UNIVERSITÀ DI PISA

Text Analytics

Bitcoin tweets analysis

Ferri Lorenzo (607828)
email l.ferri11@studenti.unipi.it

Pappolla Roberta (534109)
email r.pappolla@studenti.unipi.it

Text Analytics (635AA), Academic Year 2020/2021
Professore Andrea Esuli

Contents

1 Introduction

Il progetto è sito nella repository GitHub "lorenzoFerri95/TextAnalytics". All'interno sono presenti alcuni files più la cartella "app", che generano la pagina web, e un'altra cartella, "research", con all'interno i notebooks e files più rilevanti per il progetto, quelli delle analisi e della ricerca. L'argomento trattato riguarda il monitoraggio e l'analisi del sentimento e dei topic rilevanti discussi dagli utenti di Twitter sulle Criptovalute e, più precisamente, sul Bitcoin. Uno degli obiettivi principali è estrarre un modello per la predizione dell'andamento del prezzo del Bitcoin a brevissimo termine, come supporto all'attività di Trading Intraday. I dati per le analisi sono stati raccolti da dataset presenti su Kaggle (l'API di Twitter non permette di ottenere grosse moli di dati dal passato), mentre quelli per la pagina web sono feed di dati in tempo reale ottenuti dall'API di Twitter, tramite la libreria Tweepy. L'idea è di simulare un intero processo di Data Science, dalla ricerca alla messa in produzione.

2 Data Preparation & Understanding

In "TextAnalytics/datasets/iniziali" sono presenti i dataset di partenza: i prezzi dei Bitcoin per ogni minuto negli anni 2019 e 2021 e i testi dei tweets in due finestre temporali all'interno di questi due anni. Il progetto è stato orientato con dati al minuto perché il fine è quello di catturare un certo grado di causalità tra le discussioni che avvengono su Twitter e il prezzo nel brevissimo termine del Bitcoin. Per il dataset del 2019 è stato selezionato il periodo nel range "2019-05-06" - "2019-09-27", che è quello dove le fluttuazioni dei Bitcoin sono state più significative. Dopo aver pulito il dataset, eseguito un sample e selezionato solo i tweets in inglese esso contiene 2 milioni e 400 mila righe circa. Qui sotto si riporta il grafico dei prezzi al minuto in quel periodo (tutti i prezzi si riferiscono al prezzo di chiusura del minuto, come da convenzione).

Figure 1: Bitcoin Price 2019

Per il dataset del 2021 è stato usato tutto il periodo possibile, nel range "2021-02-05" - "2021-07-30". Qui sotto si riporta il grafico dei prezzi al minuto in quel periodo.

Figure 2: Bitcoin Price 2021

L'uso di entrambi questi due dataset deriva dal fatto che quello del 2019 ha il vantaggio di contenere molti tweets distribuiti omogeneamente su tutto il periodo, ma ha lo svantaggio di contenere solo l'attributo relativo al testo del tweet. Quello del 2021 invece contiene diversi attributi utili (come il numero di followers dell'utente), ma di contro i dati presenti sono solo addensamenti discontinui lungo il periodo. Inoltre a livello di statistiche testuali è più interessante analizzare i dati più recenti del 2021.

2.1 NLTK Analysis

Dopo aver effettuato una prima osservazione e comprensione dei dati, si è deciso di procedere ad analizzare il testo contenuto nei tweets presenti nei due dataset. Per ognuno dei due anni, prima di svolgere diversi tipi di analisi, si è proceduto con la preparazione dei dati e alla pulizia del testo: divisione del testo in tokens, così da ottenere la lista contenente tutte le parole; eliminazione delle punteggiature e stopword; ridenominazione delle diverse criptovalute riconosciute, poiché molte erano presenti sia con il nome esteso che con il code (es. bitcoin e btc). Dopo diverse prove si è deciso di non effettuare la lemmatizzazione delle parole per ottenerne la radice, perché si è notato che dopo questo processo alcune parole venivano modificate e troncate anche se non necessario, e così facendo non era più possibile riconoscere la parola stessa.

