Università degli Studi di Milano Bicocca

Scuola di Scienze

Dipartimento di Informatica, Sistemistica e Comunicazione

Corso di laurea in Informatica

Un modello ensemble basato su blockchain e reti neurali per la previsione del prezzo di Bitcoin

An ensemble model based on blockchain and neural networks for Bitcoin price prediction

Relatore: *Antonio Candelieri*

Co-relatore: *Silvio Maria Enrico Bencini*

Relazione della prova finale di:

Harshit Vikram

Matricola 865938

Anno Accademico 2022-2023

Abstract

L’individuazione di comportamenti anomali all’interno dei mercati finanziari, grazie ad opportune analisi e considerazioni può portare ad ottenere implementazioni di strategie di investimento più efficienti e profittevoli.

Nella presente ricerca, sulla base dei dati riferiti all’attività giornaliera registrata sulla blockchain di **Bitcoin** e del sentiment di mercato osservato tramite l’indicatore Net Unrealized Profit-Loss (**NUPL**), è stato progettato e realizzato un modello ensemble di tipo one-step ahead basato su reti neurali Long-Short Term Memory (**LSTM**) e Gated Recurrent Unit (**GRU**), che ha come scopo quello di prevedere il prezzo di Bitcoin.

Le sperimentazioni effettuate portano ad ottenere un’accuratezza molto positiva nelle previsioni ottenute. Inoltre, una conferma della bontà di questi risultati è rispecchiato anche dall’utilizzo dello strumento progettato, all’interno di una strategia di investimento automatizzata, dove ha portato a raggiungere un profitto netto pari a +13.3%. (escludendo eventuali costi di transazione e commissioni)

Indice

Capitolo 1
Introduzione

Lo studio delle criptovalute è divenuto un ambito di interesse particolarmente rilevante nell'ultimo decennio. Quest'affermazione è semplicemente deducibile dalla quantità sempre più crescente di paper scientifici, articoli online e post pubblicati sui social media. Tale interesse è dovuto alle numerose applicazioni in cui è possibile impiegare le criptovalute e la tecnologia blockchain che le mantiene in vita.

Ma l'aspetto più considerato in tutto ciò, è la possibilità di investire nelle criptovalute, comprandole direttamente o esponendosi su di esse tramite strumenti proposti da intermediari finananziari o aziende di invesitimento con lo scopo di guadagnare un profitto.

La presente trattazione è incentrata solamente sulla previsione del prezzo di Bitcoin, prima criptovaluta nata sulla base del WhitePaper pubblicato dal suo ideatore Satoshi Nakamoto nell'ottobre 2008. [1]

La realizzazione di modelli basati puramente sull'andamento precedente del prezzo, possono alludere ad ottime performance di previsione, come riscontrato durante le sperimentazioni condotte (come verrà illustrato nel capitolo 3) considerando diverse metriche di errore, tra cui anche il Mean Absolute Percentage Error (MAPE) che presenta un valore inferiore al 2%. Il vero significato di tale risultato consiste semplicemente nel fatto che le variazioni del prezzo risultano spesso vicine al valore precedente.

In realtà è necessario un approccio più articolato, che prenda in considerazione anche i numerosi fattori che portano il prezzo di Bitcoin a cambiare.

Infatti le fluttuazioni relative al prezzo possono verificarsi in modo molto marcato e in periodi anche molto brevi, con sbalzi che possono spingersi fino ad eventi definiti come flash crash (come ad esempio: - 49,47% in solo 11 giorni, tra il 9 Mag 2021 e il 19 Mag 2021) oppure variazioni al rialzo definibili come market rally (come ad esempio: +135,38% in 28 giorni, tra il 12 Dic 2020 e il 08 Gen 2021).

Da quanto emerge da survey e paper scientifici passati, modelli di previsione basati su algoritmi statistici, su reti neurali e sul reinforcement learning stanno venendo progettati ed impiegati con lo scopo di fornire previsioni quanto più accurate possibili sfruttando

l’immensa disponibilità di dati online per questa criptovaluta.
Un’approfondimento sullo stato attuale dei tipi di algoritmi più utilizzati attualmente verrà fornito all’interno del capitolo 2.

L’approccio adottato nella presente ricerca, risiede nell’utilizzo di dati rappresentanti:

- l’effettiva attività di investimento su Bitcoin derivabile dalla sua blockchain
- il market sentiment degli investitori tramite l’indicatore Net Unrealized Profit-Loss (NUPL), il quale riassume in modo particolare la capitalizzazione di Bitcoin, e i profitti e le perdite che si sono già avverate, osservando la blockchain.

L’impiego di tali informazioni per la preparazione di modelli basati su reti neurali di tipo Long-Short Term Memory (LSTM) e Gated Recurrend Unit (GRU), ha permesso di realizzare un modello ensemble in grado di fornire previsioni con errori pari a 676,12 $ (RMSE) e 2,17% (MAPE).

Capitolo 2
Problema e stato attuale del contesto analizzato

2.1 - Problema: Le difficoltà nella previsione del prezzo di Bitcoin

Bitcoin o BTC (ticker e nominativo abbreviato utilizzato nei mercati finanziari per riferirsi a questa criptovaluta) nato come strumento rivoluzionario per effettuare transazioni in modo decentralizzato grazie alla tecnologia blockchain, è divenuto in realtà un asset particolarmente adorato da investitori e speculatori, per merito dell'attraente volatilità del suo prezzo.

Tale fattore può permettere ad investitori attenti di trarre beneficio sia dai movimenti a rialzo che a ribasso molto intensi che si possono verificare in un arco temporale relativamente breve, come pochissime ore o giorni.

Avere a disposizione strumenti che possano usufruire ed elaborare dati provenienti da numerose fonti, e che siano in grado di mostrare con una completezza e puntualità utile l'andamento e la possibile direzionalità di un asset è divenuto un obiettivo a cui si cerca di rispondere con la tecnologia e la continua ricerca scientifico-economica.

Anche nel caso della previsione del prezzo di Bitcoin, la realizzazione di strumenti e modelli basati su Machine Learning con lo scopo di fornire un supporto aggiuntivo per le scelte di investimento, può risultare efficace, ma solamente se vengono presi in considerazione i diversi fattori causali da cui esso è continuamente stimolato.

Queste cause possono determinare movimenti di altissima volatiltià, che in primo luogo potrebbero non indicare una direzionalità ben precisa, o addirittura non avere un effetto immediato sul prezzo. Ma come verrà illustrato nei prossimi capitoli, gli effetti potrebbero addirittura verificarsi a distanza di settimane, o diversi mesi.

Prima di procedere illustrando questi fattori, è importante chiarire il modo in cui è possibile scambiare Bitcoin, ovvero il modo in cui è possibile ottenere una certa quantità di questa criptovaluta. Ogni entità che intente operare personalmente con Bitcoin deve

necessariamente possedere un **wallet** (o portafoglio). Durante la generazione di quest'ultimo vengono create 2 chiavi, una chiave pubblica o **public key** (condivisibile per poter ricevere questa criptovaluta) e una chiave privata o **private key** (strettamente personale e che permettere al suo possessore di autorizzare trasferimenti di fondi in uscita dal wallet a cui fa riferimento). Ogni transazione viene registrata e confermata tramite un processo chiamato **mining**, all'interno di un registro pubblico e distribuito organizzato a blocchi, noto come **blockchain**. [2]

In questo modo è possibile inviare una somma ben definita di questa criptovaluta da un wallet ad un altro, ma a seguito del pagamento di una piccola commissione obbligatoria e variabile necessaria per la registrazione della transazione. La modalità appena descritta è nota come transazione **peer-to-peer** [rif all'immagine], ed è la base a cui ci si riconduce ogni qualvolta una persona o entità decide di acquistare o spostare Bitcoin.

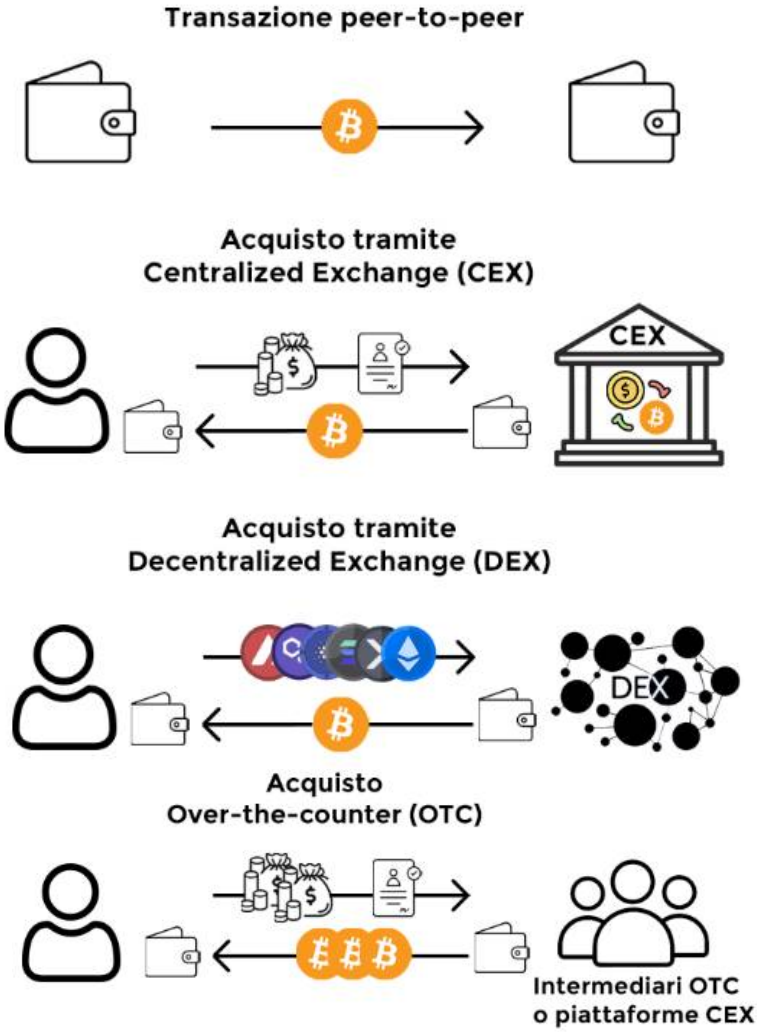


Figura 1: I diversi metodi con cui è possibile investire o acquisire Bitcoin

Come rappresentato dai vari metodi raffigurati nell'immagine Figura 1, le modalità utilizzate per scambiare, acquistare o vendere, e quindi trarre beneficio investendo in Bitcoin sono:

1. Transazioni **peer-to-peer** (descritto precedentemente)
2. Acquisto tramite piattaforma **CEX** (Centralized Exchange): un acquirente deve registrarsi con i propri dati personali su una piattaforma CEX, e può procedere all'acquisto di Bitcoin caricando il proprio profilo con valute "fiat" (come € EUR, $ USD, £ GBP, o altri...) oppure altre criptovalute. Dopodiche se la piattaforma lo consente, l'acquirente può conservare le criptovalute acquisite sulla stessa piattaforma, oppure prelevarle su un proprio wallet privato.
3. Acquisto tramite piattaforma **DEX** (Decentralized Exchange): un acquirente può acquistare Bitcoin scambiandoli con altra criptovaluta in suo possesso, senza alcun bisogno di registrarsi. Questa operazione potrebbe coinvolgere diversi portafogli per completare lo scambio, ma l'acquirente alla fine del processo otterrà sul portafoglio da lui indicato i Bitcoin comprati.
4. Acquisto **OTC** (Over-the-counter): un acquirente che intende acquistare un grande quantitativo di Bitcoin, in seguito ad accordi con entità come ad esempio piattaforme CEX, oppure intermediari OTC, potrebbe avere la possibilità di acquistare un ingente quantità di criptovalute ad un prezzo fisso concordato tra le parti, solitamente a seguito del pagamento di una commisione importante.
5. Acquisto indiretto tramite **strumenti di investimento** come strumenti derivati (futures e opzioni)

Sebbene esistano diversi modi, gli scambi effettivi che vengono effettuati al di sotto di queste operazioni vengono registrati sulla blockchain di Bitcoin, riportando informazioni

interessanti come gli importi scambiati, ma anche dettagli come le chiavi pubbliche dei due portafogli interessati dallo scambio.
Tali informazioni rappresentano una'ottima risorsa, che insieme ai dati relativi ai fattori di grande impatto sulle fluttuazioni del prezzo di Bitcoin possono aiutare a svillupare strumenti come modelli di Machine Learning (come verrà mostrato successivamente).
Di seguito si cerca di riportate in modo esaustivo, le cause più importanti che spesso portano a fluttuazione significative del prezzo di questa criptovaluta:

- trasferimenti di grandi fondi in ingresso o in uscita da piattaforme CEX o su piattaforme dove è possibile operare con strumenti finanziari come futures e opzioni
- annunci da parte di enti regolamentatori che esprimono l'intenzione di introdurre limitazioni e regole più restrittive sugli investimenti sulle criptovalute o di perseguire le piattaforme di scambio pubblico come le CEX per attività sospette
- annunci sui social media da parte di persone o entità come fondi di investimento, banche o addirittura CEO di grandi aziende che esprimono il proprio interesse o una possibile intenzione di investire in un certo modo su Bitcoin
- L'apertura e la concentrazione di numerosi ordini di tipo **stoploss** o **takeprofit** (ordini utilizzati dagli investitori per tutelare una propria posizione di investimento chiedendo di chiuderla a mercato per intero o parzialmente al raggiungimento di un livello di prezzo specifico o al verificarsi di determinate condizioni) in determinate zone di prezzo, che una volta raggiunte potrebbero scatenare la chiusura di diversi ordini e quindi scatenare movimenti impulsivi dal punto di vista del prezzo.

