

Sistema de Detecção e Classificação de Máscaras Faciais

José Gonçalves^[pg50519], Rodrigo Rodrigues^[PG50726], and Rui Monteiro^[PG50739]

Grupo 1

Universidade do Minho, Departamento de Informática

1 Introdução

O presente relatório científico tem como objetivo investigar a possibilidade de prever os retornos de criptomoedas através da análise do sentimento de *tweets*. Para isso, será utilizado um método de web scraping na rede social *Twitter* para recolher informações sobre as palavras-chave e *hashtags* relacionadas às criptomoedas. A análise de sentimento será realizada para avaliar a polaridade dos *tweets* em relação às criptomoedas, que podem ser positivos, negativos ou neutros, sendo associado o respetivo valor. Será também detalhado o processo de obter dados passados dos valores atingidos pela criptomoeda, assim como todo o processo de tratamento dos mesmos. Com base nessas informações, será possível implementar técnicas de previsão de valores de criptomoedas, utilizando técnicas de *machine learning*, mineração de dados e análise estatística. O objetivo deste estudo é analisar um possível método de obtenção de previsões, procurando contribuir para o avanço do conhecimento na área de criptomoedas e fornecer informações úteis para investidores e profissionais financeiros que desejam tomar decisões informadas em relação a esse mercado em rápida e constante evolução.

A metodologia utilizada no estudo é o CRISP-DM (Cross-Industry Standard Process for Data Mining), que é um processo padrão para mineração de dados. O CRISP-DM consiste em seis fases: entendimento do problema, entendimento dos dados, preparação dos dados, modelagem, avaliação e implementação. O uso desta metodologia permite uma abordagem sistemática e estruturada para a resolução do problema em questão.

Deste modo, o presente relatório começará por relatar num capítulo inicial o processo de obtenção de *tweets*, bem com a interpretação, tratamento e manipulação destes dados, seguindo-se de um capítulo relativamente similar, mas para os dados históricos da criptomoeda em análise. Por fim, será abordada e interpretada a análise efetuada, bem como as técnicas utilizadas e resultados obtidos.

2 Tweets

Os dados (*tweets*) obtidos são relativos a um período temporal específico, a semana anterior ao evento *Ethereum merge*, ou seja, a mudança do mecanismo de *proof of work* (PoW) to a *proof-of-stake* (PoS), que teve como objetivo reduzir 99% em custos energéticos no processamento de transações de *Ethereum*, que aconteceu em 15 de Setembro de 2022.

Este período temporal foi escolhido por ser um acontecimento que leva a discussão entre diversos utilizadores da *blockchain* e pode ser benéfico para o caso de estudo deste projeto, sendo um período sólido da história da criptomoeda.

2.1 Obtenção de dados

Inicialmente, recorreu-se a ferramentas como o *snsrape* para obter *tweets* de forma automática, recorrendo ao sistema de pesquisa avançada do Twitter, que permite considerar diversas *keywords*, fazer restrições relativas a palavras e *hashtags* usadas e ao período temporal dos mesmos para antes do evento *Ethereum merge*. De forma alheia, o *Twitter* bloqueou o acesso ao seu API gratuito, o que impossibilitou por completo a obtenção de *tweets* de forma automática e célere, tendo prejudicado a completude do conjunto de dados utilizado, mas não obstante foi possível através dos *tweets*

obtidos anteriormente à implementação desta limitação, permitindo construir um conjunto de dados coerentes e suficientes para o estudo a desenvolver, embora não de forma perfeita, uma vez que se gostaria de ter em maior quantidade e maior variedade.

Houve também a tentativa de utilizar um *script* de Selenium para continuar a obtenção dos dados após haver o bloqueio do API gratuito do *Twitter*, mas não foi suficiente para o objetivo de aumentar o conjunto de dados.

2.2 Tratamento dos dados

De modo a proceder à limpeza do texto dos tweets, é definida uma função chamada "cleantxt" que utiliza expressões regulares e substituições para remover elementos indesejados, como menções a utilizadores, símbolos # e links. Também são realizadas transformações, como a conversão para letras minúsculas e a remoção de caracteres não ASCII.

Essa função é aplicada à coluna "Tweet" do DataFrame obtido, substituindo os valores pelos textos limpos e posteriormente salvando-o num novo ficheiro com tweets limpos.

Foi realizado pré-processamento dos tweets fazendo uso de recursos da biblioteca "nltk", como tokenização, lista de *stopwords* e *stemmer*.

É definida uma função que realiza o pré-processamento de um tweet de entrada, incluindo a tokenização, remoção de stopwords e caracteres especiais, e aplicação do *stemmer*. Essa função é aplicada à coluna "Tweet" do DataFrame, criando uma nova coluna chamada "Tweet_NLTK" com os tweets pré-processados.

2.3 Obtenção do sentimento

TextBlob O TextBlob pode ser utilizado na análise de sentimentos de criptomoedas, permitindo avaliar o sentimento dos *tweets* relacionados a essas moedas digitais. Por exemplo, ao recolher *tweets* que mencionam uma criptomoeda específica, é possível usar o TextBlob para identificar a polaridade dos *tweets* e entender a perceção das pessoas face a essa moeda.

Através da sua utilização, foi retornado um tuplo (polaridade, subjetividade) onde a polaridade é uma flutuação dentro do intervalo [-1,0, 1,0] e a subjetividade é uma flutuação dentro do intervalo [0,0, 1,0] onde 0,0 é muito objetivo e 1,0 é muito subjetivo.

Após a obtenção de resultados, recolheu-se uma amostra de tweets e avaliou-se a precisão deste modelo, tendo-se chegado à conclusão de que os valores obtidos não atingiam um grau de satisfação suficiente, tendo-se optado por descartar este modelo. O facto de os resultados serem insatisfatórios poderá ser explicado pelo facto de o TextBlob não ser uma biblioteca voltada para a análise de texto informal como o apresentado em tweets, tornando-se assim inadequado ao caso em estudo.