Per avere una visione complessiva di ciò che è contenuto nei tweets e a farsi un'idea di come il pensiero degli utenti di Twitter sia variato nel tempo, si è proceduto effettuando l'analisi delle singole parole, dei bigrammi e dei trigrammi sia per l'intero arco temporale che dividendo lo stesso nei diversi mesi. Per poi, infine, procedere alla produzione delle WordCloud.

2.1.1 Analisi dell'intero arco temporale

La prima cosa che si nota guardando le frequenze dell'intero dataset è che le parole 'BITCOIN' (freq. 948'369) e 'crypto' (freq. 316'221) hanno una frequenza eccessivamente elevata rispetto alla frequenza delle altre parole, la prima parola con la frequenza maggiore che non è il nome di una criptovaluta è 'project', con frequenza 99'105 (dati relativi all'anno 2021). Sebbene questo fenomeno sia normale, per questa analisi si è proceduto all'eliminazione di tutti i nomi di criptovalute riconosciuti (ma solo dopo aver prodotto anche le loro frequenze per averne traccia). Bitcoin, Ethereum, Binance coin, dogecoin, amp, xrp, nft (APENFT), ada (Cardano), bsc (BowsCoin) sono le parole escluse e quì sotto riportiamo le frequenze delle altre parole.

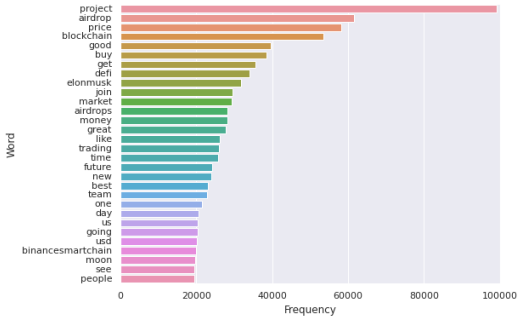

Figure 3: Frequenza parole 2021

Guardando la frequenza delle parole relative al 2021 si può percepire un sentimento generale positivo nel confronto della valuta in questione: nelle prime 30 parole con la maggiore frequenza troviamo parole come "good", "great", "like" e "best". Anche se queste singolarmente non ci dicono molto perché potrebbero essere precedute da un "not" o riferirsi a qualcosa di diverso dalla criptovaluta. Spicca subito all'occhio la seconda parola con la maggior frequenza, "**airdrop**". Ma cosa significa nel mondo delle criptovalute? questa non è altro che un'iniziativa di marketing, che ha il fine di promuovere la consapevolezza della creazione di una nuova valuta virtuale attraverso l'invio di piccole quantità della nuova valuta in questione ai portafogli dei membri attivi della comunità blockchain, gratuitamente o in cambio di un piccolo servizio, come il retweet di un post inviato dalla società che emette la valuta.
Non sorprende invece trovare la parola "**elonmusk**" tra le parole con frequenza maggiore. Nel corso del 2021 Elon Musk in due occasioni è riuscito ad influenzare il valore del Bitcoin: nel marzo 2021 ha comunicato la possibilità di acquistare macchine Tesla con l'uso di Bitcoin facendo così aumentare il suo valore, per poi, solo due mesi dopo, sospendere questa possibilità (con la promessa di un possibile ritorno in un prossimo futuro), provocando così un crollo del valore.
Osservando i bigrammi sembra esserci la conferma del sentimento positivo generale dell'utenza di Twitter. Come è possibile osservare dal grafico fig.4, troviamo nella top 4 delle parole con maggiore frequenza "good project", "great project" e "**hope project**". Si nota inoltre un nome ed un cognome: Michael Saylor, il co-fondatore di MicroStrategy. Questo probabilmente per il fatto che nel maggio/giugno dello stesso anno ha rilasciato diverse interviste a tema criptovalute dopo che la sua società ha investito significative risorse in bitcoin.

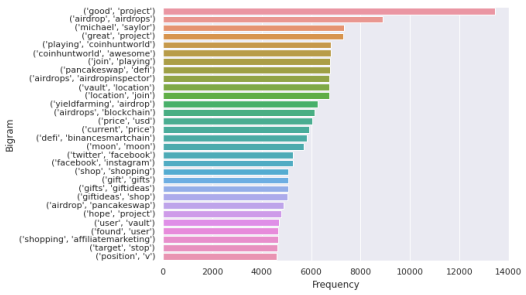

Figure 4: Frequenza bigrammi 2021

Dai trigrammi invece non evince niente di particolarmente interessante o significante.