Una previsione affidabile del prezzo di questa criptovaluta, oltre ai fattori sopraccitati richiede anche di considerare alcuni problemi non banali che rendono ancora più ostica la previsione del prezzo di Bitcoin. In particolare si fa riferimento a:

- La moltitudine dei fattori da interpretare per fornire una visione completa dell'andamento della criptovaluta
- La difficoltà nell'attribuire la causa di un determinato movimento ad fattore specifico
- La disuguaglianza dei dati in possesso dei vari investitori nel mercato delle criptovalute. Infatti in alcuni casi a seguito del pagamento di un abbonamento, è

possibile accedere ad informazioni di mercato che vengono continuamente registrate su una piattaforma CEX, come nel caso delle informazioni riferite agli ordini aggiunti, modificati o cancellati negli orderbook.

- La difficoltà nel valutare se i portafogli non attivi da molto tempo sono stati persi o se sono semplicemente tenuti per investimenti di lungo termine.
- La difficoltà nel valutare se determinate operazioni di scambio siano state effettuate tra 2 entità diverse, o dalla stessa entità ma per effettuare semplici operazioni di spostamento dei propri fondi.

2.2 - La blockchain: Indicazione per comprendere l'andamento futuro del prezzo

Una previsione solida per il prezzo futuro di Bitcoin, è un'attività che può risultare ostica, se si intende tenere in considerazione i numerosi fattori e problemi citati prima.

In questo tipo di contesto, la blockchain su cui si basa questa criptovaluta può essere utilizzata per analizzare e argomentare le ripercussioni dei fattori causali descritti in precedenza su Bitcoin. Inoltre permette di comprendere come e di quanto a seguito di determinate transazioni, il prezzo di Bitcoin si è mosso in una determinata direzione.

Nell'immagine seguente. viene mostrato come l'incremento del numero di portafogli con x ≥ 1000BTC possa fornire una chiara indicazione della direzione del prezzo futuro di Bitcoin.

Figura 2: Andamento del prezzo di Bitcoin e del numero di indirizzi con X ≥ 1000BTC tra il 2017 e Marzo 2023

Se da un lato la blockchain risulta pubblicamente consultabile, rimane comunque una forte disuguaglianza tra le informazioni a disposizione dei vari investitori. Infatti, gli operatori con a disposizione i dati più recenti (riferiti ad ordini su strumenti derivati e operazioni di scambio effettuate su piattaforme CEX) o con modelli e strumenti più avanzati hanno un notevolte vantaggio sulla visione complessiva di questa criptovaluta. Dunque è ragionevole pensare che riescano ad operare in maniera più profittevole di quelli che non possiedeno quel tipo di strumenti.

Tale intuizione e la disponibilità di informazioni derivabili dalla blockchain di Bitcoin sono le componenti fondanti su cui si basa il modello ensemble proposto in questo trattazione.
La novità che si intende introdurre alla letteratura riferita allo studio della previsione del prezzo delle criptovalute, risiede nell'analisi, nella progettazione e nel possibile utilizzo di un modello ensemble basasto su reti neurali di tipo LSTM (Long Short Term Memory) e GRU (Gated Recurrent Units), in grado di prevedere con una buona accuratezza il prezzo di Bitcoin (fare riferimento alle metriche di valutazione illustrate nel capitolo 3) facendo uso di dati di tipo serie temporale come:

- dati di mercato: prezzo di Bitcoin
- dati derivati dalla blockchain di Bitcoin:
  - numero di portafogli con $X \geq 1000BTC$
  - numero di portafogli con $100BTC \leq X < 1000BTC$
  - indicatore NUPL (Net Unrealized Profit and Loss)

Un approfondimento relativo a questi dati viene affrontato all'interno del capitolo 3.

Capitolo 3
Teoria: Stato dell'arte dei modelli usati per la previsione del prezzo di Bitcoin

In questo capitolo verranno fornite in modo sintetico le nozioni teoriche necessarie per comprendere gli esperimenti eseguiti, che hanno permesso di ottenere i risultati presenti in questa trattazione. Inoltre vengono illustrate le soluzioni adottate in paper scientifici passati, con lo scopo di prevedere il prezzo di Bitcoin o di altre criptovalute, adottando diverse tipologie di modelli di ML e DL.

3.1 - Le serie temporali: Struttura dati più usata nei mercati finanziari

I dati che vengono analizzati nelle ricerca relative ai mercati finanziari sono principalmente di tipo **serie temporale**. Questi ultimi sono definibili come sequenze di dati registrati secondo l'ordine temporale, durante un periodo di tempo ben definito. Ovvero, per ogni dato, viene salvato insieme a ciascuno di essi anche il momento in cui è stato registrato.

Inoltre i singoli punti che rappresentano la serie, risultano equidistanti nel tempo l'uno dall'altro.

Infatti, un aspetto da tenere in considerazione per questo tipo di strutture dati è la **granularità** o **timeframe**, che rappresenta l'intervallo di tempo tra la registrazione di un dato e l'altro.

Gli scopi principali per cui sono impiegate le serie temporali all'interno dei mercati finanziari, sono: [3]

- l'osservazione del valore degli asset / strumenti finanziari o variabili macroeconomiche nel tempo;
- analisi di correlazione e relazione di tipo causa-effetto tra 2 o più serie;
- preparazione di modelli di previsione basati su serie temporali e algoritmi di ML, DL o RL.

3.2 - Analisi della correlazione e della causalità tra serie temporali

La comprensione della presenza di una possibile relazione tra 2 serie temporali riferite all'osservazione di fenomeni o variabili diverse, può rivelarsi utile per comprendere meglio i fenomeni osservati o addirittura effettuare scelte di investimento più consapevoli.
Tramite l'esempio seguente viene illustrato un modo sistematico grazie al quale si possono individuare possiibli **relazioni di correlazione** e **relazioni di tipo causa-effetto** che possono presentarsi tra serie temporali riferite a dati in ambito finanziario.
Considerate 2 serie temporali (serieA e serieB), con:

- lo stesso numero di elementi,
- valori numerici continui,
- nelle quali viene usato lo stesso timeframe
- distribuzione diversa dalla distribuzione Normale
- andamento non stazionario
- presenza di valori outliers

Grazie ad una prima fase di visualizzazione grafica tramite grafici a linee è possibile notare in modo approssimativo, il range di valori all'interno del quale si muovono le serie, e il loro andamento sulla base del timeframe utilizzato.
La disposizione dei valori delle due serie, in un ipotetico caso in cui la serieA riporta i valori della serieB, ma con un lag temporale ben definito e un'intensità diversa, potrebbe in alcuni casi indurre a pensare ad una possibile presenza di una relazione di causalità, dove ad esempio la serieB influenza la serieA dopo un certo periodo temporale.
Tale supposizione in realtà può essere validata seguendo un semplice processo così articolato:

1. Verifica della correlazione delle serie considerate
2. Verifica della relazione di causalità tra le serie
3. Valutazione degli esiti ottenuti per affermare la relazione presente

La correlazione tra la serieA e la serieB permette di misurare il grado di relazione con cui gli elementi delle due serie variano insieme all'interno di un periodo di tempo ben definito.
Il tipo di correlazione individuato dipende dal valore ottenuto per il coefficiente di

correlazione calcolato, che si presenta all'interno dell'intervallo tra -1 e 1. [4]
Infatti, in caso il valore ottenuto fosse:

- 1 | indica la presenza di una correlazione perfettamente positiva ed è possibile affermare che nel momento in cui la serieA si muove in una direzione lo stesso avviene anche per la serieB.
- -1 | sarebbe il caso di una correlazione negativa perfetta, la quale indica movimenti discordi tra le serie considerate.
- 0 | assenza di una correlazione tra le serie considerate

Oltre al segno, anche il valore ottenuto permette di categorizzare ulteriormente il tipo di correlazione. Infatti valori come: [5]

- $0 < x \leq 025$: non permettono di affermare la presenza di una correlazione
- $0{,}25 < x < 0{,}5$: indicano una correlazione debole
- $0{,}5 < x < 0.75$: indicano una correlazione significativa
- $0{,}75 < x \leq 1$: indicano una correlazione forte

Questa informazione può essere ricavata tramite diversi strumenti, ma tra quelli più utilizzati vi sono: [6]

- **PCC – Pearson's correlation coefficient** – noto anche come coefficiente di correlazione di Pearson

  Questo coefficiente di correlazione è utilizzato per capire l'eventuale presenza di una relazione lineare tra le variabili considerate, insieme alle relative intensità e direzione. Per una corretta interpretazione dei risultati ottenuti, è necessario che vengano considerate le seguenti assunzioni: le variabili devono essere continue, con una distribuzione normale e che non possiedano valori outliers.
- **KCC – Kendall's correlation coefficient** – noto anche come Kendall rank coefficient

  è un coefficiente di correlazione definito come non-parametrico (ovvero non richiede assunzioni sulle distribuzioni delle variabili utilizzate) che permette di calcolare il grado di concordanza o discordanza tra le 2 variabili considerate. A differenza del PCC, è possibile utilizzarlo con variabili non necessariamente continue e non è richiesto che vengano rispettate le assunzioni indicate per il

calcolo del PCC.

- **SCC – Spearman's correlation coefficient -** noto anche come Spearman's rank correlation coefficient [7]

  è un coefficiente di correlazione non-parametrico, che permette di misurare la bontà con cui il grado di associazione delle variabili considerate può essere descritto tramite una funzione monotona. In questo modo è possibile misurare in modo più efficiente, se l'andamento dei valori delle variabili cambia in modo simile, anche se con un grado di variazione differente.

Sebbene l'utilizzo del PCC non è consigliato per la serieA e la serieB, utilizzarlo potrebbe comunque fornire un'indicazione sulla possibile presenza di una correlazione, che però andrebbe verificata opportunamente con altri strumenti.
Dunque, considerando le proprietà delle serieA e serieB risulta conveniente utilizzare i coefficienti KCC ed SCC.

Nonostante i risultati che si possono ottenere, in certi casi è addirittura possibile riscontrare problemi come quelli legatti alle **correlazioni spurie**.
Infatti, qualora il coefficiente di correlazione portasse ad una relazione significativa o forte, in certi casi non è possibile spiegare un legame logico tra le 2 variabili osservate, per questo motivo si può affermare che la correlazione sia dovuta :

- ad una semplice coincidenza
- oppure a causa della presenza di una possibile variabile terza (non considerata) che in qualche modo va ad influire sulle due variabili osservate, le quali di conseguenza tendono a comportarsi nel modo indicato dal coefficiente di correlazione [8]

Per poter verificare in modo corretto la possibile relazione di causalità tra serie temporali, è possibile utilizzare il **Granger Causality test**. [9]
Quest'ultimo consiste in un test statistico ideato e formulato dall'economista inglese Clive Granger, che non consente di affermare con fermezza la presenza di relazione di tipo causa-effetto, ma di evidenziare la presenza di una **causalità predittiva**, ovvero un possibile miglioramento nella previsione dei valori di una serieA, considerando non solo i valori

passati della serie stessa, ma anche dati riferiti ad un'altra serie B con particolari lag temporali. Nel caso in cui il test riporti un esito affermativo, si può concludere che la serie A è **granger-caused** dalla serie B.

Per poter procedere, è necessaria un'assunzione importante, ovvero quello di accertarsi che le serie su cui si intende eseguire il Granger Causality test siano **serie stazionarie**. Vengono definite in questo modo le serie che presentano caratteristiche statistiche come media e varianza costanti nel tempo e quindi prive di **trend** (andamenti con una direzione molto evidente, che può essere rialzista o ribassista), e **stagionalità** (ripetizione dell'andamento, anche se con intensità differenti in momenti specifici all'interno del periodo considerato).

