

Aula 05

*Banco do Brasil (Escriturário - Agente de
Tecnologia) Passo Estratégico de
Tecnologia de Informação - 2023
(Pós-Edital)*

Autor:
Thiago Rodrigues Cavalcanti

03 de Fevereiro de 2023

Simulado 2

Questões Comentadas 9



SIMULADO

Vamos começar nosso segundo simulado. Lembrando que os simulados do curso do Passo Estratégico são direcionados pela análise dos últimos concursos aplicados pela banca, que no nosso caso é a Fundação Cesgranrio, e servem de treinamento para nosso concurso do **Banco do Brasil**. Esse simulado terá como base o assunto das aulas 03 e 04. Nossa objetivo é ajudar a você na fixação do assunto. Espero que goste! Qualquer dúvida estou às ordens! Forte abraço!



1.

Não é considerado um caso de uso de modelagem de tópicos:

- a) Descrever sistemas interativos que permitem aos usuários interagir em linguagem natural.
- b) Usar um documento como um vetor de recursos para classificação de texto.
- c) Projetar um sistema de recomendação de texto
- d) Resumir documentos ou tweets na forma de palavras-chave com base nas distribuições de tópicos aprendidos
- e) Detectar tendências de mídia social ao longo de um período

2.

O estudo, usando distribuições estatísticas, de elementos da linguagem para caracterizar semelhanças entre documentos (por exemplo, e-mail), atos de fala (por exemplo, sentenças faladas ou escritas) ou seus elementos (por exemplo, frases, palavras) é denominado:

- a) word embeddings
- b) tokenização
- c) semântica vetorial
- d) Hipótese distributiva
- e) semântica distributiva



3.

Alguns dos algoritmos de modelagem de tópicos são mais populares. Observer a lista a seguir:

1. Latent dirichlet allocation (LDA),
2. Latent semantic analysis (LSA)
3. Nonnegative matrix factorization (NMF)
4. Probabilistic latent semantic analysis (PLSA),
5. Support Vector Machine (SVM)

Qual dos algoritmos acima são utilizados na modelagem de tópicos de forma ampla.

- a) 1, 2 e 3 apenas
- b) 1, 2, 3 e 4 apenas
- c) 2, 3, 4 e 5 apenas
- d) 1, 3 e 5 apenas
- e) 1, 2 e 4 apenas

4.

Sobre PLN, assinale a alternativa correta:

- a) os tokens são convertidos em números antes de serem entregues a qualquer rede neural
- b) o processo de conversão de uma frase ou parágrafo em tokens é referido como Stemming
- c) o processo de remoção de palavras como “a”, “e”, “até”, “uma”, “o” de uma frase é chamado de Lemmatização.
- d) TF-IDF ajuda a estabelecer uma palavra que ocorre com mais frequência no documento.
- e) as palavras representadas como vetores são chamadas de data frames de palavras neurais

5.

Qual das seguintes não é uma técnica de pré-processamento em PLN:

- a) Derivação e lematização
- b) Conversão para minúsculas
- c) Remoção de pontuações



- d) Remoção de stopwords
- e) Análise de sentimento

6.

No processo de tratamento de dados do PLN temos uma ação que vai combinar tokens que desempenham um papel semelhante que funcionam juntos da mesma maneira, essa ação é denominada:

- a) Chinking
- b) Chunking
- c) Lemmatization
- d) Stemming
- e) Tokenização

7.

Sobre o acesso aos dados em uma lista usando as linguagens Python e R.

Python:

```
lista_inteiros = [10, 20, 30, 40, 50, 60]
print(lista_inteiros[2:5])
```

R:

```
lista_inteiros <- c(10, 20, 30, 40, 50, 60)
lista_inteiros[2:5]
```

Os códigos acima vão retornar os seguintes valores, respectivamente

- a) [20, 30, 40] e [1] 20 30 40 50
- b) [20, 30, 40, 50] e [1] 30 40 50 60
- c) [30, 40, 50, 60] e [1] 30 40 50
- d) [30, 40, 50] e [1] 20 30 40 50
- e) [30, 40, 50] e [1] 20 30 40

8.

Considere o código Python, versão, na qual o comando print não requer parênteses.