Vader O VADER (Valence Aware Dictionary and sEntiment Reasoner) é uma ferramenta específica para a análise de sentimentos em texto expressos em redes sociais, que utiliza um dicionário de palavras com pontuações de sentimentos positivos, negativos e neutros.

O VADER leva em conta diversos fatores como negações, adjetivos e pontuação, CAPSLOCK, emojis, intensificadores e siglas (*p.e.* "lol", "tbh", "imo", ...). Deste modo, resulta em uma pontuação entre -1,0 (valor mais negativo) e 1,0 (valor mais positivo), sendo o valor 0 neutro. Utilizou-se então esta análise de sentimento dos *tweets* para calcular uma pontuação que representará a importância de cada *tweet*, permitindo avançar para os passos seguintes.

RoBERTa Outro método utilizado para análise de sentimentos foi o modelo RoBERTa, sendo que a versão usada foi pré-treinada especificamente para análise de sentimentos em tweets (em cerca de 58M de tweets), denominado "*cardiffnlp/twitter-roberta-base-sentiment*". O RoBERTa é um modelo de linguagem baseado em transformers, desenvolvido pela Hugging Face, que obteve resultados de ponta em várias tarefas de processamento de linguagem natural, sendo bastante

reconhecido, o que levou ao seu uso no presente projeto prático. De modo a obter o *score* correspondente a cada tweet, é aplicada uma função softmax, que permite obter um valor normalizado na pontuação de saída. Este modelo fornece um output através de um dicionário com os valores associados aos 3 tipos de sentimentos, sendo que, de modo a obter um resultado no mesmo formato e intervalo que os outros métodos utilizados, se calculou através da subtração do valor positivo pelo negativo.

OpenAI O último método a ser utilizado para obtenção de sentimentos foi através da API da OpenAI, sendo utilizado o *text-davinci-003* que é um LLM. Deste modo, tal como seria feito no próprio ChatGPT, é necessário criar um *prompt* que seja suficientemente eficaz, permitindo obter a resposta pretendida. Para tal, foi criado o seguinte prompt, ao qual foi acrescentado cada um dos tweets e enviado através da API:

"The following tweet is related to ethereum price variation, where people can face a bullish or a bearish attitude/sentiment of them. Do a sentiment analysis of the tweet, by returning ONLY the value for the tweet. Remember the whole phrase is just one tweet, which you will only give me the value of the sentiment analysis. Note that the values should range from -1 to 1, meaning that -1 is very negative and 1 is very positive, and return JUST the value of this tweet:"

Mesmo sendo expressamente dito para que fosse enviada a resposta apenas com o valor, nem sempre era assim, pelo que este problema foi tratado por meio de expressões regulares.

3 Dados Históricos da Criptomoeda

De maneira a ser possível verificar se o sentimento obtido através da análise dos *tweets* obtidos foi significativo e correto, é necessário obter dados relativos aos valores históricos da criptomoeda *Ethereum* para o período em análise.

3.1 Obtenção dos dados

Para obter os dados relativos aos preços do *Ethereum* no período que antecedeu o *"The Merge"*, foi utilizado o API da *coingecko*. Começa-se por definir a criptomoeda desejada (*Ethereum*) e, a partir de uma instância da classe *CoinGeckoAPI* interage-se com a API do *CoinGecko* para obter os dados de preços através da função *get_coin_market_chart_by_id*. Estes dados são recolhidos para um período de 300 minutos (5 horas), que define a granularidade dos dados, e armazenados num dicionário onde as chaves são os nomes das criptomoedas e os valores são os carimbos de data/hora e os preços correspondentes.

3.2 Tratamento dos dados

O *DataFrame* obtido é manipulado para incluir informações adicionais, como a data e a hora separadas. Os dados são convertidos em formato de data e hora e são extraídas as informações de data e hora separadamente, assim como são feitas algumas renomeações e agrupamentos de colunas para tornar mais explicativo.

Após o tratamento inicial, o *DataFrame* é reformatado usando o método *melt()*, que transforma as colunas de preços das criptomoedas numa única coluna "value", mantendo as informações de data, hora e nome da criptomoeda. Os dados em falta são removidos usando o método *dropna()*.

4 Resultados

De modo a obter resultados e estabelecer conclusões acerca dos dados obtidos, foram aplicadas diversas técnicas de mineração de dados, sendo efetuada também uma análise comparativa entre os modelos utilizados. Esta análise está presente no ficheiro *analysis.ipynb*. Estão também presentes os ficheiros *2analysis.ipynb* e *3analysis.ipynb*, em que o primeiro contém os dados com um

pré-processamento aplicado aos dados do Vader e o segundo com outro pré-processamento aplicado aos dados do modelo RoBERTa. Contudo, ambos os modelos revelaram piores resultados, pelo que será abordado apenas o ficheiro original.

4.1 Total de valores positivos *vs.* negativos

De modo a obter uma noção da classificação obtida por cada um dos modelos, o gráfico da figura 1 demonstra o balanço entre os valores obtidos. De referir que, por simplicidade, se considerou os valores neutros como positivos, ou seja, os positivos são todo e qualquer valor maior ou igual a zero. Na fig 1, a azul está representado o Vader, a laranja o RoBERTa e a verde o OpenAI.

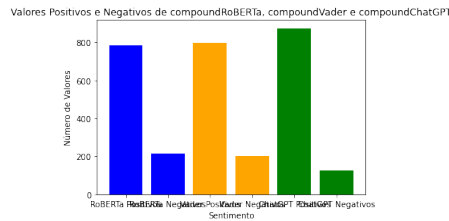


Figura 1: Balanço entre valores positivos e negativos de cada um dos modelos

Comparando o número de casos com valores superiores ou iguais a zero com os valores inferiores a zero, obteve-se um rácio de 798 para 202 com o Vader, 784 para 216 para o RoBERTa e 875 para 125 para o OpenAI, respetivamente.

