

O trabalho deverá ser realizado em **grupos de 2, 3 ou 4** (aconselhado) **elementos**.

O trabalho deve ser entregue até ao dia **26 de abril**.

As apresentações serão realizadas na **semana de 28 de abril**, durante a aula.

Análise de Sentimento

A tarefa de Análise de Sentimento a partir de um texto é amplamente conhecida na área de Text Mining. A ideia consiste em analisar o texto e identificar qual o sentimento aí presente. Exemplos:

[NEG] If you haven't seen this, it's terrible. It is pure trash. I saw this about 17 years ago, and I'm still screwed up from it. If you haven't seen this, it's terrible.

[POS] Drum scene is wild! Cook, Jr. is unsung hero of this and many movies. Fantastic actor, great flick. A few twists that keep you moving. A must-see. Drum scene is wild!

Apesar da classificação de um texto poder ser feita a vários níveis, tais como *positivo*, *negativo*, *neutro*, *muito positivo* ou *muito negativo*, este problema é muitas vezes abordado como uma tarefa de classificação binária, em que se consideram apenas as classes *positivo* e *negativo*.

Este trabalho pretende explorar vários métodos de classificação de sentimentos em textos e aplicá-los a vários tipos de dados, tais como *tweets* e *reviews* de filmes, *reviews* de produtos, etc.

1 Dados

Para realizar este trabalho estão disponíveis no Moodle vários conjuntos de dados em Inglês, previamente separados em *treino* e *teste*. Cada um dos conjuntos apresenta características e problemas diferentes, o que poderá implicar diferentes tipos de pré-processamento. Cada um dos grupos deverá utilizar um conjunto diferente, por forma a promover a diversidade de trabalhos.

1. **Tweets_EN_sentiment**: Conjunto balanceado, com cerca de 50000 tweets em Inglês, etiquetados com "pos" e "neg".
2. **amazon_reviews**: Conjunto não balanceado, com cerca de 50000 reviews de produtos da empresa *Amazon*, anotadas com as etiquetas "positive" e "negative".
3. **imdb_reviews**: Conjunto não balanceado, com cerca de 44000 reviews de filmes da IMDB, anotadas com as etiquetas "pos" e "neg".
4. **rotten_tomatoes**: Conjunto com cerca de 6500 frases de reviews de filmes, anotadas com as 3 etiquetas: "positive", "negative" e "neutral".

Para obter os dados poderá optar por uma das seguintes soluções:

1. Obter todos os dados disponíveis, seguindo as instruções que estão na secção do moodle: *Conteúdo* > *Acesso aos Exemplos Práticos*. Ao descomprimir o ficheiro **tm-data.zip**, todos os ficheiros se encontram dentro da diretoria *data/en-sentiment*.
2. Cada um dos ficheiros individuais está também disponível junto com o enunciado do trabalho na secção do moodle: *Avaliação* > *Trabalho - Análise de Sentimento*.

Observação. Não deve disponibilizar publicamente nenhum dos conjuntos de dados fornecido.

2 Tarefas

Depois de fazer a análise da literatura, que corresponde à tarefa 2.1, execute as restantes tarefas e reporte os resultados obtidos no seu relatório. Sempre que chegar a um bom resultado pode submeter informação sobre a experiência realizada no **Leaderboard**, que se encontra no endereço: <http://tiny.cc/tmcd>

2.1 Revisão da literatura

Deverá começar por fazer uma pesquisa bibliográfica e identificar uma pequena coleção com $2 \times n^o$ de elementos do grupo artigos, que possam estar relacionados com o seu trabalho. A sua pesquisa deve incluir trabalhos que utilizem a mesma fonte de dados ou fontes de dados semelhantes à que está a usar. Para cada um desses trabalhos, deverá identificar as principais abordagens utilizadas, dados utilizados e os principais resultados obtidos.

2.2 Definição de um *baseline* usando ferramentas já existentes

Verifique que os dados se encontram subdivididos em conjuntos de treino e teste. Pretende-se agora fazer uma classificação inicial **dos dados de teste** usando uma ferramenta de análise de sentimentos já existente. Para isso, deve procurar e escolher uma ferramenta que permita fazer análise de sentimento diretamente a um texto. Uma possível ferramenta é a biblioteca *TextBlob* que se encontra ilustrada nos exemplos da aula, mas poderá usar outra tal como *Vader Sentiment*¹, *spacy* ou *Stanza*², entre outras.

O n^o de elementos do grupo corresponde ao número mínimo de ferramentas a aplicar. Por exemplo, um grupo constituído por 4 elementos deve aplicar e comparar 4 ferramentas diferentes. Escolha e aplique as ferramentas aos seus dados de teste e avalie os resultados obtidos utilizando as métricas adotadas. Note que a *taxa de acerto (accuracy)* poderá ser a mais adequada para problemas binários, em que ambas as classes têm igual importância. Para outros problemas poderá fazer sentido usar outras métricas, tais como: *precisão*, *cobertura* e *medida F1*.