Per verificare la non-stazionarietà di una serie temporale un modo molto semplice è quello di utilizzare il **ADF – Augmented Dickey-Fuller test**. Questo test, appartiene alla categoria dei test che consentono di verificare la presenza o meno di **unit-root**, ovvero quanto l'andamento di una serie è definito da un trend, e quindi se possiede un'andamento che varia con il tempo. [10]
Il test considerato si articola nelle seguenti ipotesi statistiche:

- *Ipotesi nulla*: la serie considerata presenta caratteristiche non-stazionarie (unit-root)
- *Ipotesi alternativa*: la serie considerata non presenta unit-root, e quindi può essere definita stazionaria

L'esito di questo test può essere affermato a seguito del calcolo del p-value di questo test statistico, e tramite la seguente valutazione può essere stabilito il tipo di serie analizzato:

- p-value > 5% : non si può rifiutare l'ipotesi nulla e quindi la serie è definita non-stazionaria
- p-value ≤ 5% : vi è sufficiente evidenza statistica per rifiutare l'ipotesi nulla e affermare la stazionarietà della serie considerata

Siccome le serie temporali riferite a dati reali spesso non risultano stazionarie, come le serie appartenenti al settore finanziario, è possibile convertirle in maniera opportuna per soddisfare l'assunzione di stazionarietà.

Alcune delle tecniche più adottate per raggiungere questo fine sono:

- **sottrarre** i singoli valori della serie con quelli precedenti in modo consecutivo
- **rimozione** del **trend** e della **seasonality** riferiti alla serie considerata

Una volta accertata la stazionarietà delle serie è possibile procedere con l'applicazione del Granger-causality test.

Nel caso presente, si considera di voler verificare se la serieB risulta utile nella previsione dei valori della serieA, si assume:

$$X = Serie_A \qquad Y = Serie_B$$

Si considera $\mathbf{x_t}$ il valore di x al tempo t

$$x_t = a_1 \cdot x_{t-1} + a_2 \cdot x_{t-2} + ... + a_n \cdot x_{t-n} + errore_t$$

quest'ultimo può essere descritto come l'**autoregressione univariata** della variabile X, dove si considerano come valori significativi utili alla previsione:

- un coefficiente di errore al tempo t
- i valori precedenti della variabile X, fino ad lag temporale n, pesati con determinati coefficienti.

Ciascun valore con un determinato ritardo temporale viene mantenuto all'interno dell'autoregressione indicata se risulta significativo sulla base di un test statistico di tipo t.

Inoltre, tenendo presente l'obiettivo iniziale del test, si tenta di estendere l'autoregressione considerata utilizzando valori passati della variabile Y. Dunque si procede considerando e provando ad aggiungere anche valori passati della seconda variabile, verificando se tutto questo soddisfa un test statistico di tipo F. Alla fine di tutto ciò potrebbe risultare che:

- nessun valore passato di Y potrebbe risultare utile alla previsione di $\mathbf{x_t}$ –> la variabile Y non risulta significativa nella previsione dei valori della variabile X
- oppure determinati lag temporali compresi tra un lag p e q possano essere importanti per la previsione di $\mathbf{x_t}$ –> la variabile Y risulta significativa nella previsione dei valori della variabile X

Di seguito viene riportato un esempio di autoregressione univariata estesa come descritta nel secondo caso:

$$x_t = a_1 \cdot x_{t-1} + a_2 \cdot x_{t-2} + ... + a_n \cdot x_{t-n} + b_c \cdot y_{t-c} + ... + b_d \cdot y_{t-d} + errore_t$$

Nel momento in cui si intendesse verificare se diverse variabili, ad esempio Y, W, … Z risultino utili nella previsione di X, si procede in modo analogo a quanto illustrato fino ad ora, ed è possibile che l'autoregressione univariata per ottenere x_t possa divenire ancora più estesa perché potrebbe includere valori con determinati lag temporali anche riferiti alle altre variabili.

Considerando 2 variabili X e Y, in cui si intende verificare se Y risulti utile per prevedere X, è possibile riassumere il Granger-Causality test come un test statistico caratterizzato da:

- ***Ipotesi nulla*** (H_0): X non è granger-caused da Y sse valori passati di Y non contribuiscono a migliorare la previsione di X
- ***Ipotesi alternativa*** (H_1): X è granger-caused da Y

L'esito di questo test può essere affermato a seguito del calcolo del p-value di questo test statistico, e quindi se:

- p-value > 5% : non si può rifiutare l'ipotesi nulla e quindi la variabile X non è granger-caused da Y
- p-value ≤ 5% : vi è sufficiente evidenza statistica per rifiutare l'ipotesi nulla e affermare che X è granger-caused da Y

3.3 - Stato dell'arte dei modelli di previsione del prezzo di Bitcoin

Il presente paragrafo riporta in maniera sintetica alcune delle affermazioni più rilevanti emerse nel documento scientifico "***Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey***" [11] nel quale viene condotta un'analisi sistematica e completa dei diversi tipi di modelli che sono stati impiegati per la previsione del prezzo di Bitcoin tra il 2010 - 2020. In conclusione viene riportato anche il motivo per cui l'uso di modelli basati su reti neurali appartenenti alla categoria DL risulta essere più conveniente.

La previsione del prezzo di Bitcoin o altre criptovalute viene considerato un problema che appartiene alla categoria dei problemi di regressione. Diversi ricercatori negli ultimi anni hanno utilizzato tecniche basate su modelli statistici, ML, DL e RL per affrontare tale problema.

Numerose ricerche si sono concentrate nell'utilizzo di dati:

- di mercato (prezzi, volatilità, volumi scambiati, indicatori derivati da questi ultimi)
- relativi a transazioni (provenienti dalla blockchain)
- relativi alla wallet liveliness degli indirizzi riferiti agli exchange
- relativi al sentiment degli investitori sulla base dei messaggi postati sui social network

Alcune formulazioni di modelli statistici ed econometrici che puntano a produrre previsioni partendo da un'attenta analisi che ambisce ad identificare correlazioni tra serie temporali, e la conseguente creazione di modelli basati su ARMA, ARIMA, VAR, GARCH e diverse altre estensioni hanno permesso di avere risultati utili per lo scopo considerato.

Ma una loro formulazione più completa ed efficiente risulta difficile a causa della numerosità e della natura dei fattori che influenzano il prezzo di Bitcoin. Per tale motivo, ma considerando anche le performance migliori ottenute da modelli ML e DL differenti, ha portato l'interesse di questo ambito di ricerca a riporre in secondo piano i modelli statistici ed econometrici.

Inoltre, nei casi in cui si tentasse comunque l'utilizzo di questi ultimi modelli, andando a trasformare le serie in considerazione, e considerando come verificate le assunzioni richieste da questi ultimi, i risultati di previsione potrebbero risultare poco realistici e poco

accurati.
Il problema di regressione considerato è stato affrontato anche usando modelli in grado di individuare e utilizzare al meglio relazioni non lineari tra i dati analizzati e risultati di previsione prodotti.
Alcuni dei modelli più efficienti che hanno sovrapperformato gli altri sono basati su:

- ANN – artificial neural network;
- RF – random forest;
- SVM – support vector machine.

Modelli basati su SVM hanno prodotto previsioni molto accurate, specialmente per criptovalute come Ethereum [12], ma tali risultati sono fortemenți dipendenti dagli iperparametri utilizzati, come ad esempio il tipo di kernel functions e dalla categoria dei dati impiegati.
Invece, per quanto riguarda modelli basati su RF, questi ultimi hanno permesso di ottenere risultati migliori rispetto alle SVM, poiché non risulta necessario preoccuparsi della natura dei dati in input, che potrebbero presentare come nel caso considerato, diversi valori outliers e valori non separabili linearmente.
Per il problema in considerazione, i modelli che hanno fornito migliori previsioni sono stati quelli basati su reti neurali [11], in particolar modo reti neurali di tipo MLP (multi-layer perceptron), RNN (recurrent-neural network), GRU (gater recurrent units), LSTM (long-short term memory).
Tali risultati si ipotizza siano dovuti grazie al fatto che tali modelli:

- siano in grado di individuare meglio relazioni non lineari tra i dati di input e di output;
- non richiedono particolari assunzioni sui dati di input (legate ad eventuali distribuzioni e proprietà statistiche);
- grazie alla possibilità di definire diversi strati nascosti e l'utilizzo di funzioni di attivazioni non-lineari opportune è possibile aumentare la potenza delle reti neurali ed ottenere capacità di generalizzazione più ottimali per l'obiettivo richiesto.

3.4 - Metriche di valutazione

Il modo più efficace per comprendere l'ottimalità dei risultati ottenuti dai modelli di regressione e confrontare quindi diversi modelli consiste nell'utilizzo di metriche di errore. Alcune delle metriche più consone per questo fine, sono le seguenti:

$x_i = valore\ reale\ al\ tempo\ i$ $\qquad$ $\hat{x}_i = valore\ predetto\ al\ tempo\ i$

- MSE:

$$MSE = \frac{\sum_{i=1}^{n} \left(x_i - \hat{x}_i\right)^2}{n}$$

MSE – Mean Squared Error, noto anche come errore quadratico medio, permette di calcolare la disuguaglianza tra i valori reali e quelli predetti in termini di sommatoria della differenza quadratica degli scarti tra valori reali e le previsioni, rapportate al numero di elementi considerati. L'unità di misura in cui viene interpretato il valore ottenuto è il quadrato dell'unità di misura utilizzata per i singoli valori della variabile considerata. Per tale motivo non avendo un risultato nella stessa scala (unità di misura) della variabile iniziale, potrebbe non essere facile interpretare i risultati ottenuti.

- RMSE:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i - \hat{x}_i\right)^2}$$

RMSE – Root Mean Squared Error, come indicato dalla sua formula è semplicemente la radice quadrata del MSE. Per questo semplice motivo consente di ottenere un risultato nella stessa scala di misura (unità di misura) della variabile iniziale, e quindi permette di fornire una valutazione più immediata del modello che ha fornito le previsioni.

- MAPE:

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

MAPE – Mean Absolute Percentage Error, è una metrica di errore più utile rispetto al MSE e al RMSE, in quanto fornisce valori in percentuale, e quindi consente di effettuare un confronto immediato tra modelli diversi.

3.5 - Reti neurali di tipo RNN, LSMT e GRU

Le reti neurali di tipo RNN (Recurrent neural network) sono una tipologia specifica di rete neurale di tipo MLP (multi-layer perceptron), e sono utilizzate per risolvere diverse tipologie di problemi come: previsione di serie temporali, NLP (natural language processing) e speech and image recognition. [13]

I **perceptron** (o neuroni artificiali) che compongono queste reti sono organizzati in diversi strati, ovvero:

- uno strato di input,
- uno o più strati intermedi nascosti,
- uno strato di output

Inoltre le particolarità che caratterizzano questo tipo di reti neurali sono:

- la capacità di poter usare i risultati di un'iterazione corrente come input ulteriore insieme ai prossimi dati in ingresso per l'iterazione successiva, rappresentando così la possibilità di avere a disposizione una memoria che viene aggiornata ad ogni iterazione (durante la fase di allenamento della rete)
- utilizzano gli stessi parametri (pesi) per ogni strato della rete neurale

Nonostante le sue proprietà, le reti RNN soffrono di 3 problematiche:

- **short-term memory**: a causa del modo in cui funzionano le RNN, nel momento in cui vengono fornite serie di dati molto lunghe, queste reti tendono a ricordare informazioni viste di recente e dimenticare quelle lavorate inizialmente, trascurando così informazioni potenzialmente utili del passato.
- **exploding gradient problem**: i gradienti calcolati dagli strati finali per l'aggiornamento dei pesi verso gli strati iniziali, presentano valori molto grandi che potrebbero portare la rete in una fase di ipercorrezione dei pesi
- **vanishing gradient problem**: i gradienti calcolati mostrano valori sempre più piccoli (ad esempio tra 0 e 1), per cui le variazioni dei pesi della rete considerata risultano sempre più piccoli e insignificanti, portando quest'ultima a convergere molto più lentamente e in alcuni casi ad indicare soluzioni rappresentanti ottimi locali e non sempre globali

Per una progettazione opportuna di modelli di previsioni basati su reti neurali è necessario

considerare attenamente alcuni iperparametri:

- **l'architettura della rete neurale:** rappresentata dal numero di strati intermedi impiegati e **dalla dimensione** (ovvero il numero di neuroni usati) **per ogni strato** nascosto dalla rete neurale
- **le funzioni di attivazione**: funzioni lineari o non lineari che determinano il valore in uscita da ogni neurone artificiale. Queste funzioni vengono applicate sulla sommatoria dei pesi w_i ricevuti da ciascuna connessione in ingresso verso il neurone considerato
- **il learning rate**: iperparametro molto importante che determina di quanto aggiornare i pesi della rete neurale ad ogni iterazione
- **l'algoritmo di ottimizzazione**: algoritmo impiegato per l'aggiornamento dei pesi della rete neurale (alcuni esempi possono essere: Adam, Gradient Descent, e altri)

Per far fronte alle difficoltà che caratterizzano le reti neurali RNN, sono state progettate 2 diverse tipologie di reti neurali: LSTM – long short term memory [14] e GRU – gated recurrent units (citazione autore). [15]

Nella tabella seguente vengono riportate alcune delle proprietà che le contraddistinguono.