4.2 Rácio bullish

O rácio Bullish de um tweet é calculado pelo número de tweets positivos relativamente ao número de tweets negativos, sendo que um *bullish ratio* superior a 1 indica a crença do público de que o retorno da criptomoeda aumentará. Caso contrário, um valor menor que 1 mostra maior negatividade em relação à criptomoeda e o mercado apresentará uma regressão em um futuro próximo. Deste modo, obteve-se uma razão de tweets positivos para negativos com o Vader de 3.950 e de 3.629 com o RoBERTa e, surpreendentemente, de 7.0 para o OpenAI, o que será analisado nas restantes análises se irá ou não interferir no resultado global.

4.3 Variação do número de likes e de retweets

Muitas vezes, a interação com um *tweet* pode significar também os sentimentos dos outros utilizadores para com o *tweet*, pelo que se verificou a possibilidade de desenvolver uma métrica que permitisse contabilizar essas interações. Através da figura 2 verificou-se que estavam presentes estas características, pelo que se desenvolveu um algoritmo que atribuía um peso a cada uma das interações num *tweet*.

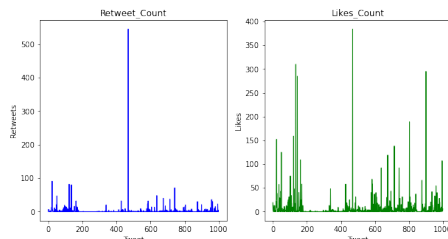


Figura 2: Variação de likes e retweets pelos tweets

Após testar os valores de sentimento com pesos, verificou-se que não se obteve qualquer vantagem, o que se poderá explicar pelo facto de, muitas das vezes, o grande número de interações estar relacionado com a quantidade de seguidores de um utilizador, assunto com que o tweet se relaciona ou até mesmo *aohype* associado ao evento, e não com a sua assertividade ou fiabilidade, pelo que se descartou esta hipótese.

4.4 Variação de cada modelo

De modo a analisar as variações do valor associado ao sentimento em cada um dos modelos, elaborou-se o gráfico da figura 3, que revelou resultados em gamas mais neutras para o Vader e o TextBlob, mas uma maior distribuição pela gama de valores por parte do roBERTa e do OpenAI.

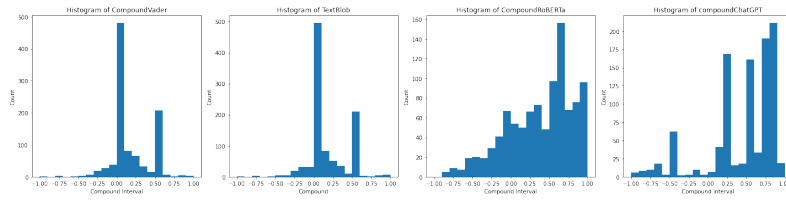


Figura 3: Variação do compound o Vader, TextBlob, RoBERTa e OpenAI

4.5 Diferença entre sentimentos resultantes de cada modelo

Uma vez que foram utilizados dois modelos de obtenção de sentimentos, foi desenvolvido um gráfico que diz respeito à variação de ambos os modelos para cada um dos tweets. De referir que os tweets foram ordenados pelo *compoundVader*, pelo que se explica a linha decrescente no mesmo.

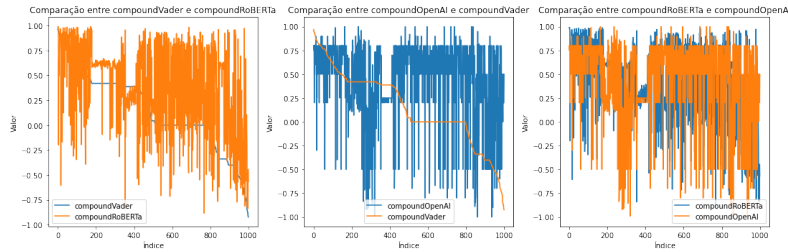


Figura 4: Comparação da variação dos valores de compound do Vader, roBERTa e OpenAI

Pela análise da figura 4, é possível verificar que não há uma relação visível entre os valores de qualquer um dos modelos, pelo que se torna difícil de compreender qual o mais correto, o que será abordado mais à frente.

4.6 Relação entre os gráficos de sentimento e os valores da criptomoeda

De modo a comparar os dados obtidos por ambos os modelos utilizados com os valores reais de ETH, serão descritos nas subsecções seguintes as várias análises efetuadas, juntamente com a sua análise. Deste modo, serão avaliados ambos os gráficos sobrepostos com diferentes variações temporais, sendo também descritas algumas das métricas aplicadas, que permitem estabelecer conclusões e previsões.

4.7 Comparação sem atraso

Tal como referido acima, foi efetuada uma análise comparativa entre os valores de *compound* obtidos e os valores históricos que o Ethereum alcançou. Assim, realizou-se uma agregação diária de ambos os valores para que fosse possível estabelecer a comparação referida, tal como demonstrada nas figuras seguintes.

