

# Trabalho Prático do Módulo 2

**Entrega** 27 de ago de 2021 em 23:59**Pontos** 25**Perguntas** 13**Disponível** até 27 de ago de 2021 em 23:59**Limite de tempo** Nenhum**Tentativas permitidas** 2

## Instruções

O Trabalho Prático do Módulo 2 está disponível!

### 1. Instruções para realizar o trabalho prático

Consulte a data de entrega no teste e em seu calendário.

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o "Fórum de dúvidas do Trabalho Prático do Módulo 2".

Para iniciá-lo clique em "Fazer teste". Você tem somente **uma** tentativa e não há limite de tempo definido para realizá-lo. Caso precise interromper a atividade, apenas deixe a página e, ao retornar, clique em "Retomar teste".

Clique em "Enviar teste" **somente** quando você concluí-lo. Antes de enviar confira todas as questões.

Caso o teste seja iniciado e não enviado até o final do prazo de entrega, a plataforma enviará a tentativa não finalizada automaticamente, independente do progresso no teste. Fique atento ao seu teste e ao prazo final, pois novas tentativas só serão concedidas em casos de questões médicas.

O gabarito será disponibilizado partir de sexta-feira, **27/08/2021**, às 23h59.

Bons estudos!

### 2. O arquivo abaixo contém o enunciado do trabalho prático

[Enunciado do Trabalho Prático – Módulo 2 – Bootcamp Engenheiro\(a\) de Dados.pdf](#)

Este teste foi indisponível 27 de ago de 2021 em 23:59.

## Histórico de tentativas

	Tentativa	Tempo	Pontuação
MANTIDO	<a href="#">Tentativa 2</a>	4 minutos	23,08 de 25
MAIS RECENTE	<a href="#">Tentativa 2</a>	4 minutos	23,08 de 25
	<a href="#">Tentativa 1</a>	604 minutos	21,16 de 25

Pontuação desta tentativa: **23,08** de 25

Enviado 27 de ago de 2021 em 14:39

Esta tentativa levou 4 minutos.

Correto!

### Pergunta 1

1,92 / 1,92 pts

A Coleta de dados é uma etapa do processo de Engenharia de Dados. Sobre essa etapa está CORRETO afirmar, que:

- ☐ Deve-se coletar dados independente do objeto a ser analisado.
- ☒ A principal fonte de dados é a externa.
- ☐ Não é possível coletar dados de redes sociais.
- ☐ Os dados coletados são sempre internos.

## Pergunta 2

1,92 / 1,92 pts

Análise as seguintes asserções:

I) No text mining a Coleta de Dados é uma etapa que independe do problema,

PORQUE

II) O text mining pode ser definido como a extração não trivial de informações, não explícitas, de grandes bases textuais, previamente desconhecidas, e potencialmente úteis.

☐ As assertivas I e II são verdadeiras, sendo que a II justifica a I.

☒ A assertiva I é falsa, mas a II é verdadeira.

☐ Ambas assertivas são falsas.

☐ As assertivas I e II são verdadeiras, mas a II não justifica a I.

Correto!

## Pergunta 3

0 / 1,92 pts

A imagem a seguir exibe um trecho do arquivo robots.txt do Instagram:

```
User-agent: Applebot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: baiduspider
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: Bingbot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: DuckDuckBot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: Googlebot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/
```

Análise as seguintes afirmativas sobre esse arquivo:

I) O arquivo robots.txt exibe as permissões de raspagem do site.

II) A pasta /api/ do Instragam está disponível para Scraping para o Googlebot.

III) O agente Applebot não pode acessar a pasta /query/.

Estão corretas as afirmativas:

☒ Apenas II e III.

Você respondeu

☐ Todas estão corretas.☐ Apenas I.

Resposta correta

☐ Apenas I e III.**Pergunta 4**

1,92 / 1,92 pts

A imagem a seguir exibe um trecho do arquivo robots.txt do Instagram:

```
User-agent: Applebot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: baiduspider
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: Bingbot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: DuckDuckBot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/

User-agent: Googlebot
Disallow: /api/
Disallow: /graphql/query/
Disallow: /publicapi/
Disallow: /query/
```

O analista de dados, usando a Linguagem R, deseja fazer scrap no Instagram. Sobre o Scraping usando a Linguagem R, está INCORRETO afirmar que:

☐ Um dos pacotes R para realização de Scraping é o rvest.☒ O comando paths\_allowed do agente Googlebot na pasta /api/ retorna TRUE.☐ O comando read\_html() faz a leitura da webpage para o R.☐ Antes de fazer o scraping é necessário verificar a permissão através do comando paths\_allowed.