LSTM	**GRU**
• consentono di sfruttare dipendenze di lungo-termine e non solamente le informazioni più recenti • grazie alla presenza di componenti come la **cell state** (che viene usata come memoria per la rete neurale) e **3 porte** (input gate, output gate e forget gate), le LSTM posono imparare di dimenticare, o preservare informazioni rilevanti per future iterazioni • tramite l'uso di oppurtune funzioni per le porte / gates, durante le fasi di training, è possibile imporre di dimenticare valori sotto una certa soglia, evitando così il vanishing gradient problem	• utilizza stati nascosti (**hidden state**) e 2 porte (**reset gate e update gate**) che regolano il flusso di informazioni preservando le informazioni più rilevanti secondo determinati criteri e le funzioni utilizzate per le porte • grazie alla loro struttura più "leggera" richiedono di stimare un numero di parametri molto inferiore rispetto a quelli richiesti per le LSTM, per questo motivo risultano più veloci da allenare

Capitolo 4
Un modello ensemble basato su reti neurali di tipo LSTM e GRU

4.1 - Dati utilizzati

I dati utilizzati per progettare il modello che verrà illustrato nei paragrafi successivi, sono serie storiche con granularità giornaliera, e mostrano valori registrati nel periodo che inizia dal 17/08/2010 al 03/03/2023.

Il dataset in questione è composto da 5 serie temporali:

- Prezzo di Bitcoin
- Numero di address con una quantità di BTC ≥ 1000BTC
- Numero di address con una quantità di BTC ≥ 100BTC
- Numero di address con una quantità di BTC ≥ 10BTC
- Indicatore NUPL (Net Unrealized Profit and Loss)

Tali dati sono stati reperiti dal sito LookIntoBitcoin.com, una piattaforma che consente la consultazione dell'andamento di numerosi indicatori, progettati da ricercatori ed esperti del settore, che mostrano una specifica prospettiva dello stato di Bitcoin, facendo uso di dati disponibili dalla blockchain di questa criptovaluta.

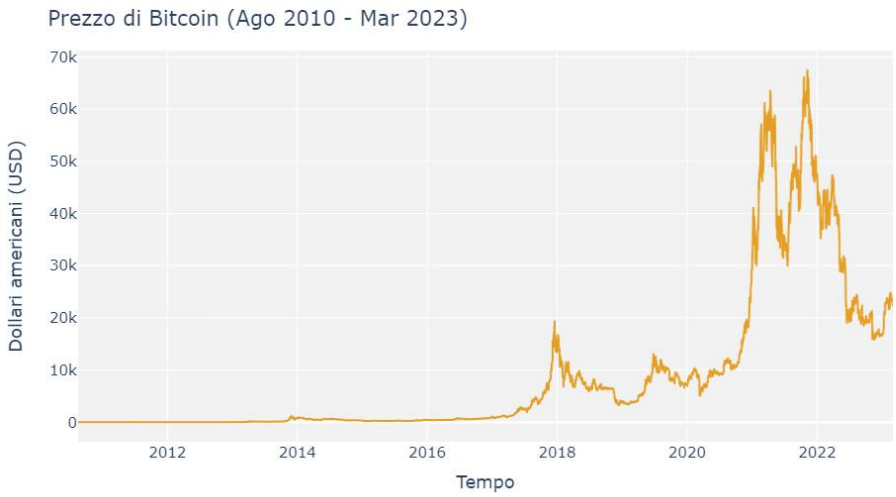


Figura 3: Prezzo di Bitcoin in USD (dollari americani) (tra Agosto 2010 e Marzo 2023)

Figura 4: Numero di indirizzi Bitcoin con un importo X ≥ 1000BTC (tra Agosto 2010 e Marzo 2023)

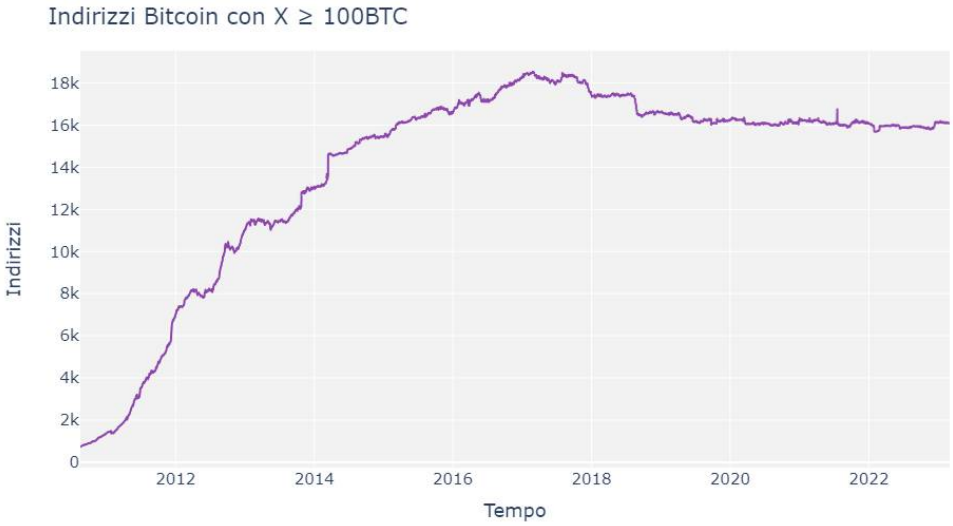


Figura 5: Numero di indirizzi Bitcoin con un importo X ≥ 100BTC (tra Agosto 2010 e Marzo 2023)

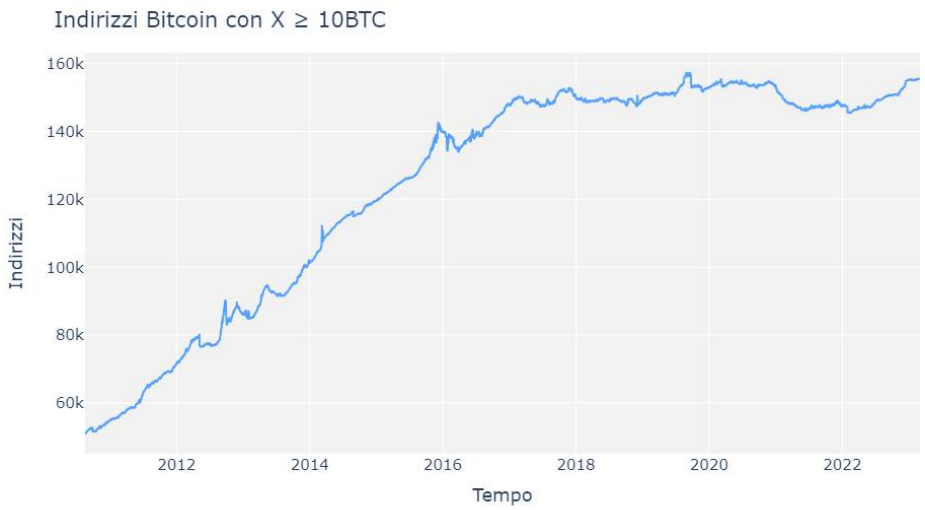


Figura 6: Numero di indirizzi Bitcoin con un importo X ≥ 10BTC (tra Agosto 2010 e Marzo 2023)

L'importanza dei dati riferiti al numero di portafogli (o indirizzi) che possiedono una determinata soglia di Bitcoin, risiede nel fatto che rispecchiano direttamente la serietà con cui individui singoli o entità più grandi investono in questa criptovaluta.

Più precisamente, un incremento o decremento in questo tipo di serie temporale indica la volontà dei vari investitori di assicurarsi di avere un pieno controllo delle criptovalute comprate, e questo trasferendole nei propri portafogli privati. In questo modo si limitano i problemi che si potrebbero riscontrare nel caso si lasciassero questi asset in gestione ad organizzazioni terze, come piattaforme CEX.

Infatti, una gestione non autonoma, ovvero il fatto di non possedere le chiavi private del portafoglio a cui sono associate le proprie criptovalute, aumenta l'esposizione a numerosi rischi. Tra questi ultimi si possono verificare problemi come:

- condivisione non autorizzata della chiave privata del portafoglio
- transazioni non autorizzate
- perdita della chiave privata da parte dell'entità che gestisce le criptovalute

Per tali ragioni, un'aumento nel numero di portafogli, e in particolare di quelli con una notevole quantità di Bitcoin associati, rappresenta in modo chiaro l'intenzione di investimento in Bitcoin, con una prospettiva di guadagno futuro.

Una primissima analisi derivata dalla semplice visione della serie temporale che rappresenta il numero di portafogli con x ≥ 1000BTC porta a dedurre una possibile

relazione particolare tra questa serie e il prezzo di Bitcoin. Tale affermazione è dovuta al fatto che i movimenti più evidenti a rialzo o al ribasso di questo indicatore si sono ripresentati con la stessa direzionalità nel prezzo di Bitcoin, anche se con un'intensità e ad una distanza temporale diversa.
Inoltre, per comprendere se veramente questo indicatore potesse causare l'andamento del prezzo, sono state considerate per completezza anche le altre serie indicate precedentemente, riferite al numero di portafogli, ma categorizzate come portafogli con una quantità di Bitcoin pari a $100BTC \leq x < 1000BTC$ e quelli con $10BTC \leq x < 100BTC$. (La motivazione di tale categorizzazione è descritta all'interno del paragrafo analisi dati)

L'indicatore **Net Unrealized Profit & Loss (NUPL)** o chiamato anche Relative Unrealized Profit & Loss, è un indicatore progettato da esperti di Adamant Capital [16] facendo uso di dati come il **market value** (o market cap) di Bitcoin e il **realized value** proveniente dalla blockchain.
Formule necessarie per il calcolo dell'indicatore NUPL:

$$marketValue = marketCap = currentBitcoinPrice \times numberOfCirculatingCoins$$

$$realizedValue = \left(\frac{\sum movedQuantity \times currentBitcoinPrice}{\sum movedQuantity}\right) \times numberOfCirculatingCoins$$

$$NUPL = \frac{marketValue - realizedValue}{marketValue}$$

Il market cap viene calcolato come il prodotto tra il numero di monete minate (prodotte) e il prezzo attuale di Bitcoin. Invece, per il realized cap s'intende il risultato ottenuto usando informazioni come la quantità di Bitcoin in uscita da un blocco, il momento in cui la transazione è stata registrata nella blockchain e il prezzo di BTC in quell'istante. In questo modo, sommando i valori delle monete spostate in un preciso istante, dagli albori di questa criptovaluta fino ad adesso, è possibile ricavare il realized cap (o realized value oppure chiamato dagli autore anche come **Realized Profit & Loss**).
L'indicatore NUPL consiste in una serie temporale, che permette di comprendere il sentiment degli investitori. Il motivo di questa intuizione, come affermano gli autori di questo indicatore, risiede nel fatto che la sottrazione tra Market cap e Realized cap mostra il profitto o la perdita non ancora "realizzata" degli investitori e quindi il valore reale di Bitcoin dal punto di vista della blockchain.

I risultati ottenuti all'interno del paragrafo relativo agli esiti dell'analisi descrittiva riguardo a queste serie temporali, illustrano la relazione di questi indicatori con il prezzo di Bitcoin e la loro utilità per lo sviluppo di modelli di previsione.

4.2 - Tools e librerie utilizzate

Gli esperimenti svolti per completare le ricerche e la creazione del modello ensemble possono essere articolati in modo semplificato nella seguente maniera:

- Download dei dati necessari per le analisi e per i modelli
- Analisi descrittiva e studio approfondito tramite metodi statistici dei dati scaricati
- Preparazione ed implementazione dei modelli basati su reti neurali per la realizzazione di un modello ensemble

Per le attività appena elencate è stato utilizzato il linguaggio di programmazione Python.
Inoltre, data la disponibilità di numerose librerie per implementare i vari steps, sono state utilizzate le seguenti librerie, con il relativo utilizzo:

- beautiful soup (usata per lo scraping e il parsing di pagine web in oggetti python),
- pandas (usata per la manipolazione di dati in forma tabellare),
- numpy (usata per la manipolazione di numerosi dati numerici e calcolo di statistiche univariate),
- matplolib, seaborn, plotly (usate per la realizzazione di grafici personalizzabili)
- scipy, scikit-learn (usata per il calcolo statistiche univariate),
  pickle (usata per salvare i modelli preparati su filesystem)
  tensorflow, keras (usate per la creazione di modelli basati su reti neurali)

4.3 - Dall'analisi al modello: Mappa complessiva degli steps

Le analisi e le fasi implementative che hanno portato al modello descritto nel presente trattato, sono riassumibili come segue:

1 Download e preparazione dei dati in file con formato .csv
   - Sono state reperite le serie temporali oggetto di analisi dalla piattaforma LookIntoBitcoin, e organizzate all'interno di un file csv per gli steps successivi

2 Analisi dati:

2.1 Analisi esplorativa:

- Sono stati realizzati grafici per mostrare ed analizzare l'andamento delle serie così come sono state ottenute, e poi suddividendole in categorie, dove ciascuna di esse rappresenta il numero di portafogli con una quantità compresa all'interno di un range ben definito di Bitcoin
- Sono stati realizzati grafici per l'analisi della distribuzione delle singole serie
- Sono state calcolate e analizzate **statistiche descrittive univariate** (media, varianza, deviazione standard, curtosi, asimmetria, quantili, valori massimi e valori minimi) per le varie serie

2.2 Analisi della correlazione e della causalità:

- Sono state calcolate e analizzate **statistiche descrittive multivariate**, come il coefficiente di correlazione tra le varie coppie di serie
- È stato utilizzato il **Granger-Causality test** tra coppie di serie, per comprendere se considerata una serie temporale con un certo numero di lag, questa potesse influenzare l'altra serie considerata. L'obiettivo di tale test è stato quello di individuare le serie che presentavano un'influenza sulll'andamento del prezzo di Bitcoin e il numero di lag dopo il quale è possibile affermare ciò.