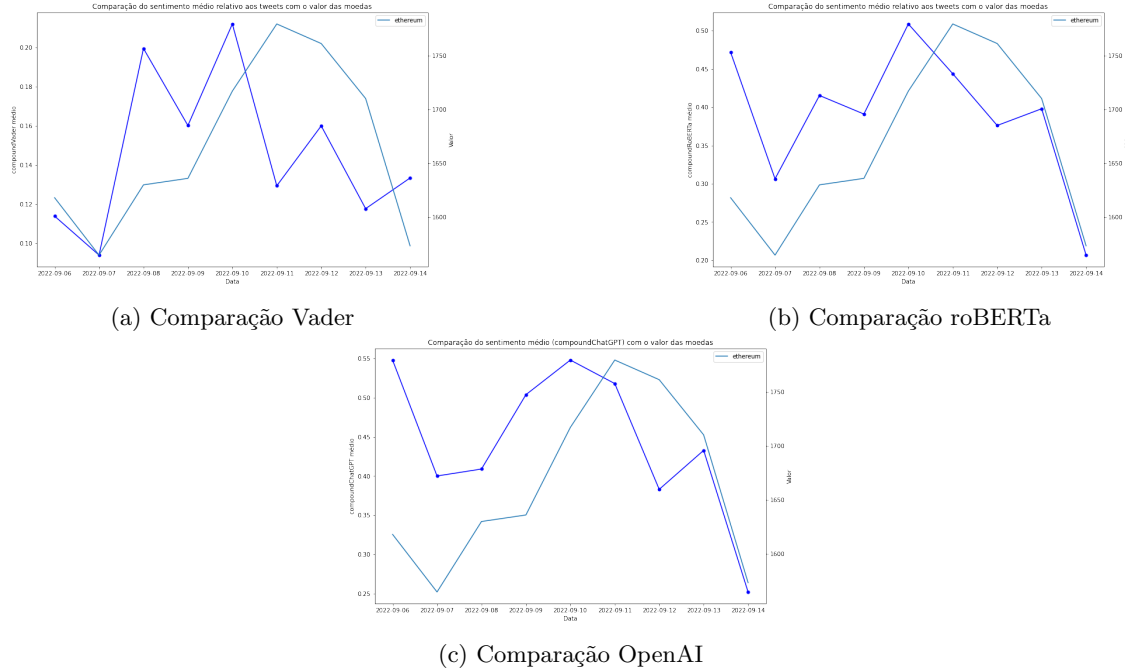


Figura 5: Comparação dos valores de sentimento com os valores reais de criptomoeda

Vader Pela análise da figura 5a, foi possível verificar uma relação significativa entre as variáveis, nomeadamente no mínimo global do dia 7, a subida e posterior descida. Porém, verificou-se um atraso em valores de sentimento face aos valores da criptomoeda, pelo que se resolveu aplicar um atraso nos dados de sentimento, que será abordado em maior detalhe em 4.8 e 4.9.

Por forma a analisar a relação estatística entre as variáveis, calculou-se a correlação, tendo-se obtido 0,29. Embora não seja uma correlação extremamente forte, ainda sugere que existe uma associação entre a percepção expressa nos tweets e o comportamento da criptomoeda. Esta relação pode indicar que as opiniões e os sentimentos expressos pelos utilizadores das redes sociais podem funcionar como fatores indicativos, em certa medida, para a demanda e o valor da criptomoeda no mercado.

RoBERTa De igual modo se fez para os valores obtidos pelo *transformer* da HuggingFace, tendo-se obtido o gráfico da figura 5b.

Novamente, é possível analisar uma determinada tendência para o gráfico de sentimento acompanhar o gráfico dos valores reais, sendo que, apesar de não haver uma correspondência nos mínimos, tal como aconteceu no Vader, há claramente uma maior proximidade entre as linhas. Assim, pelo cálculo da correlação, obteve-se um valor de 0.56. Deste modo, obteve-se uma correlação ainda mais forte, dando até aqui superioridade a este modelo (no caso em estudo), mesmo com a grande discrepância de valores de sentimento analisada em 4.5.

OpenAI De modo análogo ao referido nos anteriores, foi verificada uma certa tendência na figura 5c para acompanhar os valores reais, sendo notório um ligeiro atraso, levando novamente à secção seguinte, em que é feita uma análise com esta variante. Com este gráfico, obteve-se uma correlação de 0.40, o que demonstra um valor bastante competente e que apesar de ser mais baixa de a do RoBERTa, demonstra ainda um valor forte para o caso.

4.8 Comparação com atraso de 1 dia

Uma vez que se verificou um *delay* nos gráficos, sobretudo no Vader, decidiu-se aplicar uma deslocação nos valores de sentimento, ou seja, testar se os utilizadores tendem a embarcar em *trends* na variação da criptomoeda e, assim, adquirir uma atitude positiva ou negativa que corresponda às alterações dos valores observados num determinado dia.