A execução desse código:

```
def teste(n):
    for k in range(1, n+1):
        yield k
for x in teste(10):
    print(x)
```

- a) não tem efeito, pois nenhum comando print é acionado;
- b) provoca a exibição do número 10 na saída;
- c) provoca a exibição dos números de 1 até 10 na saída;
- d) provoca um erro de compilação.
- e) provoca um erro de execução;

9.

Na linguagem de programação Python, existem 3 estruturas para armazenar dados indexados. A estrutura cujos valores são imutáveis depois de sua criação é conhecida como

- a) lista.
- b) operador.
- c) tupla.
- d) classe.
- e) dicionário.

10.

Um analista implementou o seguinte programa em Python:

```
01 a = [1, 2, 3];
02 b = list(a);
03 print(a);
04 print(b);
05 print(a==b);
06 print(a is b);
```



Assinale a alternativa que apresenta corretamente o resultado do programa quando executado.

A)

A resposta do programa é:

[1, 2, 3]

[]

False

False

B)

A resposta do programa é:

[1, 2, 3]

[1, 2, 3]

True

True

C)

A resposta do programa é:

[1, 2, 3]

[1, 2, 3]

True

False

D)

A resposta do programa é:

[1, 2, 3]

[1, 2, 3]

True

Null

E) O programa não executa por um erro na linha 02, em que é atribuído um tipo de dado diferente daquele definido na variável b.



11.

Um analista do MJSP armazenou nomes e ID no seguinte programa Python:

```
01 dados = {  
02 'Machado de Assis': 7575,  
03 'Santos Dumont': 7475,  
04 'Rui Barbosa': 1675,  
05 };
```

Agora, o analista necessita apresentar em tela o ID 7475 referente ao nome informado. Assinale a alternativa que apresenta o código correto para imprimir esse ID na tela para o usuário.

- A) print(7475);
- B) print(dados[nome], ID==7475);
- C) print(ID, 'Santos Dumont');
- D) print(dados == 'Santos Dumont');
- E) print(dados['Santos Dumont']);

12.

Analise o seguinte código Python, escrito por um cientista de dados, e assinale a alternativa correta.

```
In [01]: nome = input("Qual  
o seu nome? ")  
In [02]: print(nome)
```

- A) O código não irá funcionar pois está faltando ';' ao final de cada linha.
- B) O uso da função input está incorreto, pois ela tem que ser chamada em uma nova linha.
- C) O resultado do comando print é apresentar na tela o nome informado pelo usuário.
- D) Falta a estrutura correta de criação de uma classe para a execução do comando.
- E) Falta definir o tipo da variável nome, por isso o programa não irá executar.

13.



Analise o seguinte código Python, escrito por um cientista de dados, e assinale a alternativa correta.

In [01]: num = 88

In [02]: print ('par' if
num%2 == 0 else 'ímpar')

- A) O resultado da execução de código é a informação ‘ímpar’ na tela.
- B) O código apresenta como resultado “1” na linha [02], pois num%2 == 0 é uma atribuição e não uma comparação.
- C) O código executado apresenta na tela o resultado ‘0’.
- D) O resultado da execução de código é a informação ‘par’ na tela.
- E) O código gera um erro, pois na linha [02] o comando if...else não está correto.



QUESTÕES COMENTADAS



1.

Não é considerado um caso de uso de modelagem de tópicos:

- a) Descrever sistemas interativos que permitem aos usuários interagir em linguagem natural.
- b) Usar um documento como um vetor de recursos para classificação de texto.
- c) Projetar um sistema de recomendação de texto
- d) Resumir documentos ou tweets na forma de palavras-chave com base nas distribuições de tópicos aprendidos
- e) Detectar tendências de mídia social ao longo de um período

Comentário: A alternativa A apresenta a definição do conceito de chatbots que não tem uma relação com os casos de uso em que a modelagem de tópicos se aplica.

Gabarito: alternativa A

2.

