

Anàlisi de correlacions i predicció del preu del Bitcoin

Oriol Vilaseca

IT Academy

Abstract

En aquest projecte, s'explora la predicció del preu del Bitcoin utilitzant tècniques d'anàlisi de sèries temporals i aprenentatge automàtic. L'objectiu és desenvolupar un model precís que pugui predir el moviment del preu del Bitcoin en funció de les dades històriques, incloent-hi variables externes rellevants. Mitjançant la utilització de diverses llibreries i mètodes d'aprenentatge automàtic, s'analitzen les dades i s'entrenen diversos models amb unes variables seleccionades segons la seva correlació.

1. Introducció

El Bitcoin és la criptomoneda més popular que existeix i ha guanyat molta atenció en el mercat financer en els últims anys. La seva volatilitat i la manca de correlació amb altres actius tradicionals en fan un objecte d'estudi interessant. Aquest projecte té com a objectiu principal desenvolupar un model que pugui predir el preu del Bitcoin amb precisió, ja que aquesta informació pot ser valuosa per als inversors.

Comprovarem la correlació dels moviments del preu del Bitcoin amb els següents elements:

- **NASDAQ:** Sent la borsa de valors de les companyies tecnològiques més grans del món, és bastant probable que tingui una relació el preu del seu índex amb el preu del Bitcoin, ja que aquest últim es tracta d'una innovació tecnològica tot i que no està sota el paraigües de cap empresa.
- **Or:** Segons diuen experts tecnològics, el Bitcoin podria convertir-se en el substitut de l'or en quant a reserva de valor si aconsegueix deixar de ser tan volàtil, fet que s'aconseguiria quan l'adopció fos massiva. S'ha de tenir en compte que un dels atractius de l'or és la seva escassetat (anualment té menys d'un 5% d'increment de quantitat total en tot el món). Bitcoin és encara més escàs, ja que està limitat a 21.000.000, xifra a la qual mai s'arribarà ja que hi ha carteres amb bitcoins de les quals s'han perdut les claus i no es podran recuperar mai els bitcoins del seu interior. Com que la pèrdua de claus seguirà passant, es pot arribar a dir que una vegada s'hagin minat tots els Bitcoins possibles, el seu supply anirà disminuint. Això el converteix en un valor més escàs que l'or, i per aquest motiu se'l bateja com l'"or digital".
- **Tipus d'interès:** Es sol dir que quan els tipus d'interès són alts, la gent inverteix menys en valors de risc i viceversa.
- **Fases de la lluna:** existeix la teoria que el Bitcoin pot tenir certa correlació amb les fases de la lluna.

2. State of art

A Internet es poden trobar projectes on han utilitzat diverses tècniques per predir el preu del Bitcoin. Algunes d'elles inclouen models de regressió lineal, tècniques d'anàlisi de sèries temporals com ARIMA i models basats en aprenentatge automàtic com Random Forest i XGBoost.

La majoria d'aquests models obtenen una precisió de menys del 52%, la qual cosa no els fa rendibles, ja que cal tenir en compte que quan es realitza una operació de compra o venda, hi ha una comissió pel servei executat. Per tal que el model sigui rendible, hauria d'aconseguir una rendibilitat d'aproximadament el 54% o superior.

En aquest projecte, abordarem la cerca de variables amb correlació amb el Bitcoin, cosa que no hem vist que es fés en cap dels models anteriors. Això ens permetrà intentar crear un model rendible i més precís en les prediccions del preu del Bitcoin.

3. Metodologia

Hem dividit el projecte en dues seccions: "Correlacions" i "Predicció".

3.1. Correlacions

Mitjançant la llibreria Yfinance, que obté dades de l'API de Yahoo Finance, hem descarregat les dades del Bitcoin, del Nasdaq i de l'or. Per obtenir les dades dels tipus d'interès i de les fases de la lluna hem descarregat un arxiu CSV amb el seu històric.