Figure 5: Word Cloud 2021

L'unica altra cosa che si può notare è l'assenza di una parola tra quelle più frequenti, una che negli ultimi due anni ha influenzato tutti i settori e l'economia di tutti i paesi: "covid". Questo probabilmente perché le criptovalute non hanno subito gli effetti negativi del coronavirus nel 2021, ma ad inizio 2020. In più, come abbiamo già osservato, ci sono state notizie più recenti ed importanti che hanno influenzato il mondo delle criptovalute nel 2021. Per quanto riguarda il 2019 l'analisi non ha portato a nulla di interessante. Le parole con maggiore frequenza sono completamente differenti da quelle viste per il 2021, ad esempio non è presente il token "project", che nel 2021 era la più frequente.

2.1.2 Analisi nel corso dei mesi

Per vedere come variano le parole utilizzate nei tweets nel corso dell'anno il dataset è stato analizzato per i singoli mesi. Come spiegato in precedenza però per l'anno 2021 i tweets non sono ben distribuiti lungo il periodo. Ad esempio il mese di marzo contiene solamente 3'202 righe, mentre il mese di luglio ne ha ben 423'735, differenza questa che renderebbe non significativo il confronto. Quindi si è deciso di confrontare solo i risultati dei mesi di aprile e giugno, che hanno rispettivamente 48'096 e 115'466 righe, nei quali è possibile osservare una grande variazione delle parole più usate.

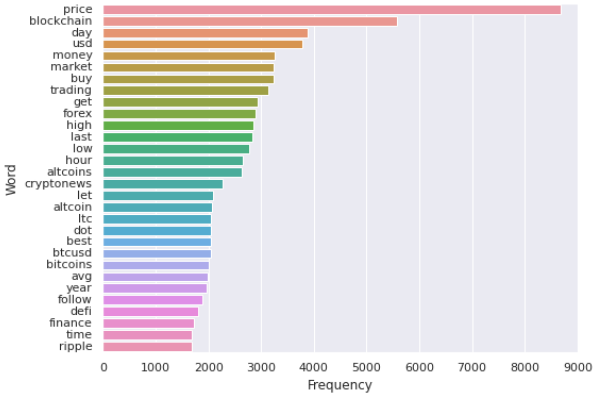

Figure 6: Frequenza parole aprile 2021

Come è possibile vedere dalla figura 6 nel mese di aprile non troviamo nessuna delle parole con frequenza maggiore che abbiamo trovato nell'intero dataset, e sarà così fino al mese di giugno. Possiamo osservare parole tipiche del traiding come "high", "low" e "price", ma nessuna parola riferita al sentimento positivo o negativo. Con una frequenza di quasi 3'000 troviamo la parola "forex", che non è altro che il mercato decentralizzato per lo scambio di valute.

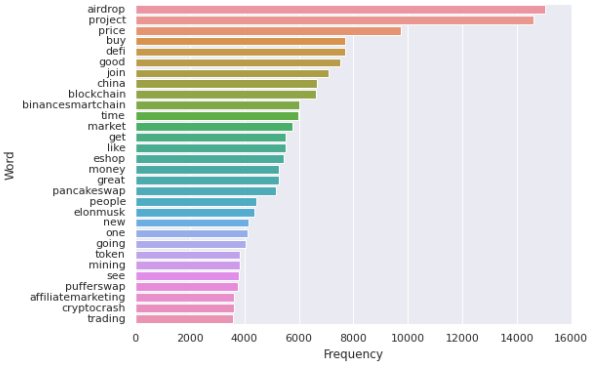

Figure 7: Frequenza parole Giugno 2021

Nel mese di giugno invece comincino a comparire parole di stampo positivo come "good" e "great". In questo mese, e ancora più in quello di luglio, l'argomento "elonmusk" inizia a farsi strada tra quelli più frequenti.
Comunque, poiché i dati dell'anno 2021 sono discontinui e la qualità di essi varia molto

lungo il periodo, non possiamo avere la certezza che la differenza notata tra i testi dei due mesi sia effettiva, oppure influenzata da questi fattori.
Non si è trovato nulla di interessante per quanto riguarda l'anno 2019.