3 Preparazione e valutazione di modelli basati su reti neurali (LSTM e GRU)

3.1 Sono stati preparati e valutati 4 modelli di tipo **LSTM e GRU** con capacità predittive di tipo **one-step ahead**. Per ciascun modello, sono state considerate come variabili di input, un gruppo ben definito di serie tra quelle oggetto di studio della presente trattatazione e come variabile target è stato considerato il

prezzo di Bitcoin.

Le fasi in cui si articola la preparazione dei singoli modelli, consistono in:

- **splitting** dei dati in 3 set di dati (training, validation e test set),
- **scaling** dei dati di ogni set, usando scalers di tipo **MinMaxScaler**,
- preparazione dei set di input in **tensor,** per poterli fornire alle reti neurali,
- training, monitoraggio delle performance e salvataggio dei modelli preparati,
- analisi delle performance ottenute durante la fase di training dei modelli
- valutazione delle previsioni ottenute usanto metriche di errore come **MSE** (mean squared error), **RMSE** (root mean squared error), **MAPE** (mean absolute percentage error)

1.2 Sono state calcolate e confrontate le previsioni dei modelli ottenuti, con la serie relativa al prezzo con un lag temporale di 1

4 Preparazione e valutazione del modello ensemble

4.1 È stato realizzato un modello ensemble, che calcola le previsioni del prezzo di Bitcoin e un range di confindenza, utilizzando le previsioni fornite dai 4 migliori modelli ottenuti nella fase 3

4.2 Sono stati valutati i possibili modi in cui il modello ensemble può essere impiegato per effettuare scelte di investimento, in particolare tramite il **backtesting** di semplici strategie basate sulle previsioni fornite dal modello e il confronto con la strategia di tipo Buy-and-Hold.

4.4 - Esiti dell'analisi descrittiva

Le serie storiche descritte nel paragrafo **Dati utilizzati** sono state analizzate a seguito di una prima fase di preparazione.

Tale operazione è stata eseguita in particolare sulle serie rappresentanti il numero di indirizzi con un certo quantitativo di Bitcoin. Siccome le serie erano disponibili sulla piattaforma citata precedentemente come dati aggregati, ovvero:

- # indirizzi con x ≥ 1000BTC
- # indirizzi con x ≥ 100BTC
- # indirizzi con x ≥ 10BTC

per analizzare e comprendere correttamente l'andamento di deteminate serie senza includere valori appartenenti alle serie già considerate è stato necessario fare una semplice separazione e ricategorizzazione di queste ultime.

Le serie considerate sono state quindi le seguenti:

- addr_gte_1000BTC = # indirizzi con x ≥ 1000BTC
- addr_100_999BTC = # indirizzi con 100BTC ≤ x < 1000BTC
  ← (# indirizzi con x ≥ 100BTC) - addr_gte_1000BTC
- addr_10_99BTC = # indirizzi con 10BTC ≤ x < 100BTC
  ← (# indirizzi con x ≥ 10BTC) - (# indirizzi con x ≥ 100BTC)

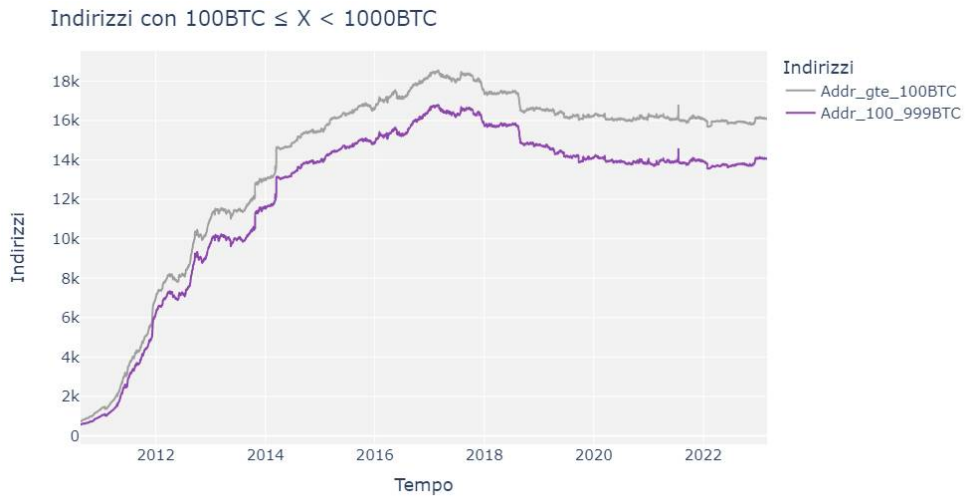


Figura 7: Confronto tra le serie rappresentanti il numero di indirizzi di Bitcoin aventi come importi X ≥ 100BTC e aventi 100 ≤ X < 1000BTC (tra Ago 2010 e Mar 2023)

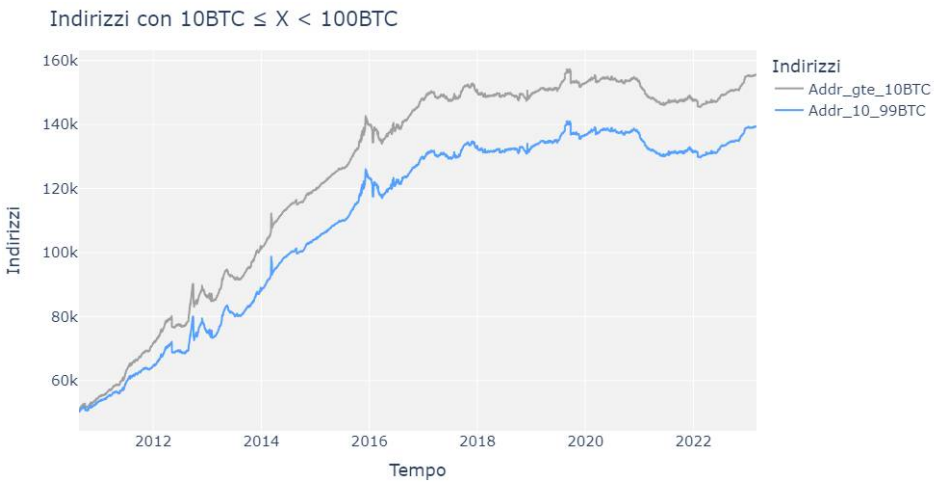


Figura 8: Confronto tra le serie rappresentanti il numero di indirizzi Bitcoin aventi come importi X ≥ 10BTC e aventi 10BTC ≤ X < 100BTC (tra Ago 2010 e Mar 2023)

NOTA: Le nomenclature usate per la seconda e terza categoria sono state scelte per semplificare il modo in cui chiamare i range dei numeri considerati. Infatti, potendo avere transazioni Bitcoin con importi fino a 8 cifre decimali, per evitare di scrivere addr_100_999.99999999BTC è stato scelto di usare nomi più abbreviati come mostrato all'inizio di questo capitolo.

Come illustrato nelle immagini soprastanti, le serie iniziali mostrano un andamento diverso da quello delle serie che rappresentano il numero di portafogli con importi presenti in range ben definiti.

Considerando questi dati, insieme al prezzo e l'indicatore NUPL, sono state calcolate statistiche descrittive univariate. Per i risultati ottenuti si notano valori numerici particolarmente insoliti da un punto di vista statistico, ma tipici dei mercati finanziari, motivo per cui possono essere definti poco sorprendenti.

	Price	Addr_gte_1000BTC	Addr_100_999BTC	Addr_10_99BTC	NUPL
count	4582.0000	4582.0000	4582.0000	4582.0000	4582.0000
mean	8843.4424	1631.2326	12344.7689	110664.2547	0.3206
std	14531.3883	530.4999	4220.8444	28516.7684	0.3139
min	0.0600	159.0000	569.0000	50231.0000	-1.5196
25%	140.2050	1444.0000	10430.0000	83241.2500	0.1498
50%	833.4350	1728.0000	13916.0000	127401.5000	0.3964
75%	9717.4300	2103.7500	14825.7500	132758.5000	0.5482
max	67492.0000	2490.0000	16794.0000	141056.0000	0.8621

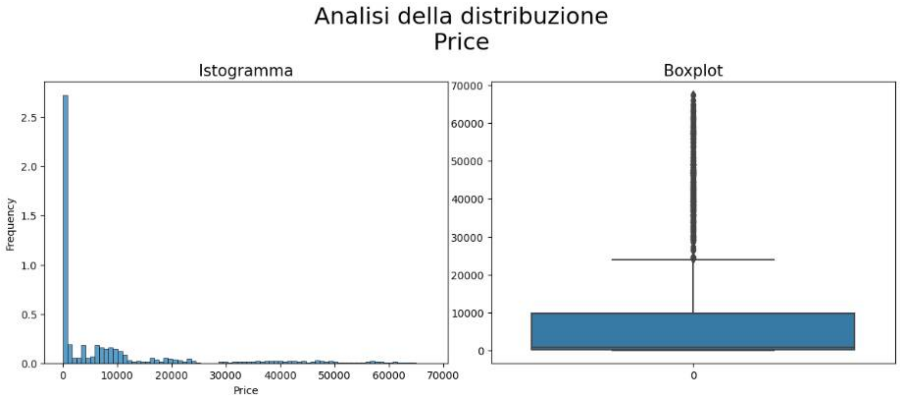
Figura 9: Statistiche descrittive univariate- 1

Price		Addr_gte_1000BTC		Addr_100_999BTC	
Media:	8843.4424	Media:	1631.2326	Media:	12344.7689
Varianza:	211115160.0882	Varianza:	281368.7547	Varianza:	17811638.8735
DevStds:	14529.8025	DevStds:	530.442	DevStds:	4220.3837
Asimmetria:	2.0509	Asimmetria:	-1.0009	Asimmetria:	-1.5206
Curtosi:	3.3979	Curtosi:	0.293	Curtosi:	1.2752
Addr_10_99BTC		**NUPL**			
Media:	110664.2547	Media:	0.3206		
Varianza:	813028604.0336	Varianza:	0.0985		
DevStds:	28513.6564	DevStds:	0.3139		
Asimmetria:	-0.7826	Asimmetria:	-1.2067		
Curtosi:	-0.9126	Curtosi:	2.3926		

Figura 10: Statistiche descrittive univariate - 2

Grazie ai valori di queste statistiche e all'utilizzo di grafici di tipo boxplot (i quali aiutano a mostrare graficamente la suddivisione in quantili dei valori appartenenti alle serie) e istogrammi (per comprendere le distribuzioni), si comprende subito che le serie analizzate mostrano:

- range di valori molto ampi
- distribuzioni fortemente asimmetriche e con code molto lunghe
- una fortissima presenza di valori anomali (outliers)
- distribuzione molto lontane da quella Normale. Per questo motivo possono essere definibili come distribuzioni di tipo leptocurtica (curtosi > 3) nel caso del prezzo e distribuzioni di tipo platicurtiche (curtosi < 3) per tutte le altre serie



