21/01/2020 - **RELATÓRIO SOBRE NOSSO TRABALHO**

* **Cheguei na Biblioteca do Senac hoje as 15h. Talvez seja mais fácil me concentrar aqui.**
* **O artigo sobre SentilexPT é excelente. Usa um léxico com palavras em português e volta 1 para positivo, -1 para negativo e 0 para neutro. A idéia é usar para verificar se as frases de nosso arquivo de treino estão com a polaridade correta.**
* **Baixei o arquivo (**<https://b2share.eudat.eu/records/93ab120efdaa4662baec6adee8e7585f> **) e contem 7.014 palavras, rotuladas.**
* **Criei na pasta \SERIQUE\_TCC o programa SentilexPT.ipynb que faz teste do Léxico.**
* **Criei na pasta \SERIQUE\_TCC o programa COMPARAR.ipynb cuja função é verificar se os rótulos das frases existentes no arquivo de treino (** **Suicidas\_NaoSuicidas\_Elizangela.csv) estão corretos.**
* **Lembrando que o arquivo Suicidas\_NaoSuicidas\_Elizangela.csv (428) contém somente a junção dos arquivos Suicidas.txt e NaoSuicidas.txt do programa da Elizângela. Mantém os parágrafos em sua forma original e o rótulo positivo para suicida e negativo para não-suicida foi colocado manualmente em função da origem do arquivo.**
* **Criando o programa ROTULADOR.ipynb que usando o léxico SentilexPT, rotula frases.**
* **Capturando textos do site** <https://www.psicologosberrini.com.br/ansiedade-sintomas-e-tratamento/> e gravando no arquivo Frases.txt para usar como entrada no Rotulador.ipynb

20/01/2020 - **RELATÓRIO SOBRE NOSSO TRABALHO**

* **A pesquisa sobre word2vec não resultou em nada prático para nosso trabalho. O que se obtem com o word2vec/doc2vec é a previsão da palavra próxima e isso aparentemente não ajuda na criação da nossa Bag of Words ou na classificação.**
* **Hoje, pretendo:**

1. **Gerar tabelas com os resultados, considerando Naive Bayes e outros 3 classificadores**
2. **Começar a redigir o artigo, informando caso de uso e metodologia.**
3. **Se der, pesquisarei sobre PANAS-T**

15/01/2020 - **RELATÓRIO SOBRE NOSSO TRABALHO**

* **Pesquisando sobre o funcionamento de word2vec e doc2vec. Parece que são 2 algoritmos com os quais é possível fazer uma classificação mais inteligente considerando contexto.**

10/01/2020 - **RELATÓRIO SOBRE NOSSO TRABALHO**

**Hoje aconteceu as 13:30h reunião com a Jerusa. Expliquei sobre o trabalho e rodamos até Nossa\_Analise\_Sentimentos.ipynb.**

**Alterei o ChermaCapturaTwitter e o Limpando\_Twitter para somente lidar com separador “;” e para apagar emoticons.**

1. **Trabalhando no Analise\_Sentimento\_Liu.ipynb que faz pipeline com 4 tipos de algoritmos**
2. **Pegar resultados com Suicidas\_NaoSuicidas\_Elizangela.csv e Suicidas\_NaoSuicidas\_Arquivo\_deTreino:**
3. **Criar “Bag of Words” para os emoticons;**
4. **Pesquisar como usar o Doc2Vec, porque parece mais rápido e mais eficiente;**
5. **Terminar de Ajustar o Analise\_Sentimentos\_Liu.ipynb**
6. **Pensar um jeito de criar dados rotulados usando a escala PANAS-T.**

09/01/2020 - **RELATÓRIO SOBRE NOSSO TRABALHO**

1. **Refiz o programa TRATA\_TWITTER e o Nossa\_Analise\_Sentimentos.ipynb**

08/01/2020 - **RELATÓRIO SOBRE NOSSO TRABALHO**

1. **Reunião de equipe, minha casa , 17 h.**

06/01/2020 - **RELATÓRIO SOBRE NOSSO TRABALHO**

1. **Voltei a trabalhar no QuebraLinha.opynb**
2. **Marcar reunião de equipe**

25/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

* **C:\Users\particular\SERIQUE\_TCC\CAPTURA SITES DA WEB trabalhando com QuebraLinha.ipynb**

