**Processamento**

* Corpus Balanceado (descrever as classes utilizadas no corpus)

|  |  |  |
| --- | --- | --- |
|  | Positivo | Negativo |
| Quantidade | 201 | 201 |

* Corpus Desbalanceado (descrever as classes utilizadas no corpus)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Positivo | Negativo | Ambas(Positivo e Negativo) | Neutra | Irrelevante | Ironia |
| Quantidade | 414 | 201 | 9 | 122 | 41 | 34 |

* Resultado Corpus Desbalanceado (Accuracy)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | Pré-processamento | Naive Bayes | Support Vector Machine | Regressão Linear |
| 1 | Tokenização  Remoção das stopwords, das URL, das hastags, das mentions, dos nomes dos candidatos e dos emog  Stemização  Bag-of-Word + Frequencia | **82,45%** | 78,05% | 79,19% |
| 2 | Tokenização  Remoção das stopwords, das URL, das hastags, das mentions, dos nomes dos candidatos e dos emog  Stemização  Bag-of-Word + TF-IDF | 73,82% | 80,49% | 75,44% |
| 3 | Tokenização  Remoção das stopwords, das URL, das hastags, das mentions, dos nomes dos candidatos e dos emog  Stemização  Bag-of-Word + Frequencia + Word-Bigram | 80,97% | 75,62% | 78,53% |

* Resultado Corpus Balanceado versão 2 (Accuracy)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | Pré-processamento | Naive Bayes | Support Vector Machine | Regressão Linear |
| 1 | Tokenização  Remoção das stopwords, das URL, das hastags, das mentions, dos nomes dos candidatos, dos emog e dos acentos  Stemização  Bag-of-Word + Frequencia | 80,95% | **82,60%** | 79,68% |
| 2 | Tokenização  Remoção das stopwords, das URL, das hastags, das mentions, dos nomes dos candidatos, dos emog e dos acentos  Stemização  Bag-of-Word + TF-IDF | 81,13% | 80,65% | 80,32% |
| 3 | Tokenização  Remoção das stopwords, das URL, das hastags, das mentions, dos nomes dos candidatos, dos emog e dos acentos  Stemização  Bag-of-Word + Frequencia + Word-Bigram | 81,30% | 80,49% | 80,97% |

Matriz de confusão.

Desbalanceada

[[100 5]

[ 13 35]]

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 100 | 2 |
| Negativo | 26 | 26 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 88 | 9 |
| Negativo | 30 | 27 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 107 | 3 |
| Negativo | 20 | 24 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 100 | 5 |
| Negativo | 13 | 35 |

Balanceado

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 37 | 9 |
| Negativo | 12 | 43 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 41 | 11 |
| Negativo | 9 | 40 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 45 | 13 |
| Negativo | 8 | 34 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Previsão | |
| Positivo | Negativo |
| Real | Positivo | 40 | 5 |
| Negativo | 3 | 52 |