Università degli Studi di Bari Aldo Moro
Corso di Laurea in Informatica - Anno Accademico 2020-2021

Analisi del mercato delle Criptovalute tramite Sentiment Analysis di dati Twitter

Laureando: Gianfranco Demarco

Relatrice: Gabriella Casalino

Introduzione

- Negli ultimi anni, il numero di tweet postati giornalmente è cresciuto ad un ritmo quasi esponenziale

- La Sentiment Analysis è un campo del Natural Language Processing (NLP) che si occupa di ricavare ed estrarre opinioni da dati in forma testuale.

- Il *sentiment* degli utenti espresso su Twitter è spesso stato utilizzato in letteratura per provare a prevedere l'andamento del mercato azionario.

Obiettivo della tesi

Raramente la Sentiment Analysis è stata applicata al mercato delle criptovalute.

Mercato delle criptovalute:

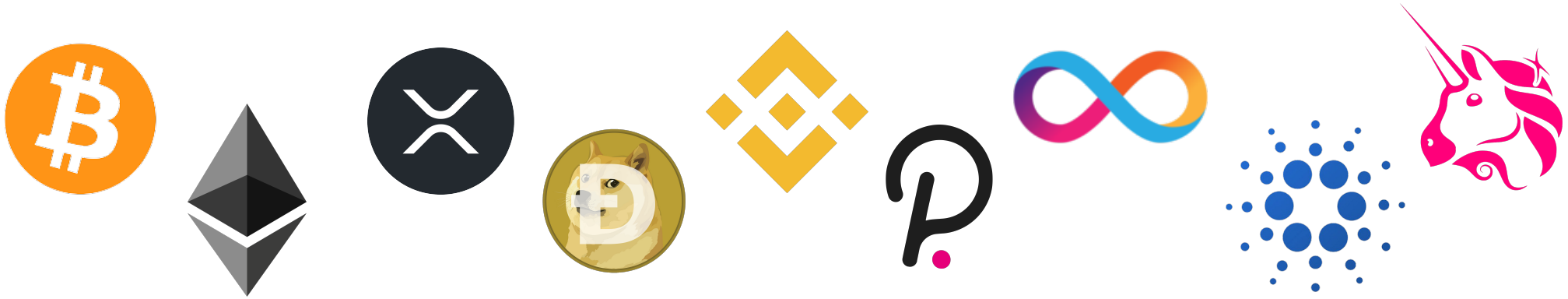
- molto discusso sul Web
- alta volatilità

L'obiettivo di questa tesi è:

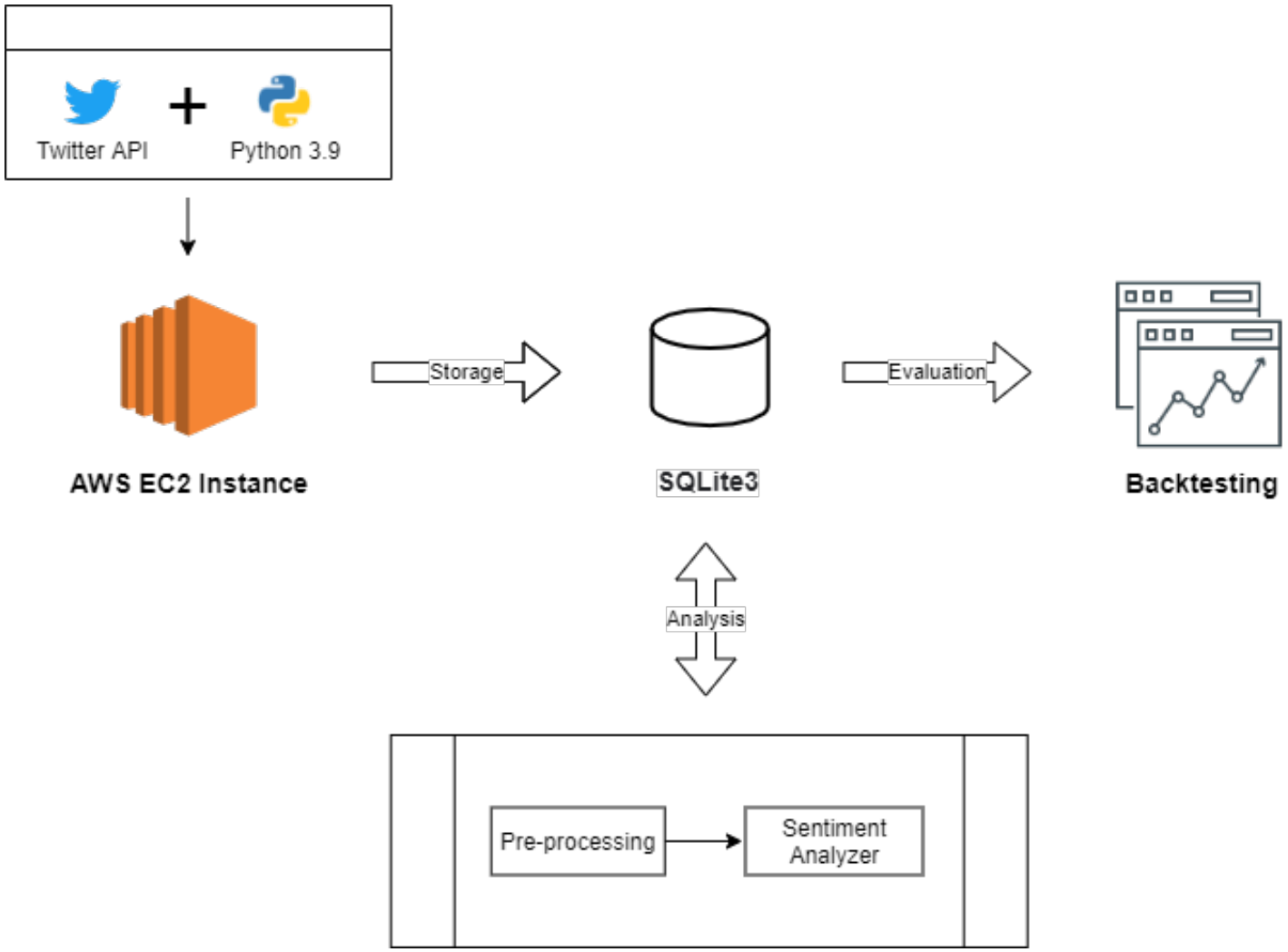
- verificare se esiste una correlazione tra sentiment e prezzi
- sviluppare modelli per generare un ritorno economico

Criptovalute e Blockchain

- Bitcoin nasce nel 2009 da Satoshi Nakamoto
- Si basa sulla *Blockchain*
- Dopo Bitcoin sono nate numerose altre monete (Altcoin)



Architettura del sistema sviluppato



Raccolta dei dati

Strumenti utilizzati:

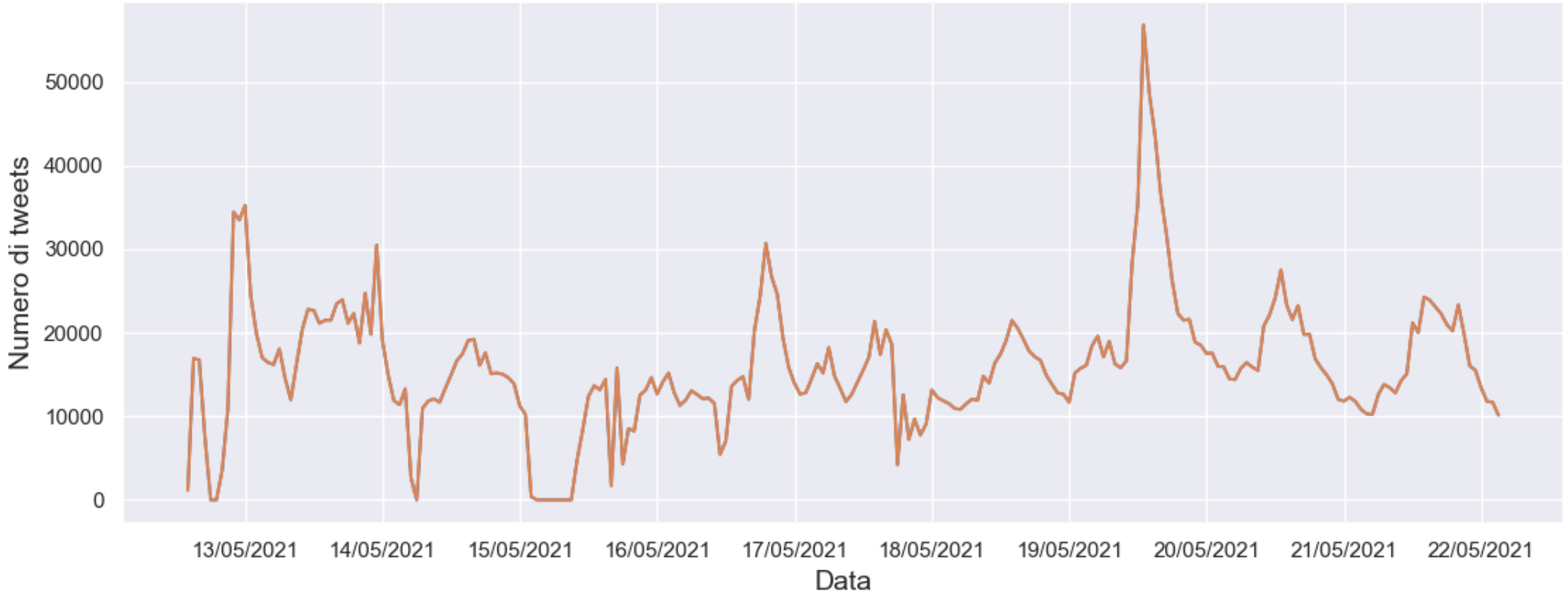
- Python 3
- Twitter API (Tweepy)
- SQLite (peewee)
- Amazon Web Services

