# T2 - Análise de Sentimentos em Textos

## Objetivo

O objetivo desta tarefa é expor os alunos aos desafios práticos envolvidos na análise de textos e na atribuição de valores de valência ou rótulos de emoções a sentenças.

Não é objetivo desta tarefa avaliar a acurácia de detecção ou a eficiência do modelo implementado mas a análise crítica do projeto e o amadurecimento em relação ao problema.

## Descrição da Tarefa

Esta tarefa deverá ser realizada individualmente ou em dupla.

Deve-se escolher um entre os dois problemas propostos abaixo.

### Problema 1 - Determinação de Valência em Manchetes de Jornais Brasileiros no 1° Semestre de 2017

* Serão fornecidas 500 manchetes dos jornais “Valor Ecônomico”, “O Globo”, “Folha de São Paulo”, “O Estado de São Paulo”, extraídas em datas específicas dos meses de dezembro de 2016 a agosto de 2017. O acesso à base deve ser realizado pelo link: <https://github.com/pdpcosta/manchetesBrasildatabase>
* O problema consiste em classificar as manchetes em valência positiva, neutra ou negativa, associado ao indicador de intensidade. Exemplo: numa escala de -100 a 100, uma manchete classificada como +80 pode ser considerada muito positiva e uma manchete classificada como 5 pode ser considerada neutra.
* A tarefa consiste em definir a abordagem ao problema, o modelo de classificação, as regras de análise e deverá realizar uma implementação prática do algoritmo definido.

### Problema 2 - Determinação de Rótulo de Emoção

* Este problema utilizará a mesma base de dados utilizada no SemEval 2007 - 4th International Workshop on Semantic Evaluations, Task 14, Affective Tests.
* A base de treinamento conta com 250 manchetes em inglês de jornais e websites (Google, CNN, etc.)
* A cada manchete está associado um score de [0 a 100] para os rótulos “anger”, “disgust”, “fear”, “joy”, “sadness”, “surprise”
* Também será fornecida uma base de testes e os rótulos “golden” fornecidos durante a conferência.
* Todos os dados podem ser acessados pelo link, abaixo. Mas atenção, considerar apenas os dados referentes a rótulos de emoções (identificados com sufixo *emotions*). A leitura dos arquivos README é essencial para o entendimento da base. <http://nlp.cs.swarthmore.edu/semeval/tasks/task14/data.shtml>
* A tarefa consiste em definir a abordagem ao problema, o modelo de classificação, as regras de análise e deverá realizar uma implementação prática do algoritmo definido.

## Ferramentas

Você pode usar qualquer linguagem de programação para implementação de seus modelos e algoritmos ou qualquer ferramenta de análise de mais alto nível (Ex.: [Weka](http://www.cs.waikato.ac.nz/ml/weka/) ou [Microsoft Azure Machine Learning](https://azure.microsoft.com/pt-br/services/machine-learning/?&wt.mc_id=AID623267_sem_RxuJwMFB) ). Mas atenção, é importante que, independentemente da ferramenta, você entenda o algoritmo de análise/rotulação implementado.

Como Pythonista, recomendo o uso da biblioteca de processamento de linguagem natural [NTLK](http://www.nltk.org/).

## Submissão

A submissão deverá ser realizada IMPRETERIVELMENTE até 14 de Setembro, via Google Classroom. Trabalhos indicados com submissão tardia pelo Classroom serão desconsiderados.

A submissão regular incluirá:

1. Relatório da tarefa em PDF.
2. Código fonte e executável da implementação realizada, incluindo arquivo README com instruções claras sobre a execução e a necessidade de instalação de pacotes específicos. Considere como ambiente de referência o Linux Mate 16.04.

Como modalidade alternativa de submissão, notebooks Jupyter, ou equivalente, unificando relatório e implementação, são bem-vindos. O link deve ser disponibilizado no Classroom.

## Como você será avaliado

Será atribuída uma nota de 0.0 a 10.0 ao seu trabalho.

O relatório, tal como um “caderno de experimentos”, deverá ser utilizado para registrar todas suas ideias e tentativas de implementação de um modelo de rotulação de valência (Problema 1) ou de emoções (Problema 2), incluindo o relato de seus possíveis fracassos e sucessos e *insights* que você tenha tido durante o processo.

Após a implementação prática de um modelo, é esperado que você analise criticamente os resultados obtidos de maneira franca, baseada em argumentos objetivos.

* Você concorda com os rótulos que seu algoritmo atribuiu?
* Foi possível descobrir algo relevante sobre os dados a partir da análise de sentimentos? (Por exemplo, no Problema 1, o modelo implementado é capaz de indicar se um jornal tem manchetes mais neutras que outro? Houve algum mês de manchetes particularmente negativas?)
* Considerando que nenhum modelo é perfeito, quais são os pontos fracos do algoritmo implementado? Quais seriam pontos fortes?

Finalize seu relatório registrando as principais lições aprendidas durante a realização da tarefa.

Dentro do razoável, adote um tom profissional e ético, dando créditos a dados utilizados e citando as principais referências utilizadas no contexto do trabalho.

O formato não é tão importante quanto a apresentação do conteúdo de maneira inteligível, objetiva e clara.

É razoável assumir que um relatório neste perfil terá entre 3 e 10 páginas.

|  |  |
| --- | --- |
| Trabalho foi submetido de maneira parcial, sem relatório ou sem código de implementação. | Nota máxima 3.0 pontos. |
| Relatório foi considerado satisfatório (o modelo adotado e a análise crítica da solução implementada estão claramente descritos), mas não foi possível avaliar a execução do código. | Nota máxima 8.0 pontos. |
| Apesar de ter sido possível executar o algoritmo implementado, o relatório apresenta problemas graves, dificultando a compreensão da solução implementada ou não apresentando uma análise crítica do algoritmo. | Nota máxima 5.0 pontos. |

Em caso de dúvidas durante o processo, encaminhe-as via email: [paula@fee.unicamp.br](mailto:paula@fee.unicamp.br)

Esteja ciente que posso demorar de um a três dias para respondê-lo.

Alternativamente, tente a sorte me procurando em minha sala, no terceiro piso, sala 311-A.

Whats App (19) 98801-3938 (por favor identifique-se).