2.3 Preparação de dados e aplicação de um léxico de sentimentos

Pretende-se agora fazer um classificador de sentimentos muito simples, baseado num léxico. Para tal, pode utilizar o NRC Word-Emotion Association Lexicon (EmoLex) que se encontra disponível dentro da diretoria `./data/NCR-lexicon.csv`. O ficheiro original é um ficheiro do tipo CSV que tem o seguinte conteúdo, embora apenas as primeiras 3 colunas do ficheiro (English, Positive, Negative) sejam suficientes para este trabalho.

English	Positive	Negative	Anger	Anticipation	Disgust	Fear	...
abacus	0	0	0	0	0	0	
abandoned	0	1	1	0	0	1	
abandonment	0	1	1	0	0	1	
abba	1	0	0	0	0	0	
...							

Mais uma vez, pretende-se fazer a classificação **dos dados de teste**. Deverá procurar alguma bibliografia que reporte a utilização de léxicos de polaridade por forma a decidir qual a melhor abordagem. No entanto, a ideia geral é quantificar quantas palavras do texto a analisar estão classificadas como *Positive* e *Negative* e, consoante a classe mais frequente, decidir qual a classificação final a atribuir ao texto.

Poderá utilizar outros léxicos alternativos e deverá também fazer algum pré-processamento por forma a melhorar o desempenho obtido. Deve fazer pelo menos duas experiências diferentes: **sem** e **com o tratamento da negação**, comentando os respetivos resultados obtidos. No caso do grupo ter mais do que 2 elementos, deverá fazer todas as experiências com dois léxicos diferentes.

¹<https://pypi.org/project/vaderSentiment/>

²<https://stanfordnlp.github.io/stanza/sentiment.html>

2.4 Treino de um modelo (aprendizagem automática)

Deverá agora tentar treinar um modelo, com o objetivo de tentar melhorar ainda mais os resultados de classificação obtidos:

1. Treine um modelo com os dados de treino, usando uma ferramenta de classificação à sua escolha, tal como por exemplo *scikit-learn*, etc.
2. Aplique o modelo que construiu ao conjunto de teste, avalie os resultados obtidos e compare-os com os resultados obtidos na tarefa anterior.

Faça várias experiências usando diferentes conjuntos ou tipos de *features* e documente os resultados obtidos. Exemplos de experiências: variar o número de *features*; usar maiúsculas e minúsculas ou apenas minúsculas; diferentes tipos de pré-processamento (e.g. normalização de palavras, *stemming*); com e sem *part-of-speech*; usando entidades mencionadas; usando GloVe (*embeddings*); etc.

Observação. Pode basear o seu trabalho em receitas disponíveis na Internet, no entanto deve fazer referência a todo o tipo de material que utilizar, incluindo o endereço onde se encontra a respetiva informação.

2.5 Utilização de transformadores para classificação

Deverá agora aplicar um modelo pré-treinado, baseado em transformadores, aos seus dados: a) numa primeira etapa poderá fazer experiências simples, recorrendo a *pipelines* já definidos, aplicando um ou mais modelos existentes; b) numa segunda etapa, use os seus dados para fazer *fine-tuning* ao modelo pré-treinado e assim conseguir ainda melhores resultados.

2.6 Usar modelos generativos para classificação de sentimento

Finalmente, deverá aplicar um modelo generativo, tal como o chatGPT ou Google Gemini ou outro, para atribuir as etiquetas “pos”, “neg” ao conjunto de dados *notebook-challenge.txt*, que é fornecido sem etiquetas e que contém 421 reviews escritas em português sobre computadores portáteis. Este conjunto encontra-se no moodle (também se encontra dentro da diretoria *data/pt-sentiment* ao descomprimir o ficheiro **tm-data.zip**). Deverá usar o sistema que lhe parecer ter melhor desempenho para atribuir automaticamente as etiquetas a estes dados. Juntamente com esse ficheiro, encontra-se o ficheiro *notebook-devel.csv*, que contém cerca de 1500 reviews previamente classificadas e que pode ser utilizado para fazer as suas experiências e que lhe permite verificar se está a obter bons resultados. Juntamente com o seu relatório, deverá submeter o ficheiro *notebook-challenge.csv*, que deve seguir o formato do ficheiro *notebook-train.csv*, com as colunas “sentiment” e “review_text”, em que a coluna “sentiment” é calculada automaticamente pelo seu sistema. Exemplo:

```
sentiment review_text
pos O note é bom, mas não superou as expectativas ...
pos Ele corresponde ao valor pago. Custando em ...
pos O notebook é muito bom, o Windows 10 ferra um pouco com a velocidade, mas...
pos O notebook tem uma tela enorme, muito útil para ...
neg Comprei para estudar, ainda não instalei um único programa ...
```

Submissões, relatório e avaliação

O trabalho deverá ser entregue através do Moodle. Cada um dos elementos do grupo deverá inscrever-se num dos grupos disponíveis para realização do *Trabalho*. O grupo deverá enviar um ficheiro ZIP contendo o relatório e o código desenvolvido. A apresentação poderá ser entregue mais tarde, até às 17h do dia da apresentação.