O estudo, usando distribuições estatísticas, de elementos da linguagem para caracterizar semelhanças entre documentos (por exemplo, e-mail), atos de fala (por exemplo, sentenças faladas ou escritas) ou seus elementos (por exemplo, frases, palavras) é denominado:

- a) word embeddings
- b) tokenização
- c) semântica vetorial
- d) Hipótese distributiva
- e) semântica distributiva



Comentário: O conceito descrito na questão é o de semântica distributiva. Outros conceitos próximos que aprendemos na nossa aula são:

Hipótese distributiva - Palavras que ocorrem em contextos semelhantes têm significados semelhantes. Por exemplo, as palavras em inglês “cachorro” e “gato” ocorrem em contextos semelhantes. Assim, de acordo com a hipótese distributiva, deve haver uma forte semelhança entre os significados dessas duas palavras.

Representação distribucional - Refere-se a esquemas de representação obtidos com base na distribuição de palavras a partir do contexto em que as palavras aparecem. Esses esquemas são baseados em hipóteses distributivas.

Representação distribuída - Esquemas de representação distribuída comprimem significativamente a dimensionalidade. Isso resulta em vetores que são compactos (ou seja, de baixa dimensão) e densos (ou seja, quase nenhum zero). O espaço vetorial resultante é conhecido como *representação distribuída*.

Semântica vetorial - conjunto de métodos de PLN que objetivam aprender as representações de palavras com base em propriedades de distribuição de palavras em um grande corpus.

Logo, temos a nossa resposta na alternativa E.

Gabarito: alternativa E

3.

Alguns dos algoritmos de modelagem de tópicos são mais populares. Observer a lista a seguir:

6. Latent dirichlet allocation (LDA),
7. Latent semantic analysis (LSA)
8. Nonnegative matrix factorization (NMF)
9. Probabilistic latent semantic analysis (PLSA),
10. Support Vector Machine (SVM)

Qual dos algoritmos acima são utilizados na modelagem de tópicos de forma ampla.

- a) 1, 2 e 3 apenas
- b) 1, 2, 3 e 4 apenas
- c) 2, 3, 4 e 5 apenas
- d) 1, 3 e 5 apenas
- e) 1, 2 e 4 apenas



Comentário: A modelagem de tópicos serve para extrair tópicos de uma coleção de documentos. É o método de mineração de texto amplamente utilizado no Processamento de Linguagem Natural para obter insights sobre os documentos em texto. Este processo altamente importante pode ser realizado por vários algoritmos ou métodos. Alguns deles são:

- **Latent Dirichlet Allocation (LDA)**
- **Non Negative Matrix Factorization (NMF)**
- **Latent Semantic Analysis (LSA)**
- **Parallel Latent Dirichlet Allocation (PLDA)**
- Pachinko Allocation Model (PAM)

Assim, temos a nossa resposta na alternativa B.

Gabarito: alternativa B

4.

Sobre PLN, assinale a alternativa correta:

- a) os tokens são convertidos em números antes de serem entregues a qualquer rede neural
- b) o processo de conversão de uma frase ou parágrafo em tokens é referido como Stemming
- c) o processo de remoção de palavras como “a”, “e”, “até”, “uma”, “o” de uma frase é chamado de Lemmatização.
- d) TF-IDF ajuda a estabelecer uma palavra que ocorre com mais frequência no documento.
- e) as palavras representadas como vetores são chamadas de data frames de palavras neurais

Comentário: Vamos analisar cada uma das alternativas:

- a) (Certo) Todas as palavras são convertidas em um número antes de alimentar uma rede neural.
- b) (Errado) A declaração descreve o processo de tokenização
- c) (Errado) O nome do processo de remoção de palavras irrelevantes é chamado de remoção de stop words.
- d) (Errado) TF-IDF ajuda a estabelecer a importância de uma palavra específica no contexto do corpus do documento. O TF-IDF leva em consideração a quantidade de vezes que a palavra



aparece no documento e balanceada pela quantidade de documentos em que a palavra aparece no corpus.

e) (Errado) As palavras representadas como vetores são chamadas de **embeddings de palavras neurais**.

Gabarito: alternativa A

5.

Qual das seguintes não é uma técnica de pré-processamento em PLN:

- a) Derivação e lematização
- b) Conversão para minúsculas
- c) Remoção de pontuações
- d) Remoção de stopwords
- e) Análise de sentimento

Comentário: A análise de sentimento não é uma técnica de pré-processamento. Esse tipo de análise é feito após o pré-processamento e é um caso de uso de PNL. Todos as outras opções listadas são usadas como parte do pré-processamento de instrução.