3 Sentiment Analysis

3.1 VADER Sentiment

Nel dataset del 2019 è stata calcolata la Polarity di ogni tweet tramite il modello VADER. Questa è poi stata trasformata in una classe di tre valori secondo i seguenti range: [0.7, 1] diventa sentiment=1 (positivo); [-1; -0.3] diventa sentiment=-1 (negativo); e [-0.3; 0.7] diventa sentiment=0 (neutrale). Il motivo del range più permissivo per il sentimento negativo e più stringente per quello positivo è che nei tweets, per qualsiasi argomento in generale, è sempre presente un Bias verso la neutralità e positività. La negatività è più rara e quindi questa trasformazione garantisce l'equilibrio delle classi: l'11% dei tweets ha classe positiva e il 10% classe negativa, a fronte di un quadruplo di classi positive rispetto alle negative che avremmo se usassimo i range [-1; -0.5], [-0.5; 0.5], [0.5, 1]. Il dataset è poi stato filtrato per contenere solo i tweets polarizzati, escludendo la classe 0 del sentimento neutrale, che non è di interesse.

3.2 Transformers Sentiment

Il dataset così ottenuto, che contiene 519.002 righe, è stato diviso in Training (dati dal 2019-05-06 al 2019-08-20) Validation (dati dal 2019-08-21 al 2019-08-31) e Test (dati dal 2019-09-01 al 2019-09-26), con il testo dei tweets in input e la VADER sentiment in output, al fine di eseguire il Fine Tuning di un modello Transformers per la predizione del sentimento, usando la libreria "simpletransformers". In particolare il modello usato è "siebert/sentiment-roberta-large-english", cioè un modello RoBERTa istruito su una grande mole di dati in inglese del 2019 (tweets, reviews ecc..), per predire la classe binaria del sentimento positivo (1) o negativo (0) (anche nel dataset sono stati trasformati i dati della VADER Sentiment per rispecchiare queste due labels). Il Fine Tuning è stato interrotto dopo poche epoche e il modello risultante è stato salvato e usato per labelizzare il dataset con il sentimento. Il modello ottenuto riesce ad individuare correttamente anche il sentimento di frasi non proprio immediate, come **"Bitcoin to the moon"**, che viene labelizzata come positiva, o **"Bitcoin flop"**, che viene labelizzata come negativa.
Lo stesso procedimento è stato adottato per il dataset del 2021, tranne per il fatto che gli estremi per la trasformazione della polarity in sentiment sono 0.8 per Positive e -0.2 per Negative e il modello Transformers usato per labelizzare il dataset è lo stesso salvato in precedenza.

3.2.1 NLTK Analysis

Al fine di valutare meglio la classificazione del sentiment è stata prodotta un'ulteriore analisi del testo dividendo il dataset in due sulla base delle labels assegnate dal modello RoBERTa. I risultati più interessanti sono stati quelli ottenuti con il dataset del 2021.

Per quanto riguarda i teets labelizzati come positivi, dei quali si può osservare la Word Cloud quì sopra a sinistra, le 30 parole con la maggior frequenza confermano la positività del loro contenuto: tra di esse troviamo "good", "great", "hope", "future", "opportunity" e "sucess". Per quanto riguarda invece i tweets labelizzati con sentiment negativo, di cui si riporta la Word Cloud quì sopra a destra, appare qualcosa di cui era stata notata l'assenza in precedenza: le ultime due posizioni dei primi 10 token con la frequenza maggiore sono occupate da "covid"

