

Bárbara Melo

Trabalho final de consultoria de Ciências de Dados

PUCSP

Mayo

introdução

O engajamento inicial dos estudantes é um fator importante para compreender seu desempenho ao longo de um curso.

A base de dados analisada reúne informações anonimizadas sobre a participação dos alunos em diferentes atividades das primeiras aulas, como pré-class, presença, exercícios e contribuições , além do resultado final na disciplina.

objetivo

Este trabalho tem como objetivo investigar se esses padrões de engajamento nas aulas iniciais estão associados ao desfecho final do aluno, contribuindo para identificar comportamentos que podem indicar maior ou menor probabilidade de sucesso acadêmico.



Passos

1. Importação de bibliotecas

Carrega ferramentas para manipulação de dados, estatística e visualização, além do prince, usado para Análise de Correspondência (ANACOR).

2. Leitura da base com duas linhas de cabeçalho

A planilha possui dois níveis de cabeçalho. Em seguida, esses dois níveis são combinados em um único nome por coluna.

3. Identificação das colunas de cada aula

A função cols_for_aula() usa uma expressão regular para encontrar todas as colunas que começam com "Aula X_".

Isso permite selecionar automaticamente todas as informações referentes a cada aula.

4. Função de detecção de engajamento

A função aula_engajada_values() avalia se o aluno mostrou engajamento em determinada aula.

Ela busca marcadores positivos nas células
"Se encontrar qualquer marcador → engajado (1)
Se não encontrar → não engajado (0)

Passos

5. Cálculo do engajamento nas 5 primeiras aulas

Para cada uma das primeiras 5 aulas, o código:

- seleciona as colunas correspondentes,
- aplica a função de engajamento,
- cria uma variável binária eng_aula_i.

8. Padronização do desfecho final

Padroniza valores como:

- “aprovado”, “aprov.” → Aprovado
- “reprovado”, “reprov.” → Reprovado

E exclui casos que não se enquadram nessas duas categorias.

6. Criação de indicadores agregados

Duas novas variáveis são criadas:

- a)** Sequência binária de engajamento
eng_sequence_5 = '10110'

Mostra o padrão do aluno nas 5 aulas.

- b)** Contagem total de engajamentos
eng_count_5 = 3

Número total de aulas em que houve engajamento.

7. Classificação em níveis (Low, Medium, High)

Uma função categoriza o engajamento total:

- 0–1 → Low
- 2–3 → Medium
- 4–5 → High

Cria a coluna:

eng_level_5

Passos

9. Geração das tabelas cruzadas (crosstabs)

Cria duas tabelas:

1. Sequência de engajamento vs. Situação Final
tabela_seq

2. Nível agregado (Low/Medium/High) vs.
Situação Final
tabela_level

Essas tabelas mostram como engajamento e desfecho estão associados.

10. Execução da ANACOR (Análise de Correspondência)

A função run_anacor_and_plot():

- Verifica se a tabela é válida
- Executa a ANACOR usando prince
- Calcula teste qui-quadrado (χ^2)
 - valor-p indica se há associação estatística entre engajamento e desfecho
- Gera o mapa perceptual com:
 - categorias de engajamento (linhas)
 - categorias do desfecho (colunas)
- Plota gráfico bidimensional com Dimensão 1 e Dimensão 2
- Retorna coordenadas e métricas do modelo

A ANACOR é executada duas vezes:

- Por sequência binária (Sequence 5)
- Por nível agregado (Level 0–5)

Isso permite avaliar a relação entre engajamento e resultado de forma detalhada ou agrupada.

Resumo

- O código transforma o engajamento nas primeiras 5 aulas em variáveis binárias.
- Cria indicadores (sequência, contagem e nível).
- Padroniza o desfecho final (Aprovado/Reprovado).
- Produz tabelas de associação.
- Realiza ANACOR para examinar graficamente e estatisticamente a relação entre engajamento inicial e o resultado final dos alunos.



Os padrões de engajamento inicial estão associados ao desfecho final do aluno no curso?