Hem creat un dataframe tenint en compte únicament els preus de tancament del dia. En el cas de les fases de la lluna, les hem dividit en dues: una amb un rang des del dia de lluna plena fins al dia abans de lluna nova i una altra a l'inrevés. L'hem indexat segons la data i hem comprovat que no tingués cap null.

Hem fet un plot per visualitzar l'evolució dels preus dels 3 valors (Bitcoin, Nasdaq i or).

Podem comprovar que el preu del Bitcoin és molt més volàtil que el Nasdaq o l'or, mostrant canvis de tendències molt bruscs.

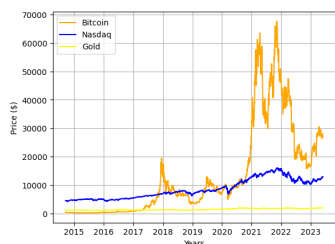


Figure 1: Bitcoin vs Nasdaq vs Gold

La tendència més estable és la de l'or, però també és el valor que menys retorn genera. Mirem-ho ara a escala logarítmica:

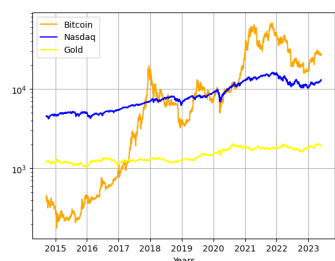


Figure 2: Bitcoin vs Nasdaq vs Gold (logarithmic)

Al posar-ho a escala logarítmica podem veure certes correlacions entre els tres valors a partir de l'any 2020, sobretot entre el Bitcoin i el Nasdaq. Fem un heatmap per comprovar-ho i veiem que el preu del Bitcoin i el del Nasdaq estan altament correlacionats. És una variable que tindrem en compte de cara a fer el model. No tindrem en compte la de l'or perquè aquesta està altament correlacionada amb la del Nasdaq i cauríem en la multicolinealitat.

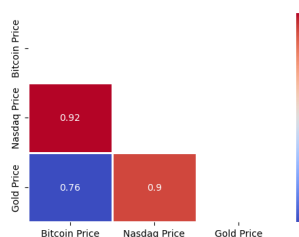


Figure 3: Heatmap Bitcoin, Nasdaq & Gold

En quant als tipus d'interès, quan són baixos els rendiments en alternatives tradicionals com ara els dipòsits bancaris o els bons del govern també són baixos. Això pot fer que els inversors busquin opcions més rendibles, que poden incloure inversions amb un perfil de risc més alt, com el Bitcoin o altres actius volàtils.

A inicis del 2020 hi ha una caiguda molt abrupta dels tres valors, provocada per l'inici de la pandèmia del Covid, i una posterior recuperació marcant una forta tendència alcista fins a finals del 2022, on comença una tendència baixista.

Entre finals del 2014 i inici del 2020 no sembla que hi hagi una correlació. No obstant, això canvia amb l'arribada de la pandèmia on es redueixen a 0,25 els tipus d'interès i el preu

del Bitcoin explota cap amunt. Al 2022, els tipus d'interès pugen de manera molt agressiva i el preu del Bitcoin s'enfonsa. Sembla que a partir de 2020 podria haver-hi una correlació.

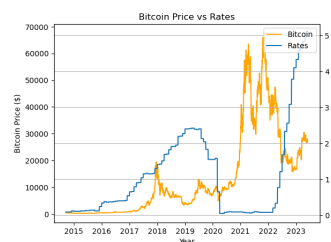


Figure 4: Bitcoin vs Interest Rates

Al comprovar-ho, obtenim els següents resultats:

Correlació entre Bitcoin i tipus d'interès en l'històric: -0.03.

Correlació entre Bitcoin i tipus d'interès a partir del 2020: -0.27.

Sembla ser que la correlació augmenta a partir del 2020. Com que els inversors s'anticipen a les pujades o baixades dels tipus d'interès, agruparem les dades en finestres de 1 any i tornarem a comprovar la correlació.