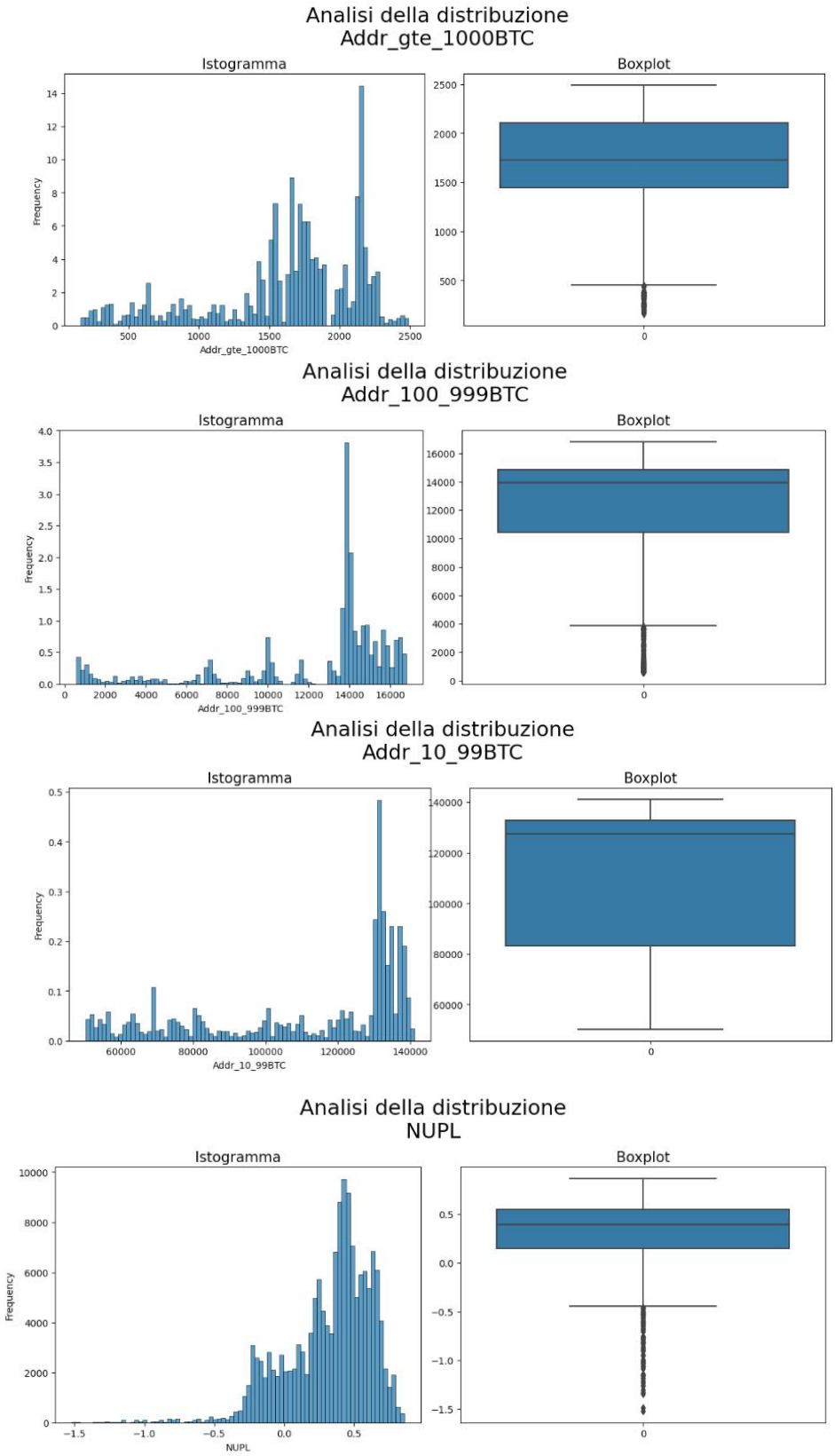


Figura 11: Distribuzioni e boxplot relativi alle serie analizzate

Tali descrizioni sono comunque poco interessanti, dato che le serie temporali riferite a mercati finanziari, soprattutto nel mercato delle criptovalute, a causa dei numerosi fattori da cui sono influenzate, tra cui anche l'operatività e stile di investimento dei vari operatori che possono portare a momenti di bassa oppure alta volatilità (ovvero fluttuazioni sempre più ripide).

Un'analisi che considera le serie in modo congiunto, ha permesso di analizzare la presenza di correlazioni e relazioni di tipo causa-effetto.
In particolare, sono state calcolate ed analizzate le correlazioni tra il prezzo e le altre serie utilizzando i 3 metodi illustrati nel capitolo 2

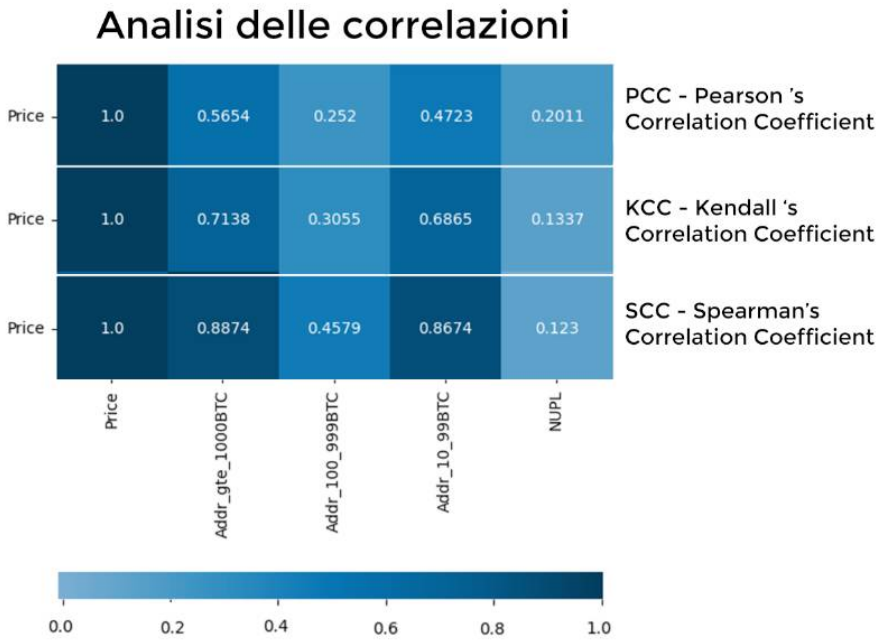


Figura 12: Coefficienti di correlazione calcolati con PCC, KCC, SCC tra prezzo di Bitcoin e le altre serie analizzate

Da tali risultati si evince che:

- Le serie con una correlazione positiva forte con il prezzo di Bitcoin sono quelle riferita a Addr_gte_1000BTC e Addr_10_99BTC.
- L'unica serie con una correlazione positiva debole con il prezzo di Bitcoin è stata quella riferita all'indicatore NUPL

Questi esiti però non fornisco un'indicazione chiara su quali possano realmente essere le

variabili che possono causare effettivamente certi andamenti del prezzo di questa criptovaluta. Tale affermazione è dovuta al fatto che la correlazione tra serie temporali, non permette di concludere la presenza di una relazione di causalità.
Per proseguire con le ricerche di una possibile relazione di tipo causa-effetto è stato utilizzato il **Granger-Causality test (GC test)**. Un test statistico dove considerando 2 serie temporali A e B stazionarie e volendo verificare se la serie A influenza B, si considerano:

- ipotesi nulla: la seria B non è causata da A
- ipotesi alternativa: la serie B è G**ranger-caused** da A

Per una spiegazione più dettagliata del funzionamento di questo test, si rimanda alla sezione specifica del capitolo 2.

Un'assunzione importante da tenere in considerazione prima di procedere con il test, è quello di verificare la stazionarietà delle serie.
A tale scopo è stato impiegato sulle serie in analisi il test Augmented Dickey Fuller test (per una spiegazione dettagliata si rimanda alla sezione dedicata del capitolo 2), il quale ha permesso di rilevare le serie non stazionarie e di poterle prepare in modo opportuno. (valore del p-value considerato poter rifiutare l'ipotesi nulla e confermare la stazionarietà: $x \leq 5\%$)

Serie considerata	P-value ottenuto dal ADF-test	Esito
Prezzo	44,97%	Non stazionaria
Addr_gte_1000BTC	< 5%	Stazionaria
Addr_100_999BTC	< 5%	Stazionaria
Addr_10_99BTC	7,5%	Non stazionaria
NUPL	< 5%	Stazionaria

Per poter proseguire, è stata utilizzata una tipica operazione usata per rendere una serie stazionaria, ovvero effettuando la sottrazione con i valori precedenti della serie stessa con un lag temporale pari a 1. In questo modo, le serie relative al Prezzo e Addr_10_99BTC sono state ripreparate e fornite nuovamente per il ADF-test, e hanno permesso di raggiungere i seguenti valori:

Serie considerata	P-value ottenuto dal ADF-test	Esito
Prezzo	< 5%	Stazionaria
Addr_10_99BTC	< 5%	Stazionaria

L'implementazione relativa al Granger Causality test che è stata impiegata (presente nella libreria Python statsmodels) forniva esiti utilizzado 4 test statistici differenti. Per questioni di semplicità, ma anche considerando che le formulazioni iniziali dell'autore di questo test adottavano una dimostrazione basata su una statistica di tipo F, sono stati presi in considerazione gli esiti dei test statistici che utilizzavano questo tipo di distribuzione statistica (ad esempio il params_ftests).

In conclusione grazie al GC-test, considerando le serie analizzate a coppie sono emerse le seguenti conclusioni:

- La serie addr_gte_1000BTC mostra un'influenza molto evidente sul prezzo (se si considerano come valori i lag temporali dal 43 poi)

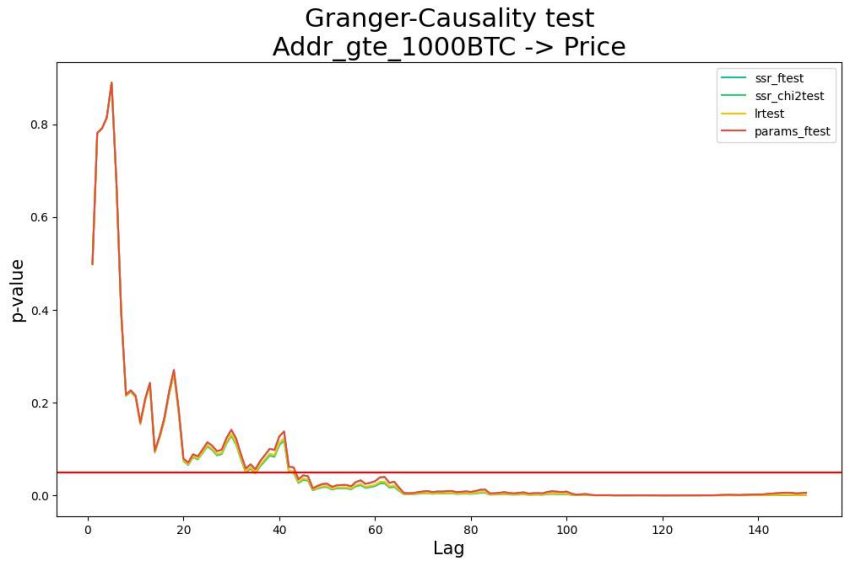


Figura 13: Granger Causality test applicato alle serie Addr_gte_1000BTC e prezzo di Bitcoin

- La serie addr_100_999BTC anche se presenta una correlazione debole con il prezzo, mostra un'influenza su di esso considerando valori di lag da 97 in poi.

- La serie addr_10_99BTC e quella riferita all'indicatore NUPL non hanno mostrato alcuna influenza sul prezzo di Bitcoin

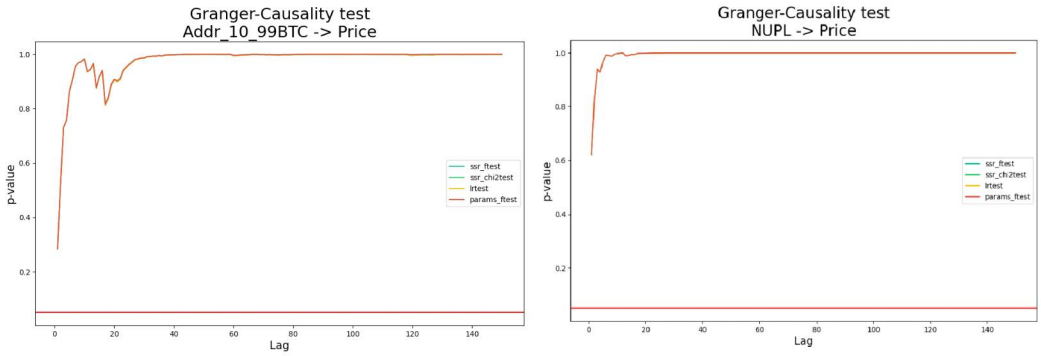


Figura 14: Granger Causality test applicato alle serie Addr_10_99BTC, NUPL e prezzo di Bitcoin

- L'indicatore NUPL influenza in modo significativo la serie addr_gte_1000BTC dopo un lag temporale che si attesta pari a 51 giorni.