Gabarito: alternativa E.

6.

No processo de tratamento de dados do PLN temos uma ação que vai combinar tokens que desempenham um papel semelhante que funcionam juntos da mesma maneira, essa ação é denominada:

- a) Chinking
- b) Chunking
- c) Lemmatization
- d) Stemming
- e) Tokenização

Comentário: a segmentação ou chunking que é **uma maneira de combinar tokens que desempenham um papel semelhante que funcionam juntos da mesma maneira**. O que é



semelhante em conceito também à ideia de n-grams. Portanto, o objetivo do chunking é identificar frases em nosso texto. "Um cordeirinho" e "planeta Terra", estes correspondem aos n-gramas que estávamos tentando identificar. **Chunking depende de partes da marcação de fala.**

Gabarito: alternativa B.

7.

Sobre o acesso aos dados em uma lista usando as linguagens Python e R.

Python:

```
lista_inteiros = [10, 20, 30, 40, 50, 60]
print(lista_inteiros[2:5])
```

R:

```
lista_inteiros <- c(10, 20, 30, 40, 50, 60)
lista_inteiros[2:5]
```

Os códigos acima vão retornar os seguintes valores, respectivamente

- a) [20, 30, 40] e [1] 20 30 40 50
- b) [20, 30, 40, 50] e [1] 30 40 50 60
- c) [30, 40, 50, 60] e [1] 30 40 50
- d) [30, 40, 50] e [1] 20 30 40 50
- e) [30, 40, 50] e [1] 20 30 40

Comentários: Ao executar o código acima em interpretadores para as respectivas linguagens vamos observar os valores presentes na alternativa D.

Gabarito: alternativa D.

8.

Considere o código Python, versão, na qual o comando print não requer parênteses.

A execução desse código:

```
def teste(n):
    for k in range(1, n+1):
        yield k
```



```
for x in teste (10):  
    print(x)
```

- a) não tem efeito, pois nenhum comando print é ação;
- b) provoca a exibição do número 10 na saída;
- c) provoca a exibição dos números de 1 até 10 na saída;
- d) provoca um erro de compilação.
- e) provoca um erro de execução;

Comentários: Observe que na linha quarta linha do código, temos um o comando **yield** alinhado com o comando da linha anterior. Pelos critérios de indentação do Python isso gera um erro de compilação. Logo, temos a resposta na alternativa D.

Gabarito: alternativa D.

9.

Na linguagem de programação Python, existem 3 estruturas para armazenar dados indexados. A estrutura cujos valores são imutáveis depois de sua criação é conhecida como

- a) lista.
- b) operador.
- c) tupla.
- d) classe.
- e) dicionário.

Comentários: A questão pede as três estruturas para armazenar dados indexados. Elas são: Listas, Tuplas e Dicionários. Uma tupla é uma estrutura bastante similar a uma lista, com apenas uma diferença: os elementos inseridos em uma tupla não podem ser alterados, diferente de uma lista onde podem ser alterados livremente.

Gabarito: alternativa C.

10.

Um analista implementou o seguinte programa em Python:

```
01 a = [1, 2, 3];  
02 b = list(a);
```



03 print(a);
04 print(b);
05 print(a==b);
06 print(a is b);

Assinale a alternativa que apresenta corretamente o resultado do programa quando executado.

A)

A resposta do programa é:

[1, 2, 3]

[]

False

False

B)

A resposta do programa é:

[1, 2, 3]

[1, 2, 3]

True

True

C)

A resposta do programa é:

[1, 2, 3]

[1, 2, 3]

True

False

D)

A resposta do programa é:

[1, 2, 3]

[1, 2, 3]

True



Null

E) O programa não executa por um erro na linha 02, em que é atribuído um tipo de dado diferente daquele definido na variável b.

Comentários: Processando o código acima, temos:

01 **a = [1, 2, 3];** → esse comando cria uma lista a com os valores 1, 2 e 3.

02 **b = list(a);** → esse comando atribui à variável b os valores que estão presentes na lista a, criada acima.

03 **print(a);** → ao executar esse comando imprimiremos os valores 1, 2 e 3.