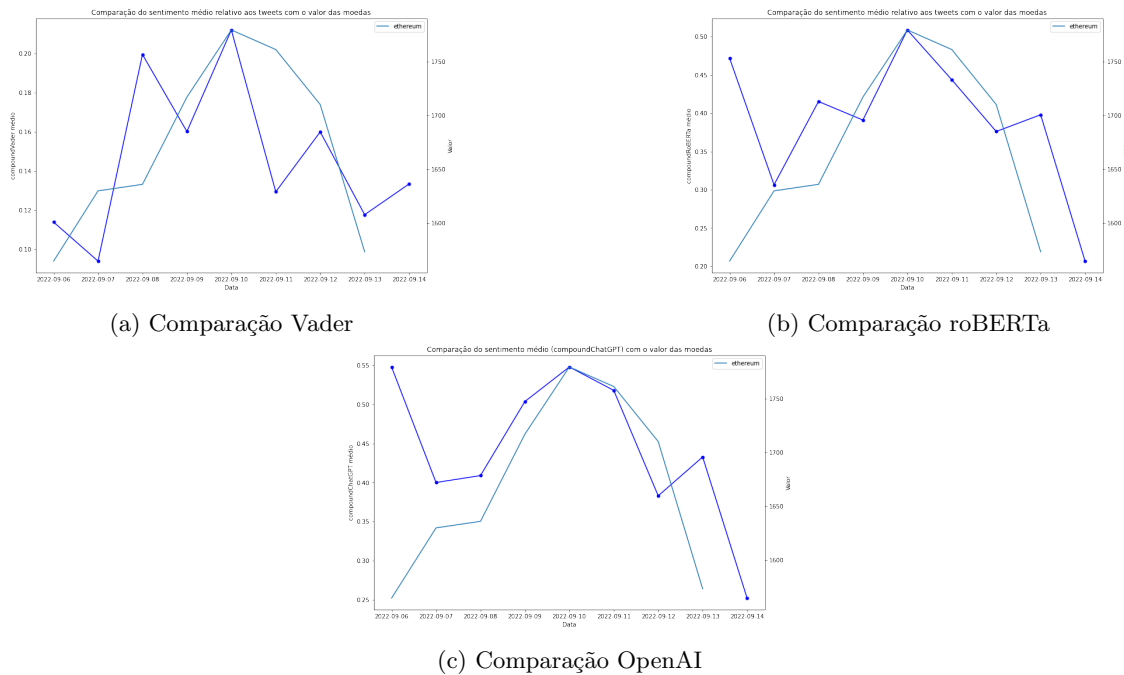


Figura6: Comparação dos valores de sentimento com os valores reais de criptomoeda com atraso de 1 dia

Vader Tal como referido acima, o gráfico 6a demonstra os resultados obtidos com a aplicação de um *shift* de um dia (i.e. atrasar um dia os sentimentos em relação ao valor monetário) no valor obtido através da análise dos tweets.

Através da análise da figura, é possível verificar que esta iniciativa resultou em uma maior similaridade entre as linhas do gráfico, sendo que através do cálculo da correlação se obteve um valor de correlação de 0.55, sendo este bastante superior e aproximando-se da marca obtida pelo RoBERTa em 5b.

RoBERTa Relativamente ao RoBERTa (figura 6b), verificou-se se a sua correlação poderia melhorar. Contudo, ao aplicar este atraso, não foi possível verificar qualquer melhoria, tendo-se obtido um decréscimo no valor de correlação para 0.29.

Com isto, verifica-se que o RoBERTa foi suficientemente capaz de extrair valores de sentimento suficientemente indicativos, por forma a indicar a sua melhor previsão sem qualquer *delay*, ao contrário do Vader que obrigou a um atraso de 1 dia para obter um valor similar a este.

OpenAI Já o OpenAI (figura 6c), verificou-se de igual modo que coincidia com o pico no dia 10, podendo assumir-se que este se revela o dia mais previsível, uma vez que todos os modelos conseguiram efetuar a melhor previsão neste dia. Com isto, obteve-se um valor de correlação de 0.29, igualando o RoBERTa. Contudo, é possível verificar que o gráfico corresponde bastante bem ao dos valores reais, mas, uma vez que o primeiro e último dia se encontram bastante dispares, baixa significativamente a correlação.

4.9 Comparação com atraso de 2 dias

De igual modo, resolveu-se verificar o que poderá ocorrer com um atraso de dois dias, tendo sido então aplicado e analisados os resultados que se obteve com o mesmo.

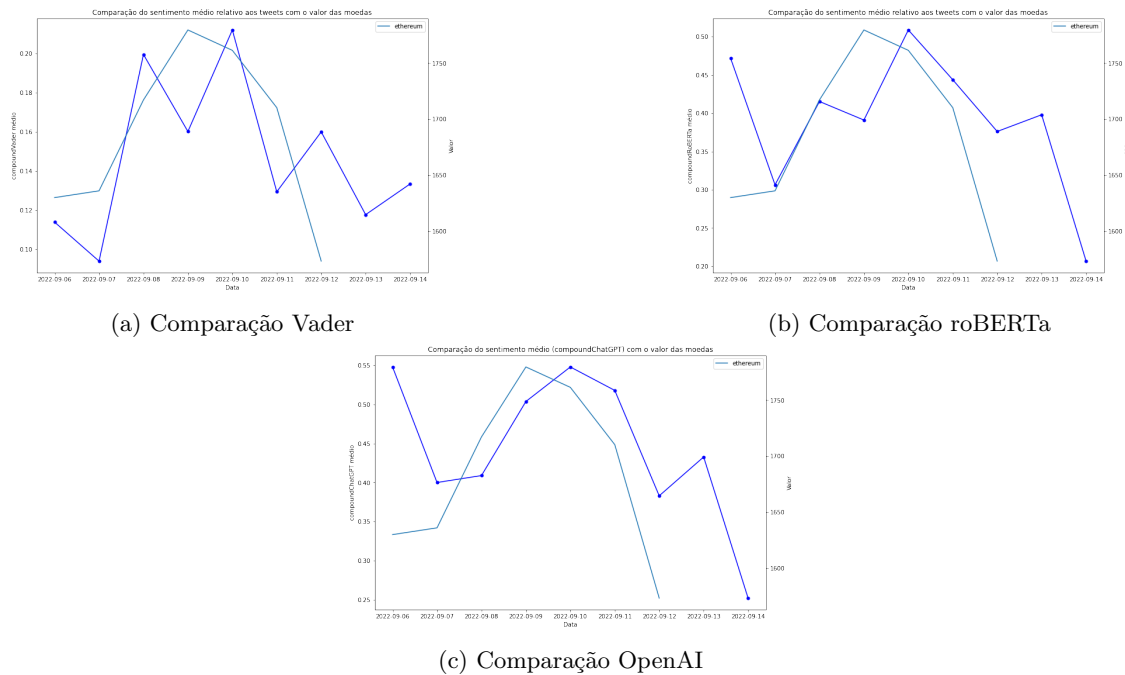


Figura 7: Comparação dos valores de sentimento com os valores reais de criptomoeda com atraso de 2 dias

Vader Analogamente, foi possível manter uma correlação positiva, o que revela um indicador de que realmente esta técnica é eficaz, mas, por meio da análise, os resultados não conseguiram acompanhar o anterior.