- $p < 10^{-45}$

O engajamento inicial NÃO é aleatório em relação à aprovação.

- Alunos com High (4–5) engajamentos:
 - 1316 aprovados
 - 93 reprovados
- Alunos com Medium (2–3):
 - 27 reprovados
 - 10 aprovados

Quanto maior o engajamento no início, maior a taxa de aprovação.

As sequências binárias (ex.: 11111, 10111, 00011) também revelaram padrões:

- Sequências com muitos 1s se agrupam com Aprovado.
- Sequências com muitos 0s se agrupam com Reprovado.

Nos mapas da ANACOR:

- High (4–5) fica próximo de Aprovado.
- Medium (2–3) e sequências com zeros ficam próximos de Reprovado.

Visualmente, os grupos estão bem separados, o engajamento inicial distingue claramente quem tende a passar ou reprovar.



```

Aula 1 cols: ['Aula 1_Pre-Class', 'Aula 1_P', 'Aula 1_Hw', 'Aula 1_CP', 'Aula 1_Bh']
[0 1]
Aula 2 cols: ['Aula 2_Pre-Class', 'Aula 2_P', 'Aula 2_Hw', 'Aula 2_CP', 'Aula 2_Bh']
[1 0]
Aula 3 cols: ['Aula 3_Pre-Class', 'Aula 3_P', 'Aula 3_Hw', 'Aula 3_CP', 'Aula 3_Bh']
[1 0]
Aula 4 cols: ['Aula 4_Pre-Class', 'Aula 4_P', 'Aula 4_Hw', 'Aula 4_CP', 'Aula 4_Bh']
[1 0]
Aula 5 cols: ['Aula 5_Pre-Class', 'Aula 5_P', 'Aula 5_Hw', 'Aula 5_CP', 'Aula 5_Bh']
[1 0]
Número de linhas após filtro de desfecho: 1446

```

Tabela (Sequência 5 aulas x Situação Final):

Aula 14_Situação Final	Aprovado	Reprovado
eng_sequence_5		
00011	1	3
00111	6	2
01001	0	3
01011	1	3
01100	0	1
01101	1	1
01110	0	3
01111	192	24
10001	0	1
10010	0	1
10011	0	3
10100	0	1
10101	0	1
10110	0	2
10111	9	4
11001	1	1
11010	0	1
11011	12	4
11101	9	4
11110	8	2
11111	1086	55

Tabela (Nível agregado 0-5 x Situação Final):

Aula 14_Situação Final	Aprovado	Reprovado
eng_level_5		

Tabela (Nível agregado 0-5 x Situação Final):