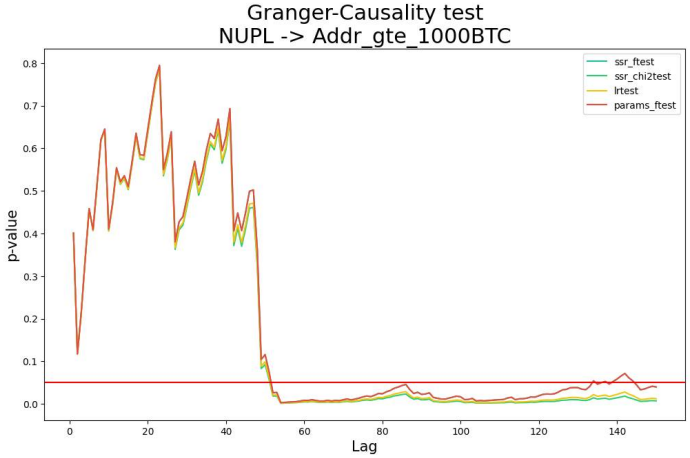


Figura 15: Granger Causality test applicato alle serie NUPL e Addr_gte_1000BTC

Considerando i risultati appena indicati e prendendo nuovamente in considerazione l'obiettivo iniziale, ovvero realizzare un modello in grado di prevedere il prezzo di Bitcoin, sono stati effettuati i seguenti acccorgimenti:

- Le serie addr_gte_1000BTC e addr_100_999BTC sono state tenute in considerazione per la preparazione dei modelli
- È stata scartata la serie addr_10_99BTC, perché non presentava alcuna relazione di causalità con il prezzo
- Nonostante il risultato del Granger-Causality test tra la serie del prezzo e l'indicatore NUPL, quest'ultima serie è stata comunque tenuta in considerazione poiché presentava una relazione di causalità interessante sulla serie

addr_gte_1000BTC (la quale a sua volta mostrava un risultato molto promettente per il GC test con il prezzo).

4.5 - Modelli LSTM e GRU

Sulla base dei risultati ottenuti nella sezione “esiti analisi descrittiva” e per le prestazioni considerevoli mostrate in diversi documenti scientifici (come riportato nel capitolo 2) dalle reti neurali ricorrenti, sono state scelte 2 categorie specifiche di algoritmi per la preparazione dei modelli: reti neurali di tipo Gated Recurrent Unit (GRU) e reti neurali di tipo Long Short Term Memory (LSTM).

La preparazione dei modelli i cui risultati verranno ampiamente discussi, è possibile articolarla in maniera sistematica come segue:

- Individuazione della categoria del modello da realizzare: selezione dei dati da utilizzare come input per prevedere il prezzo di Bitcoin
- Preparazione dei dati da fornire in ingresso al modello: divisione dei dati in 3 dataset (splitting), trasformazione dei dati di ciascun set all’interno di una scala di valori tra 0 e 1 (scaling) e infine la prepazione dei tensor da fornire alle singole reti neurali
- Scelta dei layer da impiegare nel modello
- Training della rete neurale: istanziazione della rete neurale con i layer stabiliti, e indicando gli iperparametri necessari per la sua fase di training (funzione di ottimizzazione, learning rate, numero epoche, bach_size) in base agli esiti ottenuti dall’analisi descrittiva
- Scelta del modello migliore dalla fase di training: valutazione effettuata misurando il MSE (mean standard error) ad ogni epoca e scegliendo il modello che presenta un valore per questa metrica di errore più basso possibile.
- Test di previsione: utilizzo del modello preparato per effettuare previsioni sul validation set e sul test set, trasformazione dei valori predetti (unscaling) per poter effettuare un confronto diretto con il prezzo e valutazione delle previsioni calcolando metriche di errore come MSE, RMSE, MAPE

Elenco delle caterogie dei modelli preparati

Nel caso considerato, una categoria di modelli rappresenta un gruppo di modelli per cui sono state considerate le stesse variabili di input, la stessa variabile di output e gli stessi iperparametri.

L'unico aspetto che contraddistingue i modelli della stessa categoria è il tipo di rete neurale utilizzata, in particolare ci si riferisce al tipo strato utilizzato per il primo layer.

Negli esperimenti considerati sono state preparate per ogni categoria, una rete neurale di tipo LSTM e una di tipo GRU.

Nella seguente tabella i data tensor presentano una struttura pari a (X , 43, Y).

X = numero di elementi usati per il data set considerato

Y = numero di colonne / variabili di input considerate

Dunque le categorie di modelli realizzate sono:

Categoria	**Dettagli**
Categoria 1	• Input: data tensor con struttura (X, 43, 3), dove le colonne considerate sono: Prezzo_BTC, Addr_gte_1000BTC, Addr_100_999BTC • Output: prezzo di Bitcoin al 44° giorno • Layer impiegati per le reti neurali di questa categoria: ◦ GRU \ LSTM con 64 nodi e funzione di attivazione 'relu' ◦ strato denso con 32 nodi e funzione di attivazione 'elu' ◦ strato denso per l'output con 1 nodo e funzione di attivazione 'linear'
Categoria 2	• Input: data tensor con struttura (X, 43, 1), dove le colonne considerate sono: Prezzo_BTC • Output: prezzo di Bitcoin al 44° giorno • Layer impiegati per le reti neurali di questa categoria: ◦ GRU \ LSTM con 64 nodi e funzione di attivazione 'relu' ◦ strato denso con 32 nodi e funzione di attivazione 'elu' ◦ strato denso per l'output con 1 nodo e funzione di attivazione 'linear'

Categoria 3

- Input: data tensor con struttura (X, 43, 2), dove le colonne considerate sono: Prezzo_BTC, Addr_gte_1000BTC
- Output: prezzo di Bitcoin al 44° giorno
- Layer impiegati per le reti neurali di questa categoria:
  - GRU \ LSTM con 64 nodi e funzione di attivazione 'relu'
  - strato di dropout con % di di elementi rimossi 5%
  - strato denso con 32 nodi e funzione di attivazione 'elu'
  - strato denso per l'output con 1 nodo e funzione di attivazione 'linear'

Categoria 4

- Input: data tensor con struttura (X, 43, 2), dove le colonne consideratesono: Prezzo_BTC, NUPL
- Output: prezzo di Bitcoin al 44° giorno
- Layer impiegati per le reti neurali di questa categoria:
  - GRU \ LSTM con 64 nodi e funzione di attivazione 'relu'
  - strato di dropout con % di di elementi rimossi 5%
  - strato denso con 32 nodi e funzione di attivazione 'elu'
  - strato denso per l'output con 1 nodo e funzione di attivazione 'linear'

Preparazione dei dati per le varie categorie di modelli

I dati di partenza, ovvero i dati giornalieri registrati tra il 2010-08-17 al 2023-03-03, rappresentavano un totale di 4582 righe e 4 (serie) colonne.

Questi ultimi, per la preparazione dei modelli di ciascuna categoria, sono stati opportunamente suddivisi (**splitting**) in 3 set di dati:

- 85% training set = 3894 righe di dati
- 10% validation set = 458 righe di dati
- 5% test set = 230 righe di dati

Il motivo per cui è stato scelto un training set così ampio è spiegato dal fatto che si voleva preparare i modelli, mostrando movimenti molto violenti nel prezzo di Bitcoin, come quelli che si sono presentati tra Nov 2020 e Apr 2021.

Inoltre, per facilitare e velocizzare l'apprendimento delle reti neurali e ridurre la possibilità di rimanere fermi in un punto di ottimo locale durante il processo di ottimizzazione che

avviene durante la fase di training è stato scelto di effettuare la trasformazione (**scaling**) dei dati per ogni data set considerato.
Per tale ragione sono stati utilizzati scalers di tipo MinMax, che una volta applicati alla serie considerata, la portano dal range di partenza all'interno di un range di valori tra 0 e 1.

Preparazione di ciascuna rete neurale

Per una corretta preparazione delle reti neurali, sulla base di diverse prove effettuate e gli esiti dell'analisi descrittiva è stato scelto di utilizzare:

1 data tensor in ingresso per ciascuna rete neurale pari a (X, 43, Y) come spiegato in precedenza, poiché è emerso dal Granger Causality test una relazione di causalità interessante tra Addr_gte_1000BTC sul prezzo di Bitcoin da un lag temporale dal valore 43 in poi
2 Come metrica di errore da minimizzare: MSE (utilizzata per allenare e considerare i modelli migliori)
3 i seguenti iperparametri:
   - learning rate = 0.0001 (ovvero un valore molto basso, per cercare di cogliere i miglioramento del modello in fase di training anche se minimi e nonostante questo possa comportare una durata potenzialmente più lunga per il training)
   - batch size = 10 (numero di input presi in considerazione con la struttura (43, Y) che vengono forniti per ogni epoca al modello)
   - epoche = 100 (numero di volte per cui continuare ad allenare il modello prima di fermare la fase di training)
3 una callback functions per fermare la fase di training nel momento in cui il modello non riuscisse a migliorare e quindi a ridurre l'errore MSE nelle 20 epoche successive alla rilevazione del valore migliore
4 una callback functions per salvare / sovrascrivere il modello migliore trovato in base al valore MSE rilevato sul validation set

Nella tabella seguente vengono riportate le performance osservate durante la fase di training dei modelli preparati. Inoltre per effettuare un confronto più completo, sono state calcolate anche le metriche di errore di un ipotetico utilizzo del prezzo con lag temporale pari a 1, per prevedere il prezzo stesso.

	Epoca	Scaled val_MSE	Unscaled val_RMSE	Unscaled val_MAPE	Unscaled test_RMSE	Unscaled test_MAPE	Tempo training
Gru_c1	72	0,0112	6401,15	12,1%	936,45	3,82%	12m
Lstm_c1	6	0,0209	7152,21	13,87%	1212,88	4,33%	4m
Gru_c2	57	0,001	1654,74	3,23%	557,55	1,87%	11m
Lstm_c2	89	0,0013	1964,33	3,98%	641,94	2,21%	12m
Gru_c3	69	0,0053	4969,83	10,33%	1010,69	4,08%	11m
Lstm_c3	97	0,0052	4098,53	7,8%	999,75	3,57%	13m
Gru_c4	24	0,0013	2354,39	4,84	787,73	3,06	7min
Lstm_c4	98	0,0017	2056,98	4,13%	746,2	2,42%	16min
lag1_btc					547,01	1,76%	

Per ciascuna categoria, sono stati scelti i modelli con i migliori risultati (valore più basso per la metrica di errore) ottenuti in termini di MAPE, questo poichè essendo una metrica di errore indicata in percentuale, consente di effettuare un confronto immediato tra modelli che usano dati diversi o che funzionano in modo completamente differente.
Le percentuali ottenute per il MAPE, dimostrano una capacità di previsione molto buona per i tutti i modelli ottenuti, con valori di errore al di sotto del 5%.

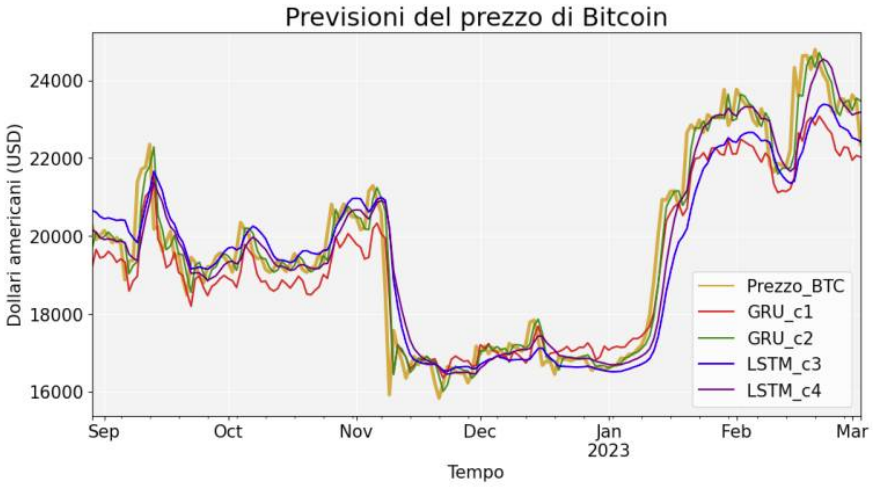


Figura 16: Previsioni ottenute dai migliori modelli basati su LSTM e GRU per ciascuna categoria

Sebbene considerare la serie con un lag di 1 (lag1_btc) permette di avere una previsione del prezzo con un MAPE pari a 1.76%, è importante ricordare che Bitcoin è in realtà influenzato da diversi fattori, per cui considerare solamente la stessa serie traslatata

orizzontalmente di 1, potrebbe non fornire una visione chiara della direzione del prezzo e quindi risultare in cattive decisioni di investimento.

Una soluzione che si ritiene possa risultare particolarmente utile, è l'utilizzo dei migliori modelli ottenuti basati su reti neurali, illustrati fino ad ora, per realizzare un modello ensemble, in grado di fornire una previsione tenendo in considerazione serie diverse, oltre al prezzo stesso.

4.6 - Preparazione e valutazione del modello ensemble

La realizzazione di un modello predittivo ensemble per il prezzo di BTC è stata effettuata a partire dai 4 migliori modelli illustrati nella sezione precedente, ovvero uno per categoria:

- Categoria 1: Modello RNN – GRU (nome modello: GRU_c1)
  Input: prezzo di BTC, addr_gte_1000BTC, addr_100_999BTC
  Output: prezzo di BTC (44° giorno)
- Categoria 2: Modello RNN – GRU (nome modello: GRU_c2)
  Input: prezzo di BTC
  Output: prezzo di BTC (44° giorno)
- Categoria 3: Modello RNN – LSTM (nome modello: LSTM_c3)
  Input: prezzo di BTC, addr_gte_1000BTC
  Output: prezzo di BTC (44° giorno)
- Categoria 4: Modello RNN – LSTM (nome modello: LSTM_c4)
  Input: prezzo di BTC, NUPL
  Output: prezzo di BTC (44° giorno)

Per comprendere i modelli tra cui scegliere, è stato necessario analizzare le previsioni ottenute da ciascuno di essi, per il periodo riferito al test set, ovvero dal 29-08-2022 al 2023-03-09.

