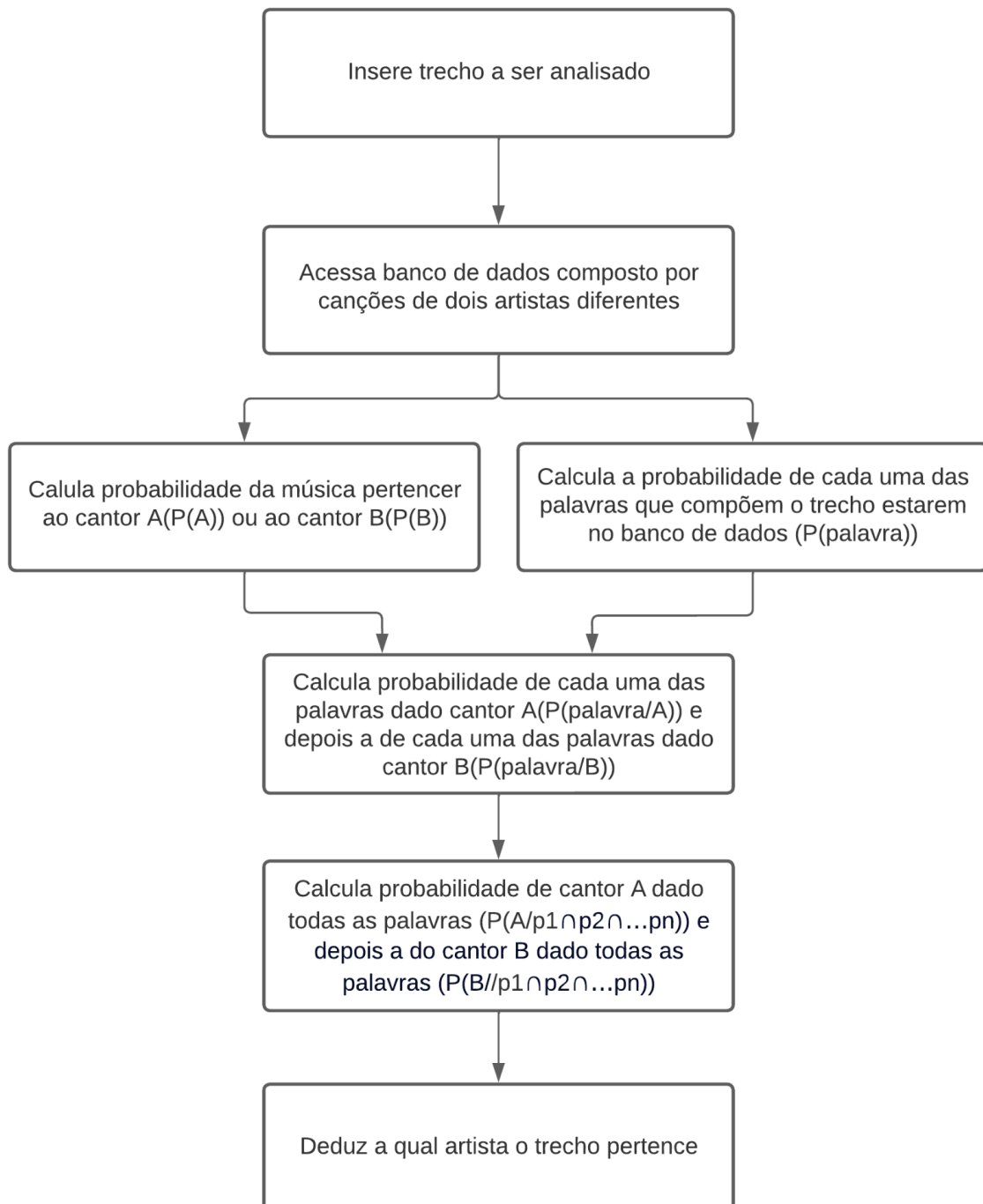


Classificador Bayesiano decifra se trecho de música pertence a Luísa Sonza ou Padre Fábio de Melo, será que você consegue também?

Giulia Gomes Vallente, Jonas Bonfá Pelegrina e Nina Savoy



O programa desenvolvido capaz de estimar a quem pertence o trecho de música inserido

foi criado e baseado no Teorema de Bayes, representado pela forma:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

O primeiro passo em nosso programa é a inserção de um trecho de música a ser analisado. Logo após a escolha do trecho, o banco de dados - composto por diversas músicas, tanto do artista A quanto do artista B - é acessado. Assim, a probabilidade da música pertencer ao artista A, ou seja, $P_{(A)}$, como referenciado na terceira linha do fluxograma, é calculada pela divisão da quantidade de músicas que são do artista A, pela

quantidade de músicas totais:

$$P(A) = \frac{\text{Quant. de musicas do artista A}}{\text{Quant. de musicas totais}}$$

Deste mesmo modo, calculamos a probabilidade da canção pertencer ao artista B, $P_{(B)}$:

$$P(B) = \frac{\text{Quant. de musicas do artista B}}{\text{Quant. de musicas totais}}$$

Além disso, também calculamos a probabilidade de cada uma das palavras estarem presentes no banco de dados, para isso, assumimos a independência entre cada uma das palavras e a cada música que a palavra aparecer, somamos um. Assim, obtemos $P_{(palavra)}$ - como também referenciado na terceira linha do fluxograma - pela divisão da quantidade de palavras pela quantidade total de músicas.