Keyword utilizzate:

Bitcoin	BTC	Ethereum	ETH
Binance coin	BNB	Dogecoin	DOGE
Cardano	ADA	Internet Computer	ICP
Polkadot	DOT	Uniswap	UNI
XRP			

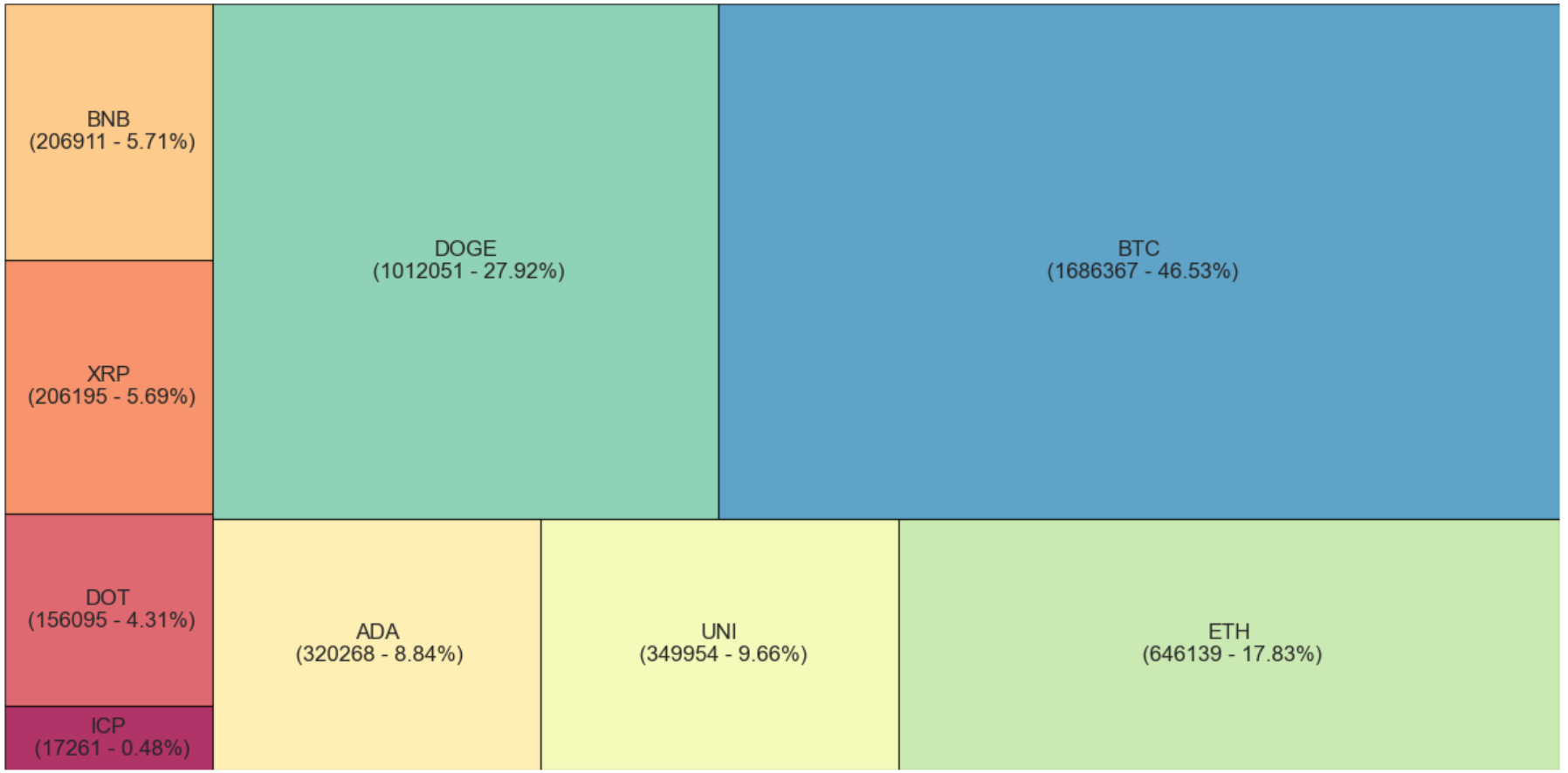
Composizione del dataset

Il dataset è composto da circa 3,5 milioni di tweet raccolti dal 12/05/2021 al 22/05/2021.



Composizione del dataset

Il dataset è composto da circa 3,5 milioni di tweet raccolti dal 12/05/2021 al 22/05/2021.

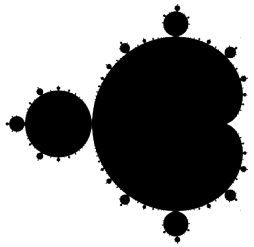


Analisi del sentiment

Per effettuare l'analisi del sentiment sono stati utilizzati 3 diversi approcci:

- Algoritmo basato su *lexicon e regole* (VADER della libreria NLTK)
- Algoritmo probabilistico basato su Naive Bayes (della libreria Textblob)
- Rete neurale di tipo ricorrente (in particolare, *Long Short-Term Memory* implementata nella libreria FlairNLP)

NLTK



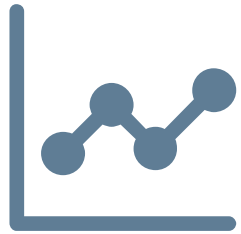
TextBlob

flair

Metodologia di valutazione

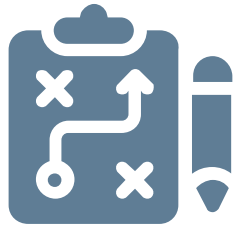
- Tecnica del *Backtesting*
- API di https://cryptowat.ch/
- Trading agent (3 varianti)

$$\begin{cases} \text{apre una posizione } \textit{long} & \text{se sentiment} > \text{threshold} \\ \text{apre una posizione } \textit{short} & \text{se sentiment} < \text{-threshold} \\ \text{non opera} & \text{altrimenti} \end{cases}$$



Risultati sperimentali

- *Trading agent* ha operato su 230 periodi (ore)