Dunque sono stati calcolati e valutati gli errori MAPE (mean absolute percentage error) considerando a coppie le previsioni fornite dai singoli modelli. Nella seguente matrice viene illustrato l'analisi eseguita.

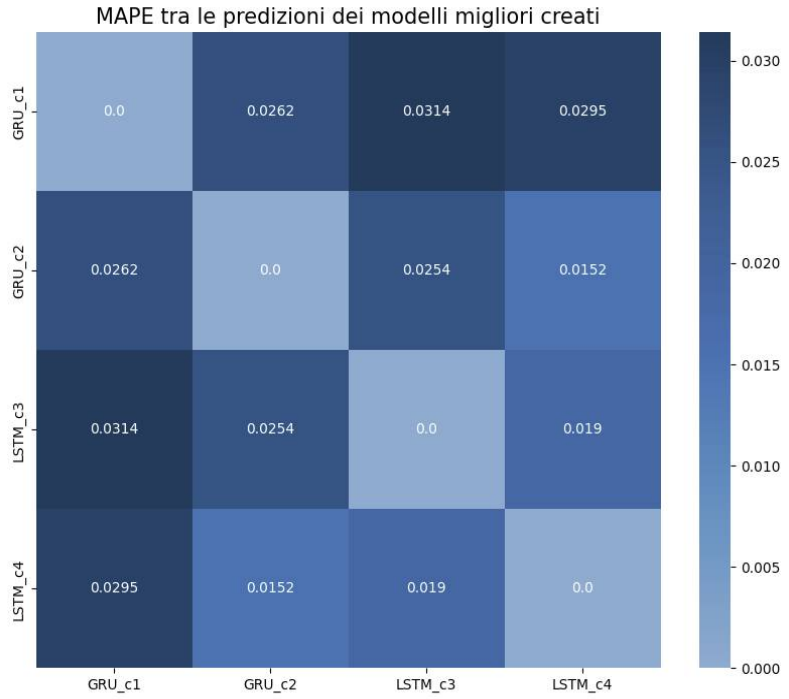


Figura 17: Matrice degli errori MAPE tra le previsioni dei migliori modelli preparati

I risultati mostrati indicano una forte similarità tra le previsioni ottenute, in quanto si discostano tra di loro con un errore massimo (MAPE) di 3.14%.

Nonostante questo esito, considerare tutti e 4 i modelli si ritiene possa risultare più efficace per la realizzazione di un modello ensemble. Tale affermazione è dovuta al fatto che una variazione violenta presente in una delle serie considerate che hanno una buona relazione di correlazione o causalità sul prezzo, siccome vengono impiegate all'interno di alcuni modelli e non in altri, potrebbero aiutare a formulare una previsione più affidabile per il prezzo futuro di Bitcoin.

Le previsioni del modello finale sono dunque realizzate calcolando i valori medi delle previsioni fornite da ciascuno dei 4 modelli ad ogni istante.

Inoltre, grazie al calcolo della deviazione standard per i valori delle previsioni sono state calcolate anche delle bande di confidenza attorno al prezzo.

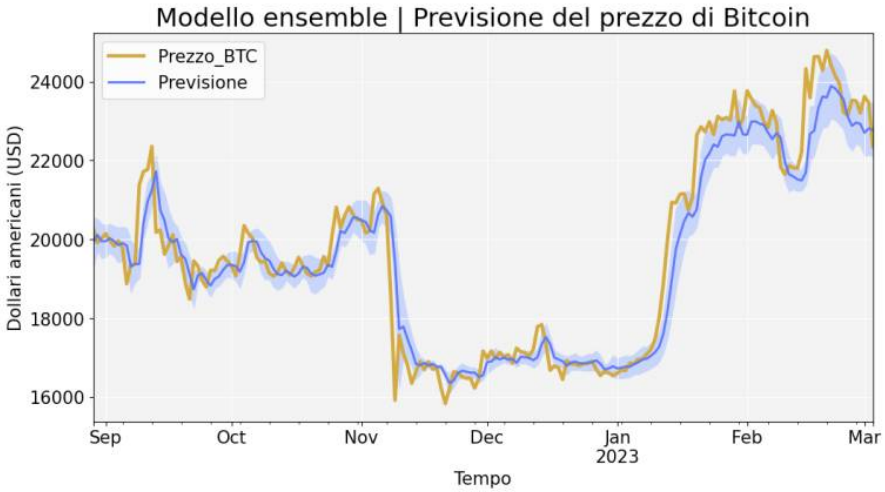


Figura 18: Previsione del prezzo di Bitcoin usando il modello ensemble

Grazie al calcolo delle metriche di valutazione su queste ultime previsioni è possibile confermare la bontà del modello ensemble realizzato.

Metrica di errore	**Valore**
RMSE	676,12
MAPE	2,17

Nella sezione successiva viene discusso il modo in cui è possibile sfruttare queste previsioni.

4.7 - Utilità del modello ensemble

Considerando il modello realizzato e le performance ottenute, un modo corretto per poter sfruttare al meglio uno strumento simile può essere :

- l'integrazione di quest'ultimo all'interno di una propria strategia di trading discrezionale, come conferma ulteriore per una propria idea di operazione che si intende eseguire
- utilizzarlo all'interno di una strategia automatizzata dove sulla base di determinate condizioni tra prezzo e la previsione stimata, si potrebbe assumere una certa posizione di investimento andando ad aprire effettivamente operazioni buy o sell

L'integrazione della previsione come indicatore all'interno di una strategia discrezionale,

non risulata facilmente stimabilie in termini di rendimenti, a causa di numerosi fattori, tra cui la psicologia del trader che sta operando e le possibili variazioni che potrebbe introdurre nella sua strategia di risk management.

Per tale motivo è stato analizzato il possibile utilizzo della previsione del modello ensemble come indicatore all'interno seguente strategia automatizzata, dove:

Posizione assunta	**Spiegazione della posizione**
LONG	si apre un'operazione BUY e si chiudono le operazioni SELL, quando il prezzo taglia al rialzo l'indicatore relativo alla previsione
SHORT	si apre un'operazione SELL e si chiudono le operazioni BUY, quando il prezzo taglia al ribasso l'indicatore relativo alla previsione

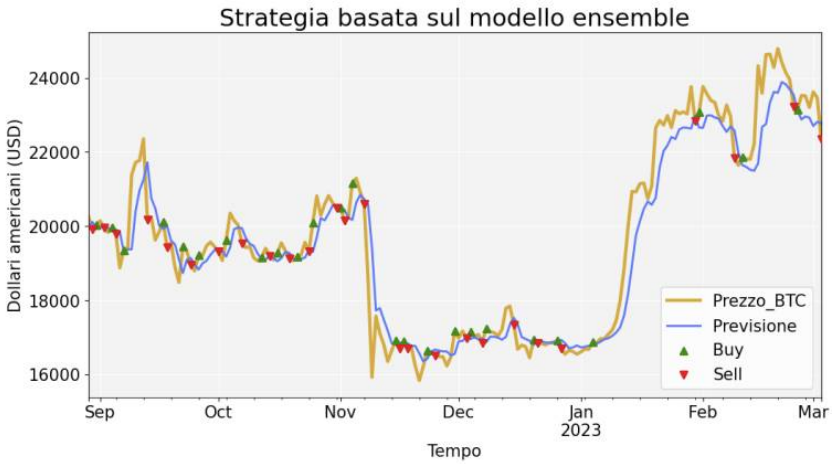


Figura 19: Utilizzo delle previsioni ensemble per individuare operazioni buy e sell

Figura 20: Confronto tra la strategia progettata e la strategia Buy and hold

La valutazione dei rendimenti per giudicare la strategia pensata, è stata effettuata senza considerare eventuali commissioni di apertura, chiusura e permanenza per gli ordini effettuati.
Come strategia di riferimento è stata considerata una tipica strategia di tipo “buy and hold”, nella quale viene effettuata un’unica operazione all’inizio del periodo, che viene chiusa al termine del periodo considerato.

Time	Buy_and_Hold	Strategia
2022-08-29	1.000000	1.000000
2023-03-03	1.102713	1.133335

Figura 21: Rendimenti lordi ottenuti dalle strategie confrontate

Sebbene la strategia basata sulle previsioni fornisca un profitto lordo che si distingue semplicemente per +3.06% per il periodo considerato, si ritiene che possa risultare più efficiente e vantaggiosa rispetto alla strategia “buy and hold” perché:

- potrebbe permettere di cogliere fluttuazioni molto evidenti del prezzo di Bitcoin per ottenere profitti più alti, senza dover per forza subire in casi sfavorevoli un andamento contrario del valore del prezzo rispetto alla posizione assunta
- considerando i dati real time o semplicemente con un timeframe più piccolo relativi al prezzo di Bitcoin e mantenendo la serie delle previsioni con il timeframe giornaliero, si potrebbe percepire l’inizio di una salita o discesa (che va a tagliare l’indicatore della previsione) più rapidamente rispetto ad aspettare i dati di chiusura della giornata relativi al prezzo.

Capitolo 5
Conclusioni

Le analisi e le sperimentazioni condotte, portano a concludere una notevole utilità di dati proveniente dalla blockchain, per effettuare una stima accurata del prezzo di Bitcoin.
Il modello ensemble realizzato permette di ottenere valori di previsione del prezzo della criptovaluta considerata, che confrontati con il prezzo reale mostrano risultati molto promettenti, ovvero valori come 676.12 $ (**RMSE**) e 2,17% (**MAPE**).
Inoltre, grazie all'integrazione delle previsioni come un indicatore per l'implementazione di una strategia automatizzata, è stato possibile raggiungere risultati sorprendenti come il potenziale profitto pari a +13.33% ottenuto eseguendo il backtesting della strategia nel periodo che inizia dal 29-08-2023 sino al 03-03-2023.

Miglioramenti futuri che ambiscono ad uno scopo identico a quello del modello proposto, ovvero in grado di fornire una visione più aggiornata e accurata del prezzo di Bitcoin, si auspica che considerino oltre ai dati provenienti dalla blockchain, anche dati rappresentanti ulteriori fattori che influiscono sulle fluttuazioni del prezzo si questa criptovaluta.
Infatti il monitoraggio e la raccolta di dati in real-time sulle attività di scambio dalle piattaforme di trading e CEX, in particolare le quantità e il valore dei contratti scambiati o i prezzi a cui si richiede di eseguire determinati tipi di ordine potrebbero fornire un vantaggio ulteriore.
Sulla base di queste informazioni e la loro elaborazione tramite criteri ben definiti, le previsioni del prezzo di Bitcoin potrebbero raggiungere miglioramenti significativi in termini di accuratezza.

Bibliografia

1 : Satoshi Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, 2008

2: Bitcoin.org, How does Bitcoin Work?, 2009, https://bitcoin.org/en/how-it-works

3: Adam Hayes - Investopedia, What is a Time Series and how is it used to analyze data?, 2022, https://www.investopedia.com/terms/t/timeseries.asp

4: Zach, Correlations in Stata: Pearson, Spearman and Kendall, 2020, https://www.statology.org/correlations-stata/

5: Zach, What is Considered to Be a “Strong” Correlation?, 2020, https://www.statology.org/what-is-a-strong-correlation/

6: David Sarmento, Correlation Types and when to use them, , https://ademos.people.uic.edu/Chapter22.html

7: Jim Frost, Spearman’s Correlation Explained, , https://statisticsbyjim.com/basics/spearmans-correlation/

8: Jim Frost, Lurking Variable: Definition & Examples, , https://statisticsbyjim.com/basics/lurking-variable/

9: Granger, C. W. J., Investigating Causal Relations by Econometric Models and Cross-spectral Methods., 1969

10: D. A. Dickey, Distribution of the Estimators for Autoregressive Time Series With a Unit Root, 1979

11 : A.M. Khedr, I. Arif, P. Raj P.V., M.El-Bannany , Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey, 2021

12 : Poongodi et al., Prediction of the price of ethereumblockchain cryptocurrency in an industrial finance system, 2020

13: IBM, What are recurrent neural networks?, , https://www.ibm.com/topics/recurrent-neural-networks

14 : Sepp Hochreiter, Jurgen Schmidhuber, Long Short-Term Memory, 1997

15 : Kyunghyun Cho et al, Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, 2014

16: Adamant Capital - Tuur Demeester et al., A Primer on Bitcoin Investor Sentiment and Changes in Saving Behavior, 2019, https://medium.com/@adamant_capital/a-primer-on-bitcoin-investor-sentiment-and-changes-in-saving-behavior-a5fb70109d32