e "covidvaccine". Mentre la presenza della prima non rappresenta una grossa sorpresa, per la seconda invece è difficile comprendere il nesso tra il vaccino contro il Coronavirus, le criptovalute e un sentimento negativo. Un'altra sorpresa è la presenza del token "elonmusk". La spiegazione più plausibile è che si tratti delle notizie riguardanti la sospensione della possibilità di acquistare vetture Tesla attraverso l'uso di Bitcoin. Infine troviamo parole con significato evidentemente negativo come "crash", "bad", "fear" "wrong", e "scam" che confermano la buona classificazione del sentiment. L'ultimo token interessante presente tra i tweets labelizzati come negativi è "china". Il motivo potrebbe essere che nel mese di luglio (che rappresenta una grossa fetta dei dati) la Cina ha apertamente dichiarato guerra alle criptovalute. La banca centrale cinese ha chiesto la chiusura di una società che "era sospettata di fornire servizi software per le transazioni di valuta virtuale". Questo porta a pensare che la presenza di questo token all'interno di quei tweets sia corretta.

4 Topic Modeling

Il Topic Modeling consiste nel trovare la lista di tokens che identificano ciascun argomento presente nei testi. Questa analisi è stata effettuata per entrambi i dataset ed ha arricchito il dataset finale per effettuare le predizioni. Sono stati applicati due modelli di LDA (Latent Dirichlet allocation) provenienti dal pacchetto "gensim": il classico LDA e il Mallet, che permette di ottenere una qualità migliore dei topic. Questi modelli necessitano in input di un *Dizionario* di tutte le parole contenute nel dataset, associate al loro relativo id, per poi utilizzarlo per ricavare il *Corpus*. Quest'ultimo è una lista di liste, ognuna delle quali rappresenta una riga del dataset e contiene l'id delle parole del dizionario presenti in quella riga e la loro frequenza, rappresentate come una tupla (id, freq). Si è deciso di non cercare i migliori parametri per il modello attraverso l'uso della RandomizedSearchCV perché per ottenere un singolo modello erano necessarie dalle 6 alle 10 ore. Per valutare i due modelli sono stati calcolati Perplexity e Coherence Score, che riportiamo qui sotto relativi a 10 topic.

Anno		Coherence
2019	0,3054	-9,527
2021	0,2776	-9,436

Per visualizzare i diversi topic e le relative parole chiave individuate dal modello si è utilizzato il il grafico interattivo del pacchetto **pyLDAvis**.

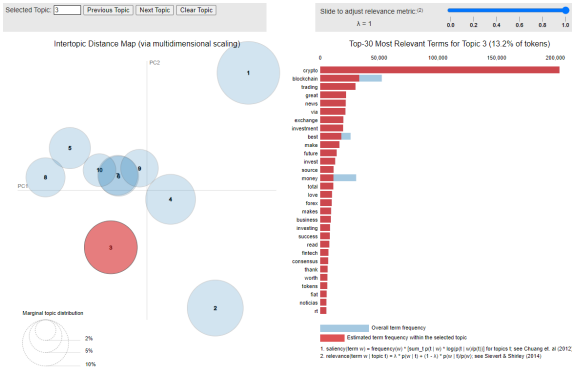

Figure 8: LDAvis, topic 3 2019

La sezione di sinistra ci mostra una visione globale del modello. Quì ogni topic è rappresentato da un cerchio: più questo è grande, più rilevante è quel topic. Le distanze tra i centri dei cerchi sono proporzionali alle distanze tra i topic, che sono calcolate mediante uno scaling multidimensionale per poterle proiettare su due dimensioni. Selezionando un cerchio le parole e le barre sul lato destro si aggiorneranno. Queste parole sono quelle chiave che formano il topic selezionato. Questa rappresentazione è anche molto utile per comprende la bontà del modello, in quanto, nel caso di modello ottimale, avrà cerchi abbastanza grandi e non sovrapposti sparsi per tutto il grafico. Quindi, come possiamo osservare in questo caso, il modello ottenuto non è buono. Infatti i cerchi non sono sparsi ma sono molto concentrati nel secondo quadrante del piano cartesiano. In più è possibile osservare una completa sovrapposizione tra i topic 6 e 7. Il modello ottenuto per l'anno 2019 è peggiore, infatti 8 topic su 10 sono in una metà del quarto quadrante.