Através da análise das linhas do gráfico, assim como a análise dos valores de correlação, foi possível verificar que se reduziu a correlação para 0.53, sendo este um valor bastante positivo e que indica uma correlação minimamente forte.

RoBERTa Já com a utilização do *transformer* da *HuggingFace*, obteve-se uma correlação de 0.40, sendo um valor relativamente alto e que poderá revelar um bom indicativo, apesar de não ser o maior valor obtido através do RoBERTa.

OpenAI Através da análise da figura 7c, é possível verificar uma maior semelhança entre as linhas, o que levou à maior correlação obtida para este atraso, obtendo-se 0.52. Com isto, foi

possível verificar (no caso em estudo) que cada modelo foi melhor em um dos dias, impossibilitando a escolha de um mecanismo de obtenção de sentimentos em concreto como sendo o melhor.

Analisando uma média dos valores obtidos em cada um dos dias, é possível verificar que este atraso de dois dias se revelou o que maiores valores de correlação ofereceu, tendo sido bastante eficaz neste estudo. Porém, não se considera este atraso muito confiável para criptomoedas que sejam caracterizadas pela sua volatilidade, ou seja, que sofram uma oscilação constante nos valores, uma vez que o atraso de dois dias poderá ser extremamente tardio nestes casos. No entanto, sendo a criptomoeda em estudo (*Ethereum*) bastante estável relativamente a outras como ADA, DOGE, etc., o atraso de 2 dias poderá ser uma alternativa eficaz, uma vez que os valores de ambos os modelos se revelaram altos, o que conduz à maior média dos vários atrasos testados.

4.10 Regressão Linear

Dep. Variable:	Value	R-squared:	0.083
Model:	OLS	Adj. R-squared:	-0.049
Method:	Least Squares	F-statistic:	0.6298
Date:	Tue, 06 Jun 2023	Prob (F-statistic):	0.453
Time:	14:51:49	Log-Likelihood:	-51.180
No. Observations:	9	AIC:	106.4
DF Residuals:	7	BIC:	106.8
DF Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	1581.6892	109.011	14.509	0.000	1323.019	1839.459
compoundVader	571.6191	720.290	0.794	0.453	-1131.595	2274.833

(a) Resultados Regressão Linear Vader

Dep. Variable:	Value	R-squared:	0.311
Model:	OLS	Adj. R-squared:	0.212
Method:	Least Squares	F-statistic:	3.155
Date:	Tue, 06 Jun 2023	Prob (F-statistic):	0.119
Time:	14:51:49	Log-Likelihood:	-49.893
No. Observations:	9	AIC:	103.8
DF Residuals:	7	BIC:	104.2
DF Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	1474.6169	109.978	13.408	0.000	1214.561	1734.673
compoundRoBERTa	488.4060	274.950	1.776	0.119	-161.748	1138.560

(b) Resultados Reg. Linear RoBERTa

Dep. Variable:	Value	R-squared:	0.157
Model:	OLS	Adj. R-squared:	0.036
Method:	Least Squares	F-statistic:	1.380
Date:	Wed, 14 Jun 2023	Prob (F-statistic):	0.292
Time:	22:29:38	Log-Likelihood:	-50.081
No. Observations:	9	AIC:	105.6
DF Residuals:	7	BIC:	106.0
DF Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	1521.7377	128.718	11.822	0.000	1217.367	1826.109
compoundChatGPT	323.9096	284.153	1.140	0.292	-347.924	995.905

(c) Resultados Regressão Linear OpenAI

Figura 8: Comparação dos resultados obtidos através da aplicação de uma Regressão Linear a cada um dos modelos

Vader Uma vez que os valores obtidos anteriormente revelaram interesse, recorreu-se a uma Regressão Linear como forma de avaliar se estes valores são significativos e, nesse caso, o quão significativos são. Assim, apresentam-se nas figuras 8a e 8b os resultados obtidos para cada um dos meios utilizados.

Pela análise da figura 8a, é importante referir que o p-value obtido é de 0.45, indicando que não há evidências estatisticamente significativas para rejeitar a hipótese nula de que não há relação entre os valores de sentimento dos tweets e o desempenho da criptomoeda. Deste modo, mesmo com a análise efetuada até aqui, não há evidências suficientes para afirmar que os valores de sentimento dos tweets têm um impacto estatisticamente significativo no comportamento da criptomoeda.

O coeficiente de determinação R-quadrado (R-Squared) mede a proporção da variabilidade na variável dependente que pode ser explicada pela variável independente. Neste caso, o R-quadrado é de 0.083, o que indica que aproximadamente 8.3% da variabilidade nos valores pode ser explicada pela variável compoundVader, considerando-se que apresenta uma relação fraca entre os sentimentos dos tweets (representados por compoundVader) e os valores da criptomoeda.

Já o F-statistic é uma medida estatística que testa a significância geral do modelo de regressão que obteve um valor de 0.6298. Este valor indica que o modelo tem baixo poder preditivo, e a probabilidade associada a esse valor (Prob (F-statistic)) é de 0.453, indicando que não há evidências estatisticamente significativas para afirmar que o modelo é significativo para o problema em questão.

RoBERTa De igual modo, foi efetuado para o RoBERTa, apresentando-se os resultados na figura 8b.

Novamente, é possível verificar que se obteve um *p-value* de 0.1, sendo que ainda não é possível afirmar com alta confiança que há uma relação estatisticamente significativa entre os valores de sentimento dos tweets e o desempenho da criptomoeda. No entanto, um *p-value* de 0.1 indica uma evidência um pouco mais forte em comparação com o *p-value* mais alto obtido anteriormente.