$$P(palavra) = \frac{\text{Quant. de musicas que possuem a palavra}}{\text{Quant. de musicas totais}}$$

Após calcular as probabilidades iniciais, o programa obtém a probabilidade de cada um das palavras dado a canção do artista A, ou seja, a probabilidade é calculada por meio da divisão da quantidade de músicas que pertencem ao artista A e que possuem a palavra selecionada em sua letra pela quantidade de músicas que pertencem ao artista A, como representado por $P_{(palavra/A)}$ na quarta linha do fluxograma:

$$P(palavra/A) = \frac{\text{Quant. de musicas do artista A que possuem a palavra}}{\text{Quant. de musicas do artista A}}$$

O mesmo raciocínio é repetido para o cálculo de $P_{(palavra/B)}$:

$$P(palavra/B) = \frac{\text{Quant. de musicas do artista B que possuem a palavra}}{\text{Quant. de musicas do artista B}}$$

Contudo, é possível que o trecho inserido possua uma palavra que não está presente em nenhuma canção de um dos artistas, mas presente em pelo menos uma música do outro. Assim, afim de impedir erros no cálculo final da probabilidade de qual cantor detém aquele trecho, é preciso calcular a probabilidade de não palavra - $P_{(não\ palavra)}$, obtida pela subtração de 1 menos a probabilidade da palavra - $P_{(palavra)}$:

$$P(não\ palavra) = 1 - P(palavra)$$

Além disso, precisamos calcular também a probabilidade de não palavra dado a canção do artista, a qual pode ser obtida seguindo a mesma lógica citada acima, ou seja, por meio da subtração de 1 menos a probabilidade da palavra dado canção do artista. Assim, precisamos realizar este cálculo tanto para o artista A, quanto para o artista B.

$$P(\text{não palavra}/A) = 1 - P(\text{palavra}/A)$$

$$\text{e}$$

$$P(\text{não palavra}/B) = 1 - P(\text{palavra}/B)$$

Por último, utilizamos o Teorema de Bayes para calcular a quem pertence o trecho analisado, para isso, calculamos a probabilidade de artista dado trecho selecionado. Deste modo, para obter o resultado precisamos multiplicar a probabilidade da música pertencer ao artista - $P_{(A)}$ ou $P_{(B)}$, dependendo de qual está sendo analisado - pela probabilidade de cada uma das palavras dado artista - $P_{(\text{palavra}/A)}$ ou $P_{(\text{palavra}/B)}$, também dependendo de qual está sendo analisado - e dividir isto pela multiplicação da probabilidade das palavras - $P_{(\text{palavra})}$.

$$P(A/p_1 \cap p_2 \cap \dots \cap p_n) = \frac{P(A) \cdot P(p_1/A) \cdot P(p_2/A) \cdot \dots \cdot P(p_n/A)}{P(p_1) \cdot P(p_2) \cdot \dots \cdot P(p_n)}$$

e

$$P(B/p_1 \cap p_2 \cap \dots \cap p_n) = \frac{P(B) \cdot P(p_1/B) \cdot P(p_2/B) \cdot \dots \cdot P(p_n/B)}{P(p_1) \cdot P(p_2) \cdot \dots \cdot P(p_n)}$$

Entretanto, é preciso verificar se há alguma palavra que não existe em nenhuma música do cantor analisado. Se existir uma palavra que atenda às condições citadas acima, precisamos substituir a probabilidade de determinada palavra - $P_{(\text{palavra})}$ - pela probabilidade de não palavra - $P_{(\text{não palavra})}$. Precisamos também realizar o mesmo com a probabilidade de palavra dado artista - $P_{(\text{palavra}/A)}$ ou $P_{(\text{palavra}/B)}$, dependendo de qual está sendo analisado. Assim, utilizamos a probabilidade de não palavra dado artista - $P_{(\text{não palavra}/A)}$ ou $P_{(\text{não palavra}/B)}$, também dependendo de qual está sendo analisado.

Desta maneira, obtemos $P_{(A/p_1 \cap p_2 \cap \dots \cap p_n)}$ e $P_{(B/p_1 \cap p_2 \cap \dots \cap p_n)}$, como referenciado na quinta linha do fluxograma e, assim, aquele artista que obter uma maior probabilidade, é dito como detentor do trecho inserido.

Através deste raciocínio construímos um programa que é capaz de identificar com 90% de precisão se um trecho selecionado pertence ao Padre Fábio de Melo ou à Luísa Sonza, mas será que você é capaz de realizar o mesmo feito? A quem pertence o trecho:

“Eu pensei que podia viver, por mim mesmo
Eu pensei que as coisas do mundo
Não iriam me derrubar”

Se você disse Padre Fábio de Melo, você acertou!

Auto-avaliação

Conceito A. O fluxograma e o texto encaixam-se na categoria A, visto que o fluxograma resume em poucos passos todo o trabalho desenvolvido, enquanto o texto explica passo por passo como calcular todas as probabilidades necessárias - utilizando fórmulas e descrevendo o processo - afim de obter o detentor do trecho analisado. Fora isso, o texto mantém-se conectado com o fluxograma por meio de referências a variáveis citadas no último. Além disso, o título do projeto e o texto instigam a curiosidade do leitor a descobrir o que é o Teorema de Bayes e de saber se conseguem descobrir também a quem o trecho pertence.