- Con *trading agent* semplice: perdita o pareggio di capitale
- Con *trading agent* più complesso: netti miglioramenti
- Risultati migliori considerando solo BTC, DOGE, ETH
- Miglior modello: LSTM



Risultati sperimentali

Crypto	Min	Max	Finale	# operazioni
BTC	0.99	1.50	1.39	207
ETH	0.98	1.99	1.40	65
BNB	0.81	1.13	1.00	121
DOGE	1.00	1.85	1.74	42
ADA	0.82	1.10	1.10	16
XRP	0.86	1.04	0.97	26
ICP	0.88	1.79	1.19	107
DOT	0.92	1.58	1.58	42
UNI	0.98	1.07	1.04	35
avg	0.92	1.45	1.27	73.44

Table 5.9: *FlairNLP: Capitale minimo, massimo, finale e numero di operazioni effettuate con soglia del sentiment*

(230 ore)

Guadagno medio: **27%**

Guadagno medio BTC, ETH, DOGE: **51%.**

Conclusioni

- Le scelte progettuali hanno influenzato in maniera simile i 3 modelli implementati
- Risultati migliori usando *trading agent* più complessi
- *Trading agent* altamente personalizzabili
- Risultati migliori usando modelli più complessi
- Con LSTM, percentuale elevata di ritorno sul capitale in meno di 10 giorni (tra 27% e 51%)

Sviluppi futuri

Miglioramento del sistema

- Parallelizzazione della pipeline creando un sistema *real-time*
- Integrazione con API esterne che permettano il testing su un vero portafoglio

Validazione dei modelli

- Raccogliere più dati per validare i risultati sperimentali
- Raccogliere più dati per poter operare su *timeframe* differenti
- Implementare ulteriori tecniche per la Sentiment Analysis