Neste caso, embora o valor não seja considerado estatisticamente significativo (0.05), sugere uma tendência ou uma possível relação entre os sentimentos dos tweets e o desempenho da criptomoeda, podendo depender do contexto específico e da importância prática do estudo. É importante observar que o valor não fornece informações sobre a magnitude ou a direção da relação encontrada. Portanto, mesmo que a correlação seja estatisticamente significativa, indica apenas uma evidência leve em favor da hipótese alternativa de que existe uma relação entre os sentimentos dos tweets e o desempenho da criptomoeda.

Neste caso, o R-quadrado é de 0.311, o que significa que aproximadamente 31.1% da variabilidade nos valores pode ser explicada pela variável compoundRoBERTa. Assim, indica uma relação moderada entre os sentimentos de tweets (representados por compoundRoBERTa) e os valores da criptomoeda. O valor de *F-Statistic* obtido é de 3.155, pelo que indica que o modelo possui maior poder preditivo, reafirmando a superioridade do RoBERTa.

OpenAI Por fim, relativamente aos resultados obtidos com o OpenAI presentes na figura 8c, é possível verificar que se obteve um *p-value* de 0.29, sendo novamente indicio de que não é suficientemente baixo para que seja possível verificar algo significativo na relação entre as duas variáveis em questão. Relativamente ao R-Squared, obteve-se um valor de 0.157, o que apresenta uma taxa bastante baixa dos valores reais que podem ser explicados pelo valor de sentimento, verificando-se bastante abaixo do valor obtido com o RoBERTa, mas significativamente acima do valor obtido com o Vader.

4.11 Previsão de Valores

Com o uso da regressão linear, criou-se uma tabela para cada modelo, que demonstra os resultados da comparação entre os valores reais e os valores previstos obtidos por meio desta técnica estatística, através do "*y predicted*". Cada linha da tabela representa uma observação correspondente a cada dia, onde são mostrados o valor real, o valor previsto e o erro percentual correspondente.

Real	Predicted	Error (%)
1617.9354554854335	1646.7531817063293	1.78%
1564.5589540302462	1635.355964290167	4.53%
1629.8832311636493	1695.7507569263455	4.04%
1635.9321760130483	1673.2301776693762	2.28%
1717.0769569840472	1702.916282543173	0.82%
1779.5190288833046	1655.6154371974646	6.96%
1761.3173558806091	1673.1298917706722	5.01%
1710.1642908515214	1648.934386929802	3.58%
1573.205213695171	1657.9089900508545	5.38%
Média do erro: 3.82%		

(a) Tabela Vader e erro associado

Real	Predicted	Error (%)
1617.9354554854335	1705.0276318967715	5.38%
1564.5589540302462	1624.2941551477675	3.82%
1629.8832311636493	1677.4351286406038	2.92%
1635.9321760130483	1665.5935657580956	1.81%
1717.0769569840472	1723.0473530684103	0.35%
1779.5190288833046	1691.2955502278091	4.96%
1761.3173558806091	1658.2990185300205	5.85%
1710.1642908515214	1668.9663918224835	2.41%
1573.205213695171	1575.633867895665	0.15%
Média do erro: 3.07%		

(b) Tabela RoBERTa e erro associado

Real	Predicted	Error (%)
1617.9354554854335	1699.2299799553884	5.02%
1564.5589540302462	1651.3259468285519	5.55%
1629.8832311636493	1654.2045298954122	1.49%
1635.9321760130483	1684.9416680461345	3.00%
1717.0769569840472	1699.2654022669926	1.04%
1779.5190288833046	1689.5368010712493	5.06%
1761.3173558806091	1645.803368234546	6.56%
1710.1642908515214	1661.9201534991678	2.82%
1573.205213695171	1603.3648131896814	1.92%
Média do erro: 3.61%		

(c) Tabela OpenAI e erro associado

Vader Analisando os resultados da tabela 9a, observa-se que a maioria dos erros percentuais é relativamente baixa, variando de 0,82% a 6,96%. Isso significa que, em geral, as previsões da regressão linear estão relativamente próximas dos valores reais. No entanto, é importante considerar que há algumas variações entre as previsões e os valores reais, como evidenciado pelos erros percentuais positivos.

A média do erro, que é calculada como a média dos erros percentuais, indica uma tendência de as previsões estarem, em média, com uma diferença de aproximadamente 3,82% em relação aos valores reais. É importante ter em conta que a precisão das previsões pode variar dependendo do tipo dos dados, das características do modelo de regressão utilizado, entre outros fatores relevantes para a análise.

RoBERTa Na tabela 9b, são demonstrados os resultados da previsão utilizando o modelo Transformer RoBERTa da Huggingface para analisar os sentimentos de tweets e obter previsões o mais precisas possível sobre os valores de uma criptomoeda.

Ao comparar os valores reais com os valores previstos, é possível observar que os erros percentuais variam de 0,15% a 5,85%, indicando que o modelo RoBERTa apresenta uma capacidade satisfatória de previsão, com a média dos erros percentuais situando-se em 3,07%.

Deste modo, obteve-se valores ligeiramente melhores com esta abordagem, sendo ainda mais baixos e revelando um resultado bastante positivo e promissor para o conjunto de dados analisado.

OpenAI Já com a API da OpenAI, obteve-se uma média de erro de 3,61%, o que reflete um resultado intermédio entre os dois referidos anteriormente, mas todos dentro dos 3%. Verifica-se também que os valores de erro diários são sempre superiores a 1,04%, ao contrário dos outros que conseguiram verificar alguns dos valores dentro dos 0%.

