**Informações da etapa de coleta para mineração de texto**

**Dados da Equipe**

|  |  |  |
| --- | --- | --- |
| Nome | Idade | Formação |
| Eliaquim Moreira do Nascimento | 22 anos | Aluno de Graduação |
| Orisvaldo Jesus | 24 anos | Aluno de Graduação |

**Dados da Coleta**

A coleta foi feita de uma plataforma de e-participação brasileira em que é permitida a coleta de dados, o período da coleta foi entre 30 de maio de 2017 até 6 de Junho de 2017. A coleta foi feita através da linguagens de programação Python em que é pode ser uma biblioteca selenium, a ferramenta geckodriver para o auxilio da manipulação do Browser para que seja feita a coleta com Web Crawler de paginas dinamica em que foi coletado do projeto com maior quantidade de cometarios da plataforma Vote na Web.

|  |  |  |  |
| --- | --- | --- | --- |
| ID Fonte | Fonte(s) | Quantidades de Opiniões  coletadas | Quantidade de palavras por Opinião |
| 01 | http://www.votenaweb.com.br/projetos/plc-1767-2015 | 7440 | 42,39 |

**Coleção Dourada – Anotação**

Foram anotados 500 comentários por dois anotadores no link a seguir pode ser baixado a coleção dourada feita:

<https://www.dropbox.com/s/ghn2o7nwx3elofa/colecao_dourada.csv?dl=0>

Com Coeficiente Kappa/Cohen’s com 12.34% em que é considerado a capacidade humana de avaliação subjetividade de texto é está em torno de 75% a 85% entre outro. Em que foi feito as regras de anotação da forma a seguir:

|  |  |  |
| --- | --- | --- |
| id | Tipo | Descrição |
| 1 | Positivo | texto subjetivo contendo apenas palavras positivas sobre o contexto |
| 2 | Negativo | texto subjetivo contendo apenas palavras negativas sobre o contexto |
| 3 | Ambas | texto subjetivo contendo palavras positivas e negativas sobre o contexto |
| 4 | Neutra | texto objetivo sobre o contexto, ou seja, sem qualquer opinião |
| 5 | Irrelevante | propagandas, textos não relacionados à temática do projeto e etc |
| 6 | Ironia | texto com opinião, mas que contém emoticon ou hastag que indica opinião contrátia. Ex.: #sqn |

Em que foi quantificado da seguinte forma:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID Fonte | Total  de Opiniões | Positiva | Negativa | Neutra | Ambas | Irrelevante | Irônica |
| 1 | 500 | 96 | 322 | 6 | 61 | 6 | 9 |