4.1 LDA Mallet

Visti gli scarsi risultati si è proceduto applicando la versione Mallet dell'algoritmo LDA (LDAMallet) fornita da Gensim. Essa necessita della creazione di un **"environment variable"** per indicare al computer dove trovare tutti i vari componenti dei suoi processi quando è in esecuzione. Si è cercato di comprendere quale fosse il miglior numero di topic da ricercare all'interno del dataset cercando il modello che portasse ad ottenere la Coherence più alta. I grafici ottenuti per i due anni sono molto diversi. Mostreremo quello relativo all'anno 2021 perché è il più interessante:

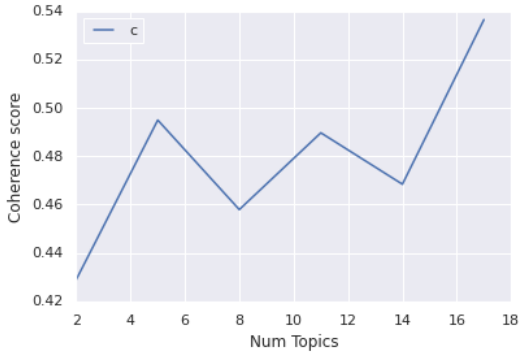

Figure 9: Optimal topic number 2021

Come è possibile osservare la Coherence non incrementa in maniera regolare (sono presenti diversi massimi e minimi locali). Vediamo, per i due dataset, quale è il miglior numero di topic per ottimizzare la Coherence secondo l'algoritmo:

Anno	n_topic	Coherence
2019	12	0.5596
2021	11	0.4897

Come si nota è stato ottenuto un netto miglioramento della Coherence. C'è da evidenziare che questo algoritmo calcola la Coherence partendo da un numero di topic minimo, fino a raggiungere un valore massimo, ed è possibile settare uno step. Lo step settato è stato 3, perché per via della ridotta potenza di calcolo e memoria con uno step minore l'algoritmo non riusciva a terminare. La conseguenza di questa limitazione è che applicando il modello con un numero di topic scelto da noi abbiamo ottenuto un modello con maggiore Coherence nel 2021. Questo però non è vero per l'altro dataset.

Si è in fine trovato il topic dominante per ogni riga del dataset dei tweets del 2019, la riga che maggiormente rappresenta i diversi topic e infine come sono distribuiti i diversi topic. Queste informazioni sono disponili rispettivamente nei file: df_dominant_topic_2019.csv, df_representative_row_topic_2019.csv e df_topic_distribution_2019.csv.
In particolare la labelizzazione presente nel primo di questi tre dataset è stata usata a supporto delle predizioni spiegate nella prossima sezione.

5 Predictions

5.1 Data Preparation

Il dataset finale del 2019 per le predizioni è stato ottenuto raggruppando per minuto il dataset dei tweets del 2019 (quello filtrato per Polarity e labelizzato con i sentiment e il Topic) e calcolando: il volume di tweets in quel minuto, come conteggio dei testi distinti; la mediana della polarity assegnata con il modello VADER; le due mode delle labels del Sentiment assegnate dai modelli VADER e Transformers; e la moda del Topic. E successivamente facendo un Join tra questo dataset raggruppato e il dataset dei prezzi del Bitcoin per minuto del 2019. Prima però anche quest'ultimo dataset è stato processato generando i nuovi attributi: "returns", che sono i log-returns ottenuti dal prezzo di chiusura di ogni minuto ("close"); "up_down", con valore -1 se in quel minuto il prezzo è sceso o +1 se è salito; e infine la "volatility", calcolata come standard deviation dei returns in una finestra temporale di 4 ore (240 minuti) precedenti al minuto corrente.