4.12 Testes de Causalidade

Uma vez que correlação não implica causalidade, foi aplicado o teste de causalidade de Granger, que visa superar as limitações do uso de simples correlações entre variáveis. Este teste procura determinar o sentido causal entre duas variáveis, estipulando que X "Granger-causa" Y se valores passados de X ajudam a prever o valor presente de Y.

```
Granger Causality
number of lags (no zero) 1
ssr based F test:      F=1.6989 , p=0.2492 , df_denom=5, df_num=1
ssr based chi2 test:   chi2=2.7183 , p=0.0992 , df=1
likelihood ratio test: chi2=2.3401 , p=0.1261 , df=1
parameter F test:      F=1.6989 , p=0.2492 , df_denom=5, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:      F=0.2846 , p=0.7784 , df_denom=2, df_num=2
ssr based chi2 test:   chi2=1.9923 , p=0.3693 , df=2
likelihood ratio test: chi2=1.7532 , p=0.4162 , df=2
parameter F test:      F=0.2846 , p=0.7784 , df_denom=2, df_num=2
```

Figura 10: Testes de causalidade aplicado aos resultados do Vader

```
Granger Causality
number of lags (no zero) 1
ssr based F test:      F=0.0443 , p=0.8417 , df_denom=5, df_num=1
ssr based chi2 test:   chi2=0.0708 , p=0.7901 , df=1
likelihood ratio test: chi2=0.0705 , p=0.7906 , df=1
parameter F test:      F=0.0443 , p=0.8417 , df_denom=5, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:      F=1.2957 , p=0.4356 , df_denom=2, df_num=2
ssr based chi2 test:   chi2=9.0698 , p=0.0107 , df=2
likelihood ratio test: chi2=5.8172 , p=0.0546 , df=2
parameter F test:      F=1.2957 , p=0.4356 , df_denom=2, df_num=2
```

Figura 11: Testes de causalidade aplicado aos resultados do RoBERTa

```
Granger Causality
number of lags (no zero) 1
ssr based F test:      F=0.3878 , p=0.5607 , df_denom=5, df_num=1
ssr based chi2 test:   chi2=0.6206 , p=0.4308 , df=1
likelihood ratio test: chi2=0.5977 , p=0.4395 , df=1
parameter F test:      F=0.3878 , p=0.5607 , df_denom=5, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:      F=1.5421 , p=0.3934 , df_denom=2, df_num=2
ssr based chi2 test:   chi2=10.7947 , p=0.0045 , df=2
likelihood ratio test: chi2=6.5309 , p=0.0382 , df=2
parameter F test:      F=1.5421 , p=0.3934 , df_denom=2, df_num=2
```

Figura 12: Testes de causalidade aplicado aos resultados do OpenAI

Nas figuras 10, 11 e 12 são apresentados os resultados obtidos com este teste. Com os testes efetuados face ao Vader, não foram encontrados quaisquer resultados significativos em nenhum dos lags, mas no entanto, o p-value para o *ssr based chi2 test* está próximo do limite de 0.05, sugerindo a possibilidade de uma relação causal fraca.

Já o RoBERTa não verificou qualquer evidência significativa para 1 lag, mas para o caso de 2 lags o valor de F aumentou em relação ao teste com 1 lag. O *ssr based chi2 test* mostra um valor de chi2 de 10.7947 e um p-value de 0.0045, indicando que o modelo com 2 lags é estatisticamente significativo, ao permitir rejeitar a Hipótese Nula. O *likelihood ratio test* também indica significância estatística, com um valor de chi2 de 6.5309 e um p-value de 0.0382. Portanto, ao considerar 2 lags, os testes de causalidade de Granger encontram evidências de uma possível relação causal entre a análise de sentimentos de Tweets e os valores da criptomoeda Ethereum.

Por fim, o OpenAI não encontrou qualquer resultado significativo para 1 lag, mas com 2 lags obteve-se no *ssr based chi2 test* um valor de chi2 de 10.7947 e um p-value de 0.0045, que revela significância. Também o *likelihood ratio test* indica significância estatística, com um valor de chi2 de 6.5309 e um p-value de 0.0382. A inclusão de lags revela-se importante, uma vez que permite capturar o efeito do tempo no problema em questão. A ideia é que a informação contida nos valores passados de X possa ajudar a prever Y. Portanto, ao adicionar lags de X no modelo, investiga-se se esses *lags* têm poder preditivo sobre Y além do que já é explicado pelas próprias defasagens de Y.

5 Conclusão

Com a conclusão do presente projeto prático, considera-se que foi possível atingir com distinção os objetivos propostos, envolvendo a recolha, processamento, análise e mineração de dados publicamente acessíveis com o objetivo de extrair conhecimento útil e não óbvio ao integrar dados de duas fontes distintas.

Ao longo do desenvolvimento, foi possível superar algumas dificuldades, nomeadamente o uso e manipulação das ferramentas de obtenção e mineração, mas, pelo outro lado, não foi possível enfrentar as limitações da API do Twitter, que limitam o conjunto de dados, impedindo a obtenção e corroboração de resultados através de aplicação, por exemplo, em outras criptomoedas ou até mesmo outros períodos de tempo. No entanto, foi possível lograr o objetivo do projeto, sendo este relacionado com a análise da possibilidade de efetuar previsões através de tweets. Apesar de alguns dos valores não serem significativos, fenómeno este que ocorre frequentemente em análise de DataMining, conseguiu-se obter resultados significativos que permitem evidenciar esta relação e verificar indícios positivos de que é possível prever as variações em criptomoedas (neste caso no ETH) com recurso a sentimentos de tweets.

Em suma, considera-se bastante satisfatório o resultado obtido, tendo permitido consolidar os conhecimentos obtidos na UC de Mineração de Dados ao longo do semestre, assim como aplicar e expandir os mesmos em uma temática atual e que poderá ser útil quer a nível pessoal como para um futuro estudo na área.