



**Universidade Federal da Bahia
Instituto de Matemática e Estatística**

Projeto da disciplina Mineração de Texto

Bruno A. de Carvalho, Carla N. C. da Costa, Davi P. da Silva, Tiago Carneiro

Prof. Crystian Paixão

Salvador, junho de 2019

Sumário

- 1. Introdução**
- 2. Objetivo**
- 3. Mineração de Texto**
- 4. Conclusão**
- 5. Referências**

Mineração de Texto com as legendas de Game of Thrones.

Bruno Carvalho¹, Carla Caldeira¹, Davi Paiva¹, Tiago Carneiro¹

¹Departamento de Estatística do Instituto de Matemática e Estatística– Universidade Federal da Bahia (UFBA)

Avenida Ademar de Barros s/n, Campus de Ondina– 40.170-110 – Salvador – BA – Brazil

bruno.carvalho@live.com, carlacaldeira@hotmail.com.br,
davipaivadasilva@gmail.com, carneiro.tiago@yahoo.com.br

***Resumo.** O presente artigo tem como objetivo aplicar conhecimentos em Mineração de Texto, através da descrição de seus conceitos e definições aplicadas a análise das legendas de Game of Thrones.*

1. Introdução

Game of Thrones (GOT) é uma série de televisão norte-americana baseada na série de livros A Song of Ice and Fire. A serie centra-se no Trono de Ferro dos Sete Reinos, onde duas famílias dominantes estão lutando pelo seu controle e segue um enredo de alianças e conflitos entre as famílias nobre dinásticas. Visando extrair os padrões e tendências sobre a série foi utilizado a técnica de mineração de texto.

O estudo de mineração de texto utiliza técnicas de análise e extração de dados a partir de textos, frases ou apenas palavras utilizando algoritmos e técnicas que extraem de dados estruturados ou semi-estruturados informações uteis.

Para extração dos dados foram utilizadas legendas em inglês das oito temporadas da série, utilizando a biblioteca VADER (Valence Aware Dictionary e sEntiment Reasoner) que retorna um escore após a análise de sentimento.

2. Objetivo

Realizar uma análise de sentimento de todas temporadas GOT através da biblioteca VADER para quantificar os sentimentos quanto a intensidade de sua emoção e qualificar quais episódios apresentaram perfil positivo e negativo na série.

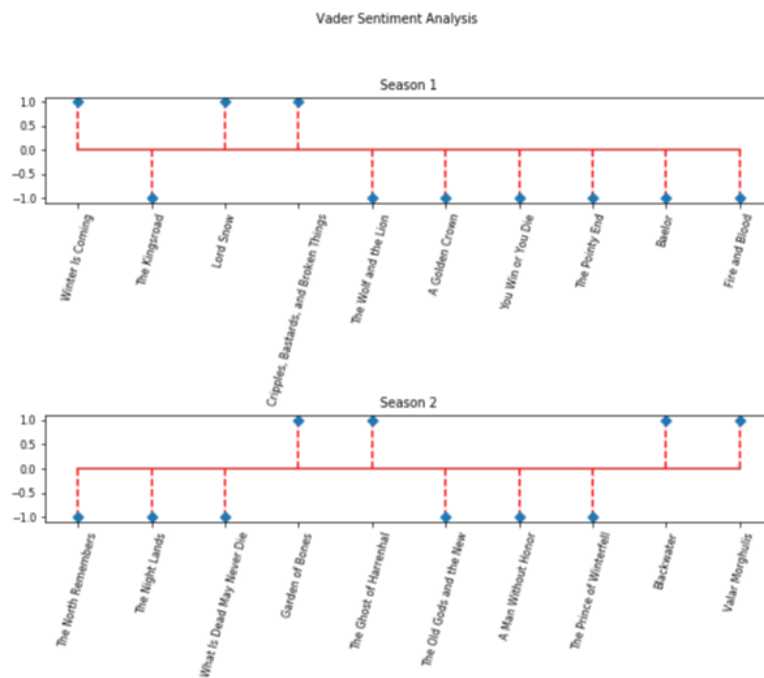
3. Mineração de texto

Como fonte de dados, foram utilizados arquivos de legenda em inglês, no formato .str e extraídas do site <http://www.tvsubtitles.net/tvshow-911-1.html>. Cada arquivo é composto por falas de um determinado episódio da série GOT, em cima desses arquivos foi realizado um pré-processamento, retirando os marcadores de temporais das falas, bem como a substituição da grafia contraída para a forma não contraída.

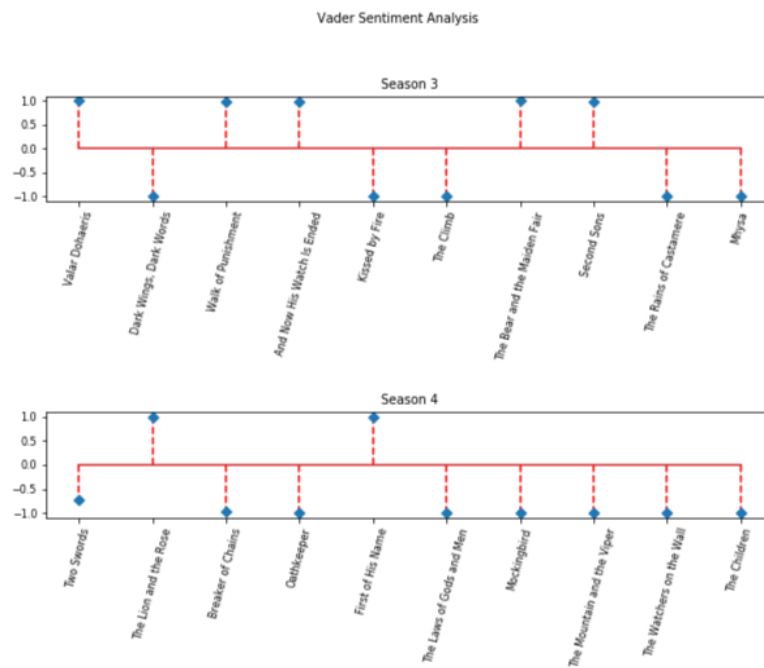
As “StopWords” também podem ser palavras que apresentam uma incidência muito alta em uma coleção de documentos por isso foram incluídas algumas palavras/termos característicos desse universo, como: “*world*”, “*lord*” e “*lady*”.

A word cloud shaped like a map of Westeros from the Game of Thrones series. The words are in various sizes and orientations, representing terms related to the show. The largest words include 'sister', 'sword', 'years', 'honor', 'last', 'stark', 'lannister', 'hand', 'life', 'house', 'kill', 'die', 'name', 'house', 'sister', 'years', 'honor', 'last', 'stark', 'lannister', 'hand', 'life', 'house', 'kill', 'die', 'name', 'house'.

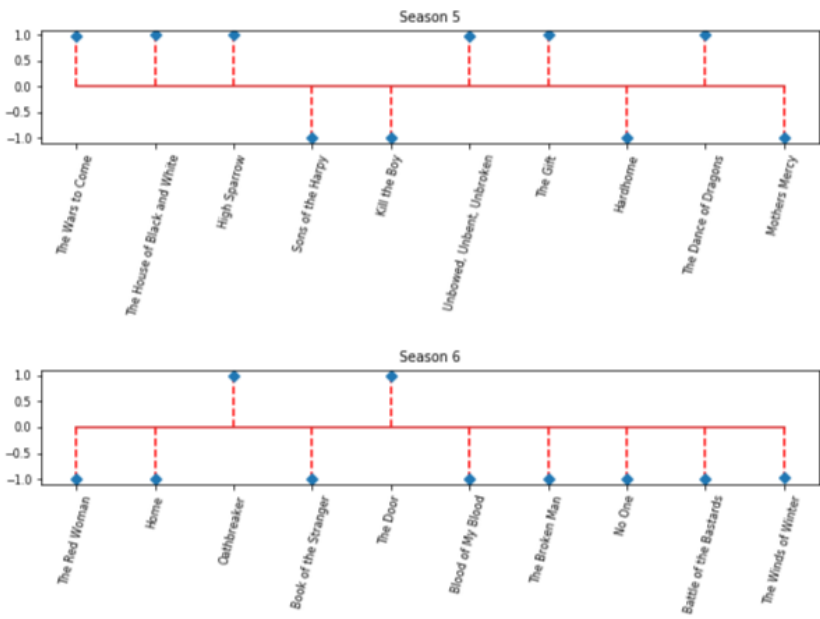
De posse dos “*compound scores*” obtidos foram gerados gráficos no formato *lollipop* ou *stem* como forma de melhor representar a pontuação obtida com a análise de sentimento.



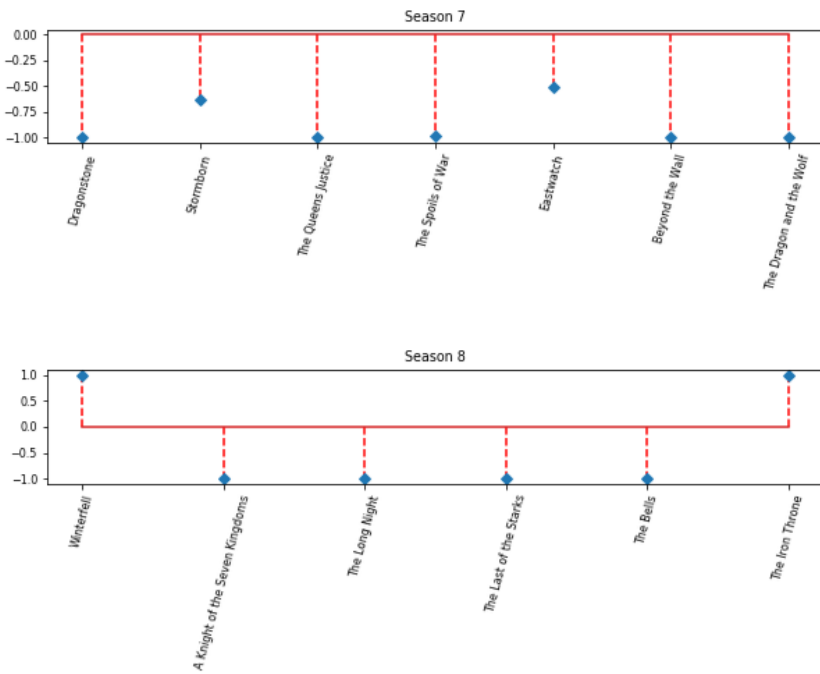
Os dados foram divididos por nome do episódio, marcados com a pontuação de sentimento obtida.



Vader Sentiment Analysis



Vader Sentiment Analysis



4. Conclusão

Levando em consideração que a série trata de temas pesados, como estupro, incesto e tortura, nos causou surpresa a presença de episódios com pontuação positiva, acreditamos que por se tratar de uma transcrição somente das falas, é possível que não seja totalmente expresso na fala, o que está sendo apresentado em vídeo. No entanto, os episódios com escores positivos representam aproximadamente 33% de episódios da série. A temporada mais negativa foi a 7ª, com todos os episódios com escores negativos.

5. Referências

- [1] BARION, Eliana Cristina Nogueira; LAGO, Decio. MINERAÇÃO DE TEXTOS: Text mining. Revista de Ciências Exatas e Tecnologia, Valinhos, São Paulo, v. 3, n. 3, p.123-140, 08 dez. 2008.
- [2] HUTTO, Cj. Descrição do Projeto: VADER-Sentiment-Analysis. 2018. Disponível em: <<https://pypi.org/project/vaderSentiment/>>. Acesso em: 13 jun. 2019.