Correlació en l'any 2020: -0.29.

Correlació en l'any 2021: -0.48.

Correlació en l'any 2022: -0.89.

Correlació en l'any 2023: 0.77.

Obtenim informació contradictòria que no ens permetria fer prediccions. Entre 2020 i finals del 2022, veiem que la correlació (negativa) augmenta cada any, sent molt forta l'any 2022 (els tipus d'interès van pujar agressivament i el preu del Bitcoin va col·lapsar). No obstant, a l'any 2023, els tipus d'interès han seguit pujant i el preu del Bitcoin s'ha començat a recuperar, el que ens dona una correlació positiva del 0.77.

Podem concloure que aquesta variable no la tindrem en compte a l'hora de fer el model.



Figure 5: Preu del Bitcoin i fases de la lluna l'any 2021

Com podem veure en aquest gràfic del preu del Bitcoin durant l'any 2021, quan hi ha lluna nova el preu del bitcoin cau, quan hi ha lluna plena el preu del bitcoin puja. A partir d'un dataframe agrupant de manera individual els trams entre lluna plena i lluna mirarem si hi ha una correlació entre les fases i el preu. Els resultats estan al voltant del 50%, per tant podem afirmar que no hi ha cap tipus d'influència de la lluna en el preu.

Descartarem aquesta variable

De les 4 variables que hem analitzat, només el Nasdaq té correlació amb el preu del Bitcoin. Per tant, és l'única que tindrem en compte.

Total Full Phases	29
Full Up Count	15
Full Down Count	14
Total New Phases	30
New Up Count	14
New Down Count	16
Full Up %	52
Full Down %	48
New Up %	47
New Down %	53

Figure 6: Resultats de l'anàlisi de la lluna

3.2. Predicció

Per fer la predicció hem utilitzat els models Random Forest, SVM i LSTM.

Mitjançant una API ens descarreguem dades del "Fear and Greed Index". Aquest índex té com a objectiu proporcionar una indicació del sentiment dominant entre els participants del mercat, ja que les emocions com la por i l'avarícia poden tenir un gran impacte en els preus del Bitcoin.

L'índex es calcula a partir de diversos factors i indicadors relacionats amb Bitcoin, com ara la volatilitat del preu, el volum de les operacions, les tendències de Google, les notícies i altres indicadors de sentiment del mercat. Quan l'índex es troba en valors baixos, indica que predomina la por en el mercat, i els inversors poden estar més cauts i reticents a comprar Bitcoin.

Per contra, quan l'índex mostra valors alts, significa que l'avarícia és l'emoció dominant i que els inversors estan més propensos a comprar Bitcoin amb la creença que el preu seguirà augmentant. Això pot ser un indicador d'una sobrecompria o d'una bull market.

Hem creat un pipeline per a fer el preprocesat aplicant un MixMaxScaler al model SVM i al LSTM, ja que al Random Forest no li cal.

Hem evaluat cada model utilitzant TimeSeriesSplit pel back-testing.

Posteriorment hem aplicat PCA a dues variables per veure si milloraven els resultats.

4. Resultats

Els resultats han sigut decebedors, ja que no hem aconseguit obtenir una rendibilitat que arribi al 54%, que era l'objectiu. Tots els resultats han rondat el 50% amb els tres models, inclús després d'aplicar PCA els resultats s'han mantingut similars.

5. Conclusions

A l'hora de fer prediccions sobre inversions, hem de tenir en compte l'impacte de les emocions humanes en aquestes. Moltes vegades veiem accions d'empreses que es disparen degut al FOMO (Fear Of Missing Out), és a dir, degut a la por de les persones de no entrar a temps per beneficiar-se del benefici que obtindran si segueix pujant aquella determinada acció. Això genera que el valor de certes empreses pugi fins a nivells molt per sobre del que seria el seu valor real basant-nos en indicadors com