Correto!

**Pergunta 5**

1,92 / 1,92 pts

Considere as seguintes afirmativas sobre o Text Mining:

I. A mineração de textos utiliza técnicas diferentes da mineração de dados.

II. Os textos representam um tipo específico de dado.

III. Pode ser definido como a extração não trivial de informações, não explícitas, de grandes bases textuais, previamente desconhecidas e potencialmente úteis.

Estão corretas as afirmativas:

Correto!

- ☐ Apenas I e II.
- ☐ Apenas I e III.
- ☐ Apenas II e III.
- ☒ Todas estão corretas.

**Pergunta 6**

1,92 / 1,92 pts

Análise as seguintes asserções:

I) O Text Mining é um processo completamente diferente do Data Mining

PORQUE

II) Os dados utilizados no Text Mining são textos.

Correto!

- ☐ As assertivas I e II são verdadeiras, mas a II não justifica a I
- ☐ As assertivas são falsas.
- ☒ A assertiva I é falsa, mas a II é verdadeira.
- ☐ As assertivas I e II são verdadeiras, sendo que a II justifica a I

**Pergunta 7**

1,92 / 1,92 pts

A tokenização, análise léxica, do texto abaixo, produz quantos tokens?

"Considere o seguinte texto: esse texto refere-se ao enunciado da questão 5".

Correto!

- ☒ 13..
- ☐ 11.
- ☐ 14.
- ☐ 12.

**Pergunta 8**

1,92 / 1,92 pts

Associe a coluna da esquerda com a coluna da direita

- |                 |                             |
|-----------------|-----------------------------|
| A) Steming.     | ( ) Análise Léxica.         |
| B) Tokenização. | ( ) Análise Morfológica.    |
| C) Stopwords.   | ( ) Compõem o Bag of Words. |

A CORRETA associação entre as colunas é representada pela seguinte sequência:

☐ C, B, A.

☒ B, A, C.

☐ A, B, C.

☐ B, C, A.

Correto!

### Pergunta 9

1,92 / 1,92 pts

Análise as seguintes asserções:

I) A stemização pode reduzir o número de termos distintos

PORQUE

II) Esse processo consiste na redução das palavras ao seu radical.

☒ As assertivas I e II são verdadeiras, sendo que a II justifica a I.

☐ A assertiva I é falsa, mas a II é verdadeira.

☐ As assertivas I e II são verdadeiras, mas a II não justifica a I.

☐ As assertivas são falsas.

Correto!

### Pergunta 10

1,92 / 1,92 pts

Considere as seguintes afirmativas sobre as Stopwords:

I) São Termos Irregulares que normalmente possuem alta frequência nos textos.

II) Normalmente são constituídas de verbos auxiliares, artigos, preposições.

III) Agregam valor na análise.

Estão CORRETAS as afirmativas:

☐ Todas estão corretas.

☒ Apenas I e II.

☐ Apenas II e III.

☐ Apenas I e III.

Correto!

### Pergunta 11

1,92 / 1,92 pts

Analise as seguintes asserções:

I) As stopwords não devem ser descartadas

PORQUE

II) Agregam valor ao processo de Text Mining.

- ☐ As assertivas I e II são verdadeiras, sendo que a II justifica a I.
- ☐ As assertivas I e II são verdadeiras, mas a II não justifica a I.
- ☒ As assertivas são falsas.
- ☐ A assertiva I é falsa, mas a II é verdadeira.

Correto!

### Pergunta 12

1,92 / 1,92 pts

A imagem a seguir exibe um pequeno trecho de código, após a execução desse código serão removidas quantas stopwords?

```
texto_token = paste0("Esse texto após a remoção das stopwords conterá as seguinte palavras")
corpus <- Corpus(VectorSource(texto_token))
corpus <- tm_map(corpus, removewords, stopwords('pt'))
tdm <- as.matrix(TermDocumentMatrix(corpus, control=list(stemming = TRUE)))
```

- ☐ 2.
- ☐ 4.
- ☐ 5.
- ☒ 3.

Correto!

### Pergunta 13

1,96 / 1,96 pts

Sobre o texto “recuperação recuperam alunos que precisam ser recuperados”, está CORRETO afirmar que:

- ☐ A normalização morfológica não impacta no número de termos.
- ☐ “Ser” não faz parte da lista de stopwords.
- ☐ A análise léxica resulta em cinco termos.
- ☒ Recuperação, recuperam e recuperados possuem o mesmo radical “recuper”.

Correto!

Pontuação do teste: **23,08** de 25