Relatório

O grupo deverá produzir um relatório a descrever o trabalho realizado, com um máximo de 12 páginas, em formato PDF, baseado no template disponível em: *Information for Authors of Springer Computer Science Proceedings* (Microsoft word e Latex).

O relatório deverá *obrigatoriamente* seguir a seguinte estrutura: *título; resumo; 1) introdução; 2) dados utilizados* (descrição e caracterização); *3) trabalho realizado*, descrevendo os procedimentos efetuados, as opções tomadas, os recursos utilizados e os resultados obtidos (deve fazer uma subsecção por cada uma das etapas + análise de resultados); *4) trabalho relacionado; 5) conclusões*, incluindo aspetos a melhorar no futuro; e, bibliografia utilizada. O resumo deve indicar os dados e os métodos utilizados, as opções mais relevantes e as principais conclusões obtidas.

Observação. O relatório não deverá ter mais do que seis tabelas com resultados. O resultado de cada uma das experiências poderá ser apresentado, juntamente com uma breve descrição das condições da experiência. Exemplo:

	Acc.	Prec.	...
Baseline (TextBlob)	0.660		
...			
SVM, C=1.0, tokenization, lowercase, stopwords, lemmatization, POS, negation, spell correction , 10K features, ...	0.956	...	

O último parágrafo da última secção do relatório, deve indicar uma estimativa da percentagem de contribuição de cada elemento para o trabalho, juntamente com uma justificação. Por exemplo: manuel: 30%, pedro: 30%, joaquim: 40%.

Código

O grupo deverá também enviar todo o código que permita replicar os resultados obtidos.

Apresentação

O grupo deverá também preparar uma apresentação para fazer durante a aula (as apresentações serão presenciais). A apresentação deverá ser projetada para um máximo de 8 minutos, devendo cobrir: *descrição dos dados, abordagens e opções mais relevantes, resultados obtidos, principais conclusões* e propostas para *trabalho futuro* que podem levar a melhores resultados. Os ficheiros da apresentação devem ser submetidos até às 16h00 do dia das apresentações através do link: <https://t.ly/3uDM9>. O grupo deve criar a pasta GrupoXX, em que XX é o número do grupo, e colocar os respetivos ficheiros dentro dessa pasta.

Observação. Caso assim o entendam, os membros do grupo poderão gravar a sua apresentação previamente (usando o software Zoom), por forma a que no dia das apresentações possam usar o vídeo em vez de fazerem a sua apresentação ao vivo, garantindo assim que cumprem o tempo.

CrITÉrios de avaliação

Na avaliação serão tidos em conta os seguintes critérios:

1. (10v) Correção e multiplicidade de soluções propostas
2. (7v) Estrutura e conteúdo do relatório
3. (3v) Apresentação (é obrigatória).

Eventuais atrasos na entrega do relatório ou apresentação serão penalizados.

Política em caso de fraude

Os alunos podem partilhar e/ou trocar ideias entre si sobre os trabalhos e/ou resolução dos mesmos. No entanto, o trabalho entregue deve corresponder ao esforço individual de cada grupo. São consideradas fraudes as seguintes situações:

- Trabalho parcialmente copiado
- Facilitar a copia através da partilha de ficheiros
- Utilizar material alheio sem referir a sua fonte.

Em caso de deteção de algum tipo de fraude, os trabalhos em questão não serão avaliados, sendo enviados à Comissão Pedagógica ou ao Conselho Pedagógico, consoante a gravidade da situação, que decidirão a sanção a aplicar aos alunos envolvidos. Serão utilizadas as ferramentas *Moss* e *SafeAssign* para deteção automática de cópias.

Recorda-se ainda que o Anexo I do Código de Conduta Académica, publicado a 25 de Janeiro de 2016 em Diário da República, 2ª Série, nº 16, indica no seu ponto 2 que:

Quando um trabalho ou outro elemento de avaliação apresentar um nível de coincidência elevado com outros trabalhos (percentagem de coincidência com outras fontes reportada no relatório que o referido software produz), cabe ao docente da UC, orientador ou a qualquer elemento do júri, após a análise qualitativa desse relatório, e em caso de se confirmar a suspeita de plágio, desencadear o respetivo procedimento disciplinar, de acordo com o Regulamento Disciplinar de Discentes do ISCTE - Instituto Universitário de Lisboa, aprovado pela deliberação n.º 2246/2010, de 6 de dezembro.

O ponto 2.1 desse mesmo anexo indica ainda que:

No âmbito do Regulamento Disciplinar de Discentes do ISCTE-IUL, são definidas as sanções disciplinares aplicáveis e os seus efeitos, podendo estas variar entre a advertência e a interdição da frequência de atividades escolares no ISCTE-IUL até cinco anos.