 106 77]  
 [ 57 97 509]]

--------------------------------------------------

Arquitetura: (100,50)

Acurácia: 0.7018

Matriz de Confusão:  
[[299 72 56]  
 [ 77 82 79]  
 [ 47 65 551]]

--------------------------------------------------

## 4.3 Tabela de Resultados – Arquitetura vs. Acurácia

|  |  |
| --- | --- |
| Arquitetura | Acurácia |
| (20,) | 0.7161 |
| (50,) | 0.7003 |
| (100,) | 0.7018 |
| (20,20) | 0.7048 |
| (50,20) | 0.7011 |
| (100,20) | 0.6920 |
| (100,50) | 0.7018 |

## 4.4 Código: studens\_dropout\_test.py

Teste com um novo aluno fictício para prever a classe (Dropout, Enrolled, Graduate)  
Saída do programa:  
Classe prevista (código): 0  
Classe prevista (rótulo): Dropout

# 5. Conclusão

A rede neural MLP apresentou acurácia entre 69% e 72%, dependendo da arquitetura.  
Redes com uma camada ou duas camadas ocultas tiveram desempenho semelhante.  
O modelo consegue classificar novos alunos em Dropout, Enrolled ou Graduate, com base nos 36 atributos.  
Esse estudo demonstra a aplicabilidade de redes neurais para previsão de sucesso acadêmico usando dados do mundo real.