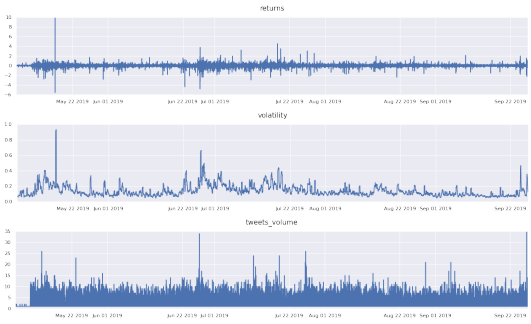

Figure 10: returns, volatility and tweets volume

Il dataset finale dopo il Join contine 127'575 righe, di cui 26'245 sono quelle labelizzate con sentimento negativo dal Transformer. Quì sopra nella figura 10 riportiamo le serie temporali di "returns", "volatility" e "tweets_volume" e notiamo che c'è una discreta correlazione tra il volume dei tweets e la volatilità del prezzo.
L'attributo più interessante su cui fare predizioni è quello binario "up_down" relativo all'andamento del prezzo nel minuto. Ovviamente la predizione non deve essere fatta sull'attributo così com'è, ma dobbiamo predire una versione "laggata" di esso. In ottica di trading predire cosa sta facendo adesso il prezzo non è utile, ma vogliamo sapere cosa farà in futuro. per scoprire il numero di lag ottimali di cui shiftare indietro le time series per effettuare le predizioni è stata calcolata la Partial Autocorrelation Funcion del prezzo di chiusura del minuto, che riportiamo quì sotto.

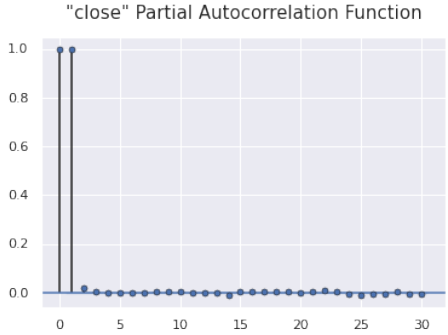

Figure 11: PACF Plot

E' evidente che il prezzo presenta un comportamento autoregressivo, con un'elevata autocorrelazione con i valori della serie shiftata fino a 2 lag. Lo shift di due minuti in avanti è quindi stato applicato a ciascuno dei tre attributi "returns", "volatility" e "up_down" per generare i 3 nuovi attributi su cui effettuare le predizioni. Questo ci permette di mantenere le serie originali nel dataset di input per i modelli predittivi, i quali saranno utili per prevedere l'andamento del prezzo dopo 2 minuti. Da notare, come è ovvio che sia, il comportamento estremamente autoregressivo della volatilità.

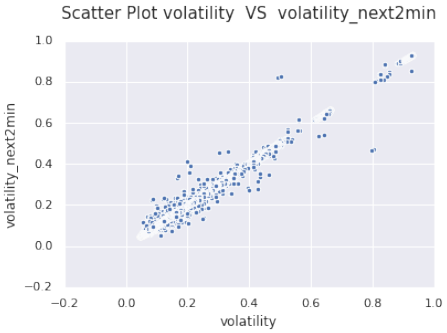

Figure 12: Volatility Scatter Plot

Le stesse trasformazioni sono state applicate al dataset del 2021, tranne per il fatto che è stata aggiunta una versione pesata della polarity per il numero di followers di ogni utente proprietario del tweet. E' stato notato comunque che il dataset del 2021 risulta troppo frammentario e non discutiamo i risultati.

5.2 Up or Down Prediction

Il dataset finale del 2019 è stato diviso in training e test set nella data "2019-09-05". Il test set contiene il 50.8475% di valori pari a +1 (=up, che è quindi la classe maggioritaria) nell'attributo da predire ("up_down_next2min"). Quindi questa percentuale è quella di riferimento da superare in termini di Accuracy per avere un modello che performa meglio di predire semplicemente sempre "up". Sono stati testati alcuni modelli con diverse configurazioni di scaling e PCA, ottimizzando i parametri con una random search con Cross Validation di 5 folds. Alla fine la migliore pipeline è uno Scaling più una Random Forest con parametro "max_depth"

pari a 12, senza applicare la PCA, ed è stata salvata. Riportiamo quì sotto i risultati.