Aula 14_Situação Final Aprovado Reprovado

eng_level_5

High (4-5) 1316 93

Medium (2-3) 10 27

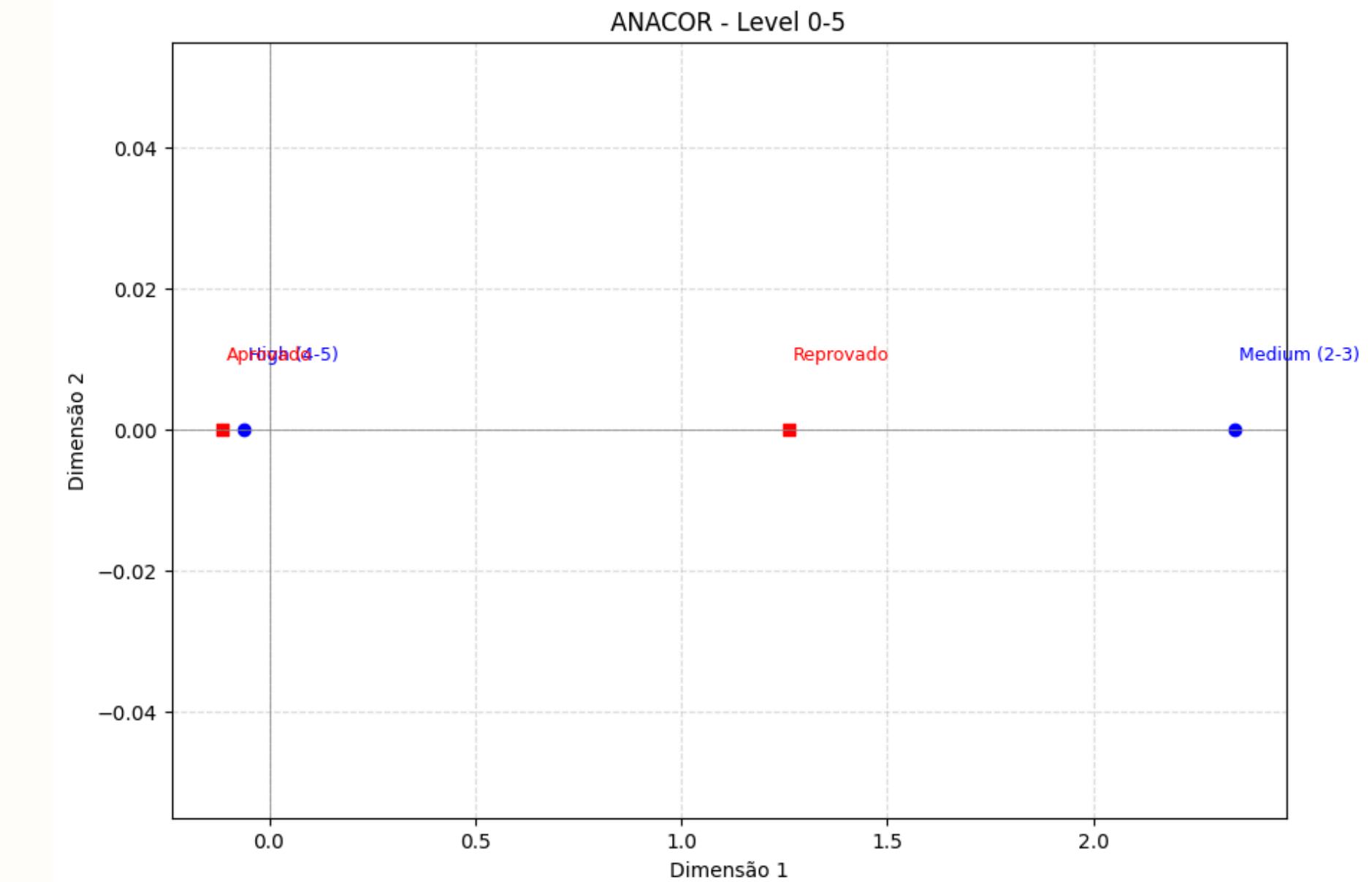
--- ANACOR: Sequence 5 ---

Chi2: 291.9934506566267 p-value: 3.47680789070553e-50 dof: 20

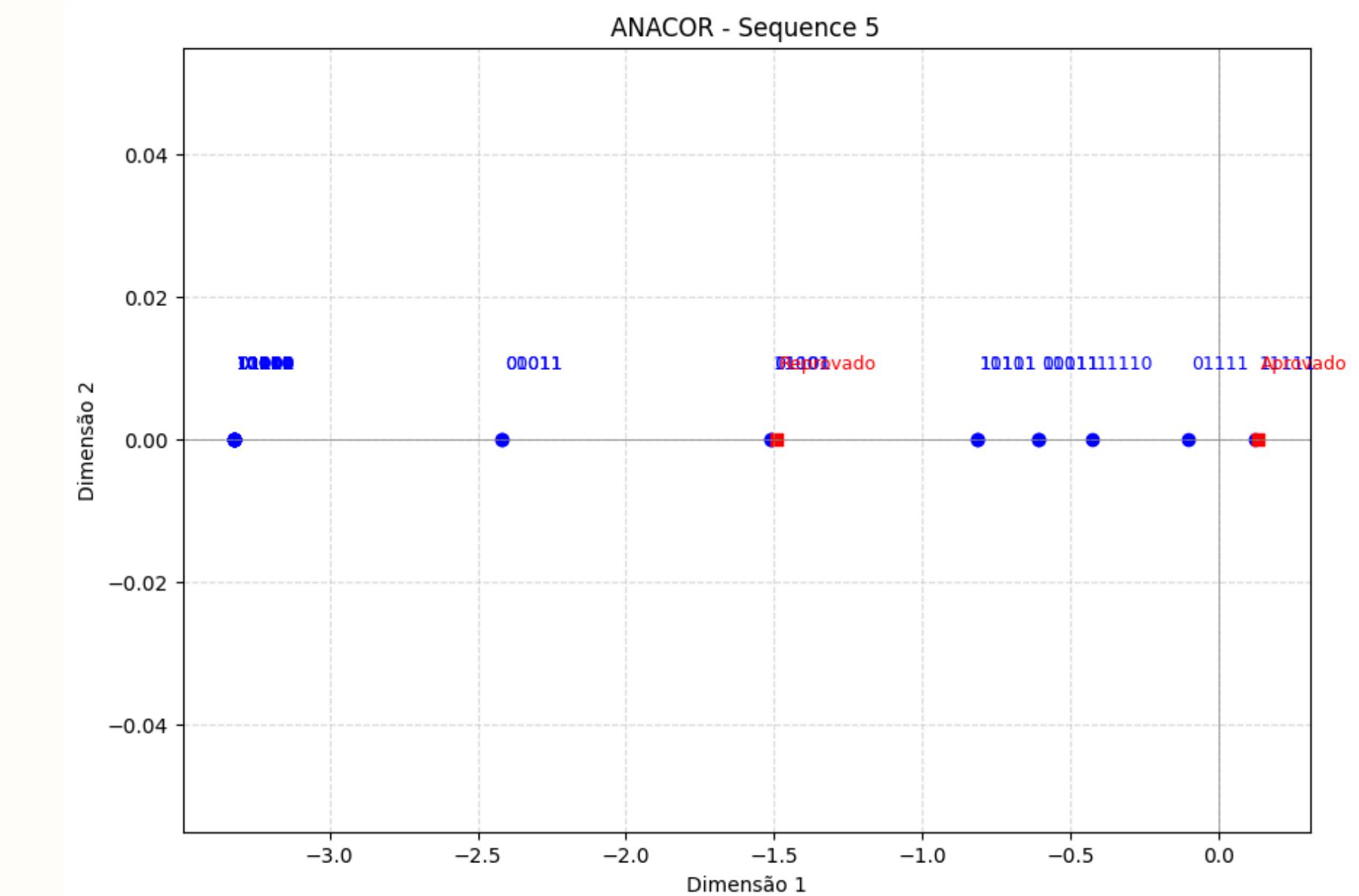
--- ANACOR: Level 0-5 ---

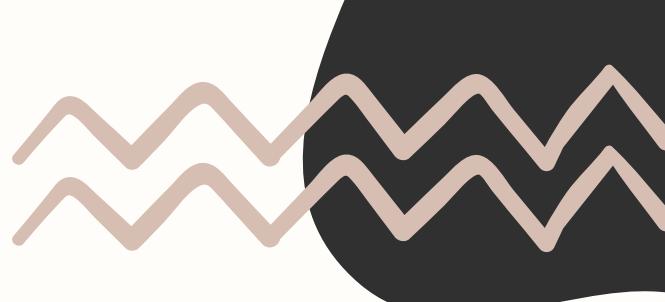
Chi2: 200.07460648855567 p-value: 2.0116439104119896e-45 dof: 1

Resultados



Resultados





Conclusão

Quem engaja forte nas primeiras aulas → aprova em massa.

Quem mostra engajamento instável ou baixo cedo → tende
fortemente a reprovar.

Os gráficos ANACOR evidenciam isso visualmente: os grupos
estão completamente separados.

*thank you for
your attention!*