04 **print(b);** → ao executar esse comando imprimiremos os valores 1, 2 e 3.

05 **print(a==b);** → essa comparação nos retorna true, pois os valores de a e b são iguais.

06 **print(a is b);** → essa comparação nos retorna false, pois a variável a é diferente da variável b.

Gabarito: alternativa C.

11.

Um analista do MJSP armazenou nomes e ID no seguinte programa Python:

```
01 dados = {  
02 'Machado de Assis': 7575,  
03 'Santos Dumont': 7475,  
04 'Rui Barbosa': 1675,  
05 };
```

Agora, o analista necessita apresentar em tela o ID 7475 referente ao nome informado. Assinale a alternativa que apresenta o código correto para imprimir esse ID na tela para o usuário.

- A) **print(7475);**
- B) **print(dados[nome], ID==7475);**
- C) **print(ID, 'Santos Dumont');**
- D) **print(dados == 'Santos Dumont');**
- E) **print(dados['Santos Dumont']);**



Comentários: A questão apresenta uma estrutura de dicionário em Python. Essa estrutura armazena pares de chave e valor. Assim, para recuperar determinado valor, é necessário especificar sua chave.

Na questão, temos um dicionário denominado "dados". O primeiro registro possui chave 'Machado de Assis' e valor 7575.

A questão pede a impressão do valor 7475, que possui por chave 'Santos Dumont'.

Em Python, para recuperar um valor de um dicionário, dada a sua chave, utilizamos a seguinte estrutura:

```
nome_do_dicionario[chave]
```

Portanto, para recuperar o valor 7475, utilizamos dados['Santos Dumont'].

Gabarito: alternativa E.

12.

Analise o seguinte código Python, escrito por um cientista de dados, e assinale a alternativa correta.

```
In [01]: nome = input("Qual  
o seu nome? ")  
In [02]: print(nome)
```

- A) O código não irá funcionar pois está faltando ';' ao final de cada linha.
- B) O uso da função input está incorreto, pois ela tem que ser chamada em uma nova linha.
- C) O resultado do comando print é apresentar na tela o nome informado pelo usuário.
- D) Falta a estrutura correta de criação de uma classe para a execução do comando.
- E) Falta definir o tipo da variável nome, por isso o programa não irá executar.

Comentários: ATENÇÃO! In [01] e In [02] servem-nos aqui apenas de marcação para as linhas. O código mesmo não os possui, pois caso eles existam gera erro de compilação.

Esse código tem a missão de pedir ao usuário para informar o seu nome e em seguida será impresso o nome digitado na tela.

Gabarito: alternativa C.

13.



Analise o seguinte código Python, escrito por um cientista de dados, e assinale a alternativa correta.

In [01]: num = 88

In [02]: print ('par' if
num%2 == 0 else 'ímpar')

- A) O resultado da execução de código é a informação 'ímpar' na tela.
- B) O código apresenta como resultado "1" na linha [02], pois num%2 == 0 é uma atribuição e não uma comparação.
- C) O código executado apresenta na tela o resultado '0'.
- D) O resultado da execução de código é a informação 'par' na tela.
- E) O código gera um erro, pois na linha [02] o comando if...else não está correto.

Comentários: No código acima temos a variável num sendo declarada e recebendo como atribuição (=) o valor 88.

Na linha 2, dentro do comando print, temos uma estrutura de seleção que nos retorna par, caso o resultado buscado seja verdadeiro, e ímpar, caso falso.

Para saber se o resultado é par ou ímpar foi utilizado o operador matemático %, que representa o resto de uma divisão entre números inteiros.

Se dividirmos qualquer número par do universo por 2 o resto será 0 e se o fizermos a divisão de qualquer número ímpar por 2 o resto será 1.

A questão nos deu a seguinte operação:

$\text{num}\%2 == 0 \rightarrow$ Vamos alterar a variável pelo seu valor, com isso temos
 $88\%2 == 0 \rightarrow$ 88 dividido por 2 dá 44, com resto 0.

Assim, o código acima nos retorna par na impressão.

Gabarito: alternativa D.



ESSA LEI TODO MUNDO CONHECE: PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.



1

Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.



2

Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).



3

Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).



4

Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).



5

Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).



6

Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).



7

Concursado(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.



8

O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.



Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.