Anno	2019	2021
Accuracy	0.5091	0.5076
F1	0.4968	0.4933

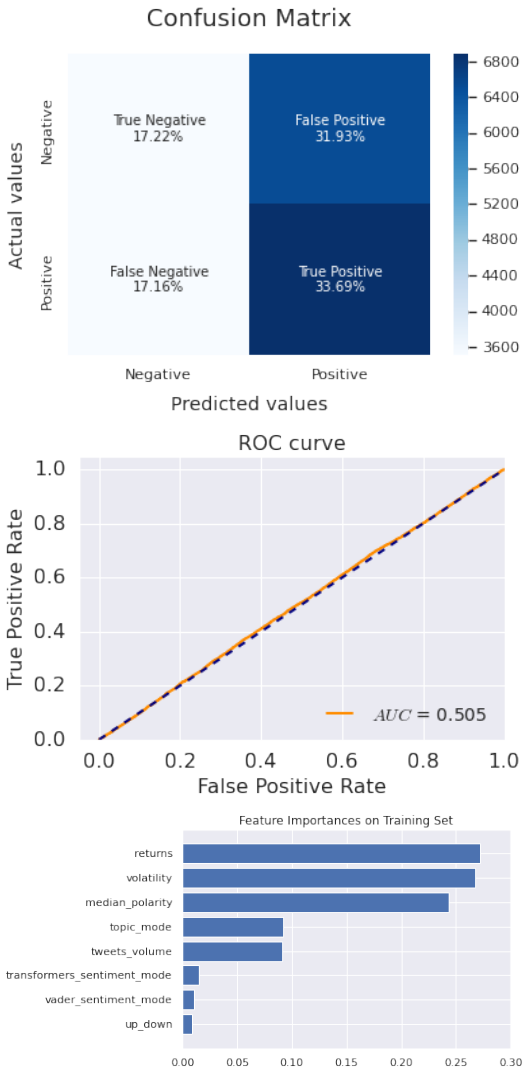

Figure 13: Predictions Results for 2019

Un'Accuracy del 50.911% significa che seguire il modello per investire nel periodo temporale a cui si riferisce il Test Set avrebbe reso le performances solo leggermente migliori rispetto ad andare semplicemente sempre al rialzo. Ci si può aspettare però che sia sempre auspicabile seguire il modello. Diversificare l'investimento andando sia al rialzo che al ribasso infatti previene le perdite ingenti che si avrebbero nei casi in cui il mercato spingesse costantemente nella direzione opposta a quella scelta se non si seguisse il modello. E questo rende la performance più stabile anche nel breve periodo. Come si nota dalla feature importance estratta dal modello gli attributi più importanti sono "returns" e "volatility", ma anche la mediana della VADER Polarity, la moda del Topic e il volume di Tweets rivestono un'importanza significativa. Lo stesso modello è stato fittato su un dataset contenente solo gli attributi estratti dal dataset dei tweets (di derivazione testuale) e su un dataset che invece conteneva solo gli altri attributi (autoregressivi). Si è notato che i soli attributi riferiti ai tweets non bastano per superare l'Accuracy minima sopra citata, ma i risultati sono migliori del fitting con solo gli attributi autoregressivi. Questo risultato dimostra che i dati testuali possono essere usati efficacemente per migliorare o stabilizzare i modelli previsionali. Per certe configurazioni di parametri e Test Set si è riusciti a raggiungere anche Accuracy del 57.0235%, cosa che sarebbe praticamente impossibile seguendo sempre la stessa direzione sul mercato. Sono stati effettuati anche tentativi di predire la Volatilità dei due minuti successivi e come ci si può aspettare i risultati sono quasi perfetti: R2 = 0.998.

6 Web Page

Poiché l'attributo relativo alla mediana della VADER Polarity lungo il minuto è risultato quello più significativo per la predizione del prezzo è stata prodotta una pagina web che rende possibile il suo monitoraggio in tempo reale. Le tecnologie e tools utilizzati sono state: Flask, SQLAlchemy, Twitter StreamListener, Jinja Template, HTML, CSS, Bootstrap, Plotly Express, GIT. La app funziona tramite due Entry-Point, uno per popolare il DB SQL Server con i dati dei Tweets in tempo reale (che simula dati provenienti dagli utenti), e l'altro per runnare la pagina web. Quest'ultima mostra un solo grafico della mediana della polarity al minuto che si aggiorna tramite una richiesta GET, con anche la possibilità di cancellare tutti i dati dal DB. Il codice è stato organizzato secondo il pattern architetturale MVC.