



UNIVERSIDADE FEDERAL DE SANTA MARIA

CENTRO DE TECNOLOGIA

Peso: 5

Disciplina: Mineração de Dados (ELC1098)

Professor: Dr. Joaquim Assunção

## Trabalho prático 2

Grupos de até 4 componentes.

### Descrição

Você terá acesso a dados, parcialmente anonimizados, da UFSM. Estes possuem registros de disciplinas do CT, juntamente com alguns dados relacionados. Você também terá *datasets* com estatísticas de ingressantes X formandos por centro. Sua tarefa é definir um plano para analisar e minerar os dados visando descobrir padrões ou *insights* nos dados usando técnicas mostradas na disciplina (pelo menos 2 tipos).

Há dois tipos de arquivos disponíveis:

1. **Dados das disciplinas**, cujo nome do arquivo é composto por um código e o ano (ex. D200888\_2021.xlsx). Estes são compostos pelos seguintes campos:

```
a. $ Ano : num [1:45] 2021 2021 2021 2021 2021 ...
b. $ Semestre : chr [1:45] "1. Semestre" "1. Semestre" "1. Semestre" "1. Semes
tre" ...
c. $ Cód. Disciplina : chr [1:45] "D200888" "D200888" "D200888" "D200888" ...
d. $ Cód. Turma : chr [1:45] "10" "10" "10" "11" ...
e. $ Situação : chr [1:45] "Aprovado" "Dispensado" "Reprovado" "Aprovado" ...
f. $ % : chr [1:45] "15" "3" "82" "31" ...
g. $ Alunos : num [1:45] 5 1 27 17 37 1 20 17 3 1 ...
h. $ Professor : chr [1:45] "P888402148" "P888402148" "P888402148" "P148855320"
...
i. $ Cód. Curso : chr [1:45] "139" "139" "139" "521" ...
j. $ Curso : chr [1:45]
```

2. **Ingressantes e formandos por sexo** (diversos centros).

```
a. $ COD_CURSO : chr [1:607] "640.EAD.08" "640.EAD.08" "640.EAD.08" "640.EAD.08" ...
b. $ NOME_UNIDADE : chr [1:607] "Curso de Formação de Professores para Educação Profissional/Ead/Agudo/RS" "Curso d
e Formação de Professores para Educação Profissional/Ead/Agudo/RS" "Curso de Formação de Professores para Educaçã
o Profissional/Ead/Agudo/RS" "Curso de Formação de Professores para Educação Profissional/Ead/Agudo/RS" ...
c. $ NIVEL_CURSO : chr [1:607] "Graduação" "Graduação" "Graduação" "Graduação" ...
d. $ ANO : chr [1:607] "2018" "2018" "2019" "2019" ...
e. $ SEXO : chr [1:607] "M" "F" "M" "F" ...
f. $ INGRESSANTES : num [1:607] 14 24 0 0 14 24 14 15 0 0 ...
g. $ FORMADOS : num [1:607] 0 0 7 15 7 15 0 0 6 9 ...
```

### Algumas dicas:

- Explore os dados, obtendo estatísticas básicas e mostrando características visualmente (plot).
- Crie um plano baseado em alguma hipótese. Altere os dados (transformação) para estar no formato adequado.
- Crie um relatório descrevendo o processo de modo claro e convincente.

### Prazo

O trabalho deve ser entregue até dia 2 de dezembro (11:59) e apresentado nos dias 3 ou 5 de dezembro (10 a 20 minutos).

### Entrega e Apresentação

Deve ser entregue:

1. Um arquivo PDF (ou HTML) com toda a descrição do trabalho (mínimo 4 páginas),
2. Um PDF (ou HTML) com a apresentação.
3. O script usado na análise e mineração dos dados.

O trabalho deve ser entregue por **e-mail** até dia 2/12, às 11:59 (assunto: “[DM] alunoA, alunoB, alunoC...”).

A avaliação será dada pelos itens entregues:

1. A qualidade do volume de descrição e análise do processo (4 pontos)
2. A apresentação (2.5 pontos)
3. O script (1 ponto)
4. Uso de técnicas, análises e outras descobertas (2.5).

### Alternativa (Individual, máximo 9.5 pontos)

Leia um dos artigos selecionados, cujo tema envolve *data mining* ou *machine learning*, e faça um resumo do mesmo.

### Regras

- Você deverá apresentar seu artigo em, no máximo, 10 minutos.
- Seu resumo deve ser claro o suficiente para que seus colegas entendam o artigo.
- Caso não entenda algum conceito, você deve ler as referidas referências bibliográficas.
- Os artigos devem estar na lista de artigos selecionados.
- A entrega deve conter a apresentação e os resumos (apresentação será dia 3).