**Tentando quebrar as linhas dos parágrafos para usar como entrada para algoritmo (não tá funcionando).**

* **A Leila ficou de editar o arquivo suicidas\_nao\_suicidas.csv**

15/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

**Reunião de domingo 16h em minha casa**

14/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

**Lembrar sequência:**

1. **Rodar ChermaCapturaTwitter (diversas vezes) – colocando como parâmetro = (suicídio, suicídio, depressão, etc) - alternar buscar no grupo do CVV e nos outros;**
2. **Rodar LIMPANDO\_TWITTER – este programa agrupa todos os twitters em um único chamado “arquivo\_todos.csv” e depois faz processamento para retirar todos os caracteres e palavras desnecessárias e deixar pronto para entrar para classificação**
3. **Nossa\_Analise\_Sentimentos.ipynb**

* Ontem tivemos reunião com o Serique para orientações do Qualify.
* Pediu que fizéssemos um classificador simples, apenas capturando twitters do CVV e classificando entre (depressivo/não depressivo)
* Pediu que fizéssemos pré-processamento dos twitters
* Preparar apresentação informando interesse em estudar msg que transitam no grupo do CVV
* Enfocar situações em que o classificador não consiga avaliar corretamente o contexto, por exemplo:
* “Tranquilize-se: a depressão não vai durar para sempre.” E “A tranquilidade sumiu e a depressão dura para sempre”.
* Feito hoje:
  + - ChemaCapturaTwitters.ipynb – funcionando ok – gravando palavras com acento no diretório \DADOS
    - LER\_E\_CONCATENAR.ipynb – funcionando ok – testes para concatenar os arquivos em um único
    - LIMPANDO\_TWITTER.ipynb – concatena em um único arquivo os twitters que existirem no diretório \DADOS e a seguir faz a preparação e limpeza dos textos para repassar para o algoritmo de classificação.
    - Nossa\_Analise\_Sentimentos = está dando erro!!!
    - Analise\_de\_Sentimentos – funcionando ok – esse é igual ao Comentarios\_Parte\_Dois.ipynb

T

13/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

* Tentando criar arquivo csv com mensagens para treino, mas tá dando erro na leitura no pandas.
* Criei o Depressao\_msg\_Treino.csv e o Suicidio\_msg\_Treino.csv
* Criei o **Analise\_de\_Sentimento**.ipynb (cópia do Comentarios\_Parte\_Dois.ipynb)

12/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

* Adaptei o programa Comentarios\_Parte\_Dois.ipynb em 2 partes.
* a primeira tem o algoritmo Naive Bayes e classifica o arquivo Comentarios\_test.csv em elogio/reclamação
* A segunda usa arquivo de treino que fiz na mão (poucas linhas). Informa a classe e gera classificação por tipo de classe
* A classe é a informação do assunto, baseado nos tópicos que o LDA mostrou.

11/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

Elabore modelos preditivos que classifiquem um comentário de restaurante com as seguintes classes:

- elogio ou reclamação, com base no arquivo "comentarios.csv".

Para cada tópico identificado na tarefa anterior, crie uma classe para ser identificada pelo modelo preditivo.

Aplique o modelo no arquivo comentarios\_test.csv e mostre os resultados de performance

do modelo vencedor (acurácia, precisão, especificidade e f-measure)

PROBLEMA = Estou empacada porque não sei como lidar com o arquivo. Imagino que o target deva ser igual

A “Classificacao” (se elogio/comentário)

10/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

ONDE PAREI:

* Estou na pasta do Rogério trabalhando em RemoverAcentos.ipynb
* Desisti. Editei o arquivo.csv na mão e retirei os acentos.
* Fiz o programa Comentarios\_Parte\_dois.ipynb que é a segunda parte do trabalho. Usa como entrada um arquivo csv com as colunas (ID, Avaliacao,**Classe**,Titulo,Texto)
* Na coluna Classe fica especificado ao que se refere o comentário (tópicos)
* Estamos trabalhando com Classe(Atendimento/Ambiente/Comida/Preco)

09/12/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

Relembrando: nosso maior problema é que precisamos de dados rotulados suficientes para efetivar o treinamento da máquina.

Começamos a princípios fazendo rotulagem à mão. Pegamos mensagens de sites positivistas e rotulamos como positivo. Igualmente de sites depressivos e rotulamos como negativo.

A ídeia de usar o algoritmo LDA para extrair tópicos de mensagens capturadas e facilitar o processo de rotulação não deu certo, porque os tópicos são extraídos mas o trabalho de análise humana ainda continua necessário para definir as classes em função dos tópicos apontados.

1. Rodei no Jupyter o programa ChermaCapturaTwitter.ipynb – estamos usando a senha de desenvolvedor da Cherma e capturamos mensagens do twitter que apresentem a palavra suicídio.
2. Revendo o vídeo sobre análise de sentimentos com dados do twitter, que o Serique indicou do site “Minerando Dados”, vemos que ele trabalha com uma base de twitters rotulados como positivos/negativos/neutros.
3. A classificação que ele faz é binária. Precisamos pensar se o que queremos é simplesmente isso, ou fazer uma classificação do tipo Multi-classe: Depressivo, Tendência ao Suicídio, etc.

TAREFA DE HOJE:

* Objetivo: Pegar twitter capturados e usá-los como entrada no algoritmo do Classificação\_tweets.ypinb
* Estou trabalhando no **\Usuarios\Particular\SERIQUE\_TCC\TCC\_Classificacao\_tweets.ypinb**
* No diretório **\Usuarios\Particular\SERIQUE\_TCC\CAPTURA\_DADOS,** guardo os programas :

**- ChermaCapturaTwitter.ypinb**

* Para baixar arquivo do github, quando não aparece o botão de download:

utilize o botão “Raw” e depois a opção “Save Page As” de seu navegador.

* Organizando

06/11/2019 - **RELATÓRIO SOBRE NOSSO TRABALHO**

Pensei em descrever o que fiz, para melhor contextualizar vocês. Ainda temos muito pela frente e queria muito que nós três estivéssemos com as ferramentas funcionando para tentar caminhos diferentes. Cada um pode tentar um algoritmo e trazer o que encontrou. Pode ficar bacana.

Aquela indicação do SERIQUE, que é o vídeo do cara do Minerando Texto(<https://www.youtube.com/watch?v=MV-X-a84BFs&feature=youtu.be>) é ótima! Não sei se lembram, mas ele trabalha com 8.199 mensagens de twiter, e elas estão rotuladas como positivas, negativas ou neutras.

Essas são as mensagens que ele usa como entrada para treinar o machine learning.

Com o programa da Cherma, podemos capturar mensagens, mas elas não estão ainda rotuladas como positivas, negativas ou neutras. Talvez possamos usar as mensagens capturadas para testar nossa máquina depois que ela tiver sido treinada.

Eu achei um site com mensagens prontas (<https://www.mundodasmensagens.com>). Ele apresenta mensagens relativas a Amor, Solidariedade, Saudade, etc. Fica fácil rotular essas mensagens, só que eu percebi que a maior parte delas é positiva. Pouca coisa como neutra ou negativa. Precisávamos capturar mensagens depressivas e neutras para criarmos uma base nossa de treinamento.

Acabei de achar um site com idéias negativas que podemos utilizar (<https://www.pensador.com/frases_depressivas/>)

Trabalhando juntos

Vou criar no GitHub um projeto nosso chamado TCC-SENAC. Vocês podem acessar, alterar e tudo o mais com relação aos nossos programas. Eis o link de acesso (<https://github.com/janesd/TCC-SENAC>)

Sobre a Captura de Mensagens

O programa da Cherma captura tweets, mas eles não estão classificados. Vou colocar no GitHub o programa dela com as adaptações que eu fiz, porque aí, ele não pesquisa apenas no site do CVV, mas em tudo. Além disso, a cada rodada ele cria um arquivo com data e hora da captura.

O outro programa que vou colocar no Github é o que captura mensagens do Mundo das Mensagens. Ele pode facilmente ser adaptado para capturar de outros sites, se a gente conseguir descobrir mais algum. Esse programa, classifica as mensagens de forma fixa. Criei uma tabela de-para, assim, mensagens que contenham a palavra Amor, fica positiva, começando com Reflexão é neutra, e assim por diante.

Sobre o nosso Artigo

Nada impede que comecemos a escrever, apesar de nosso estudo de caso ainda não estar pronto. Mas, sugiro que primeiro a gente fique junto, com relação a esse estudo de caso e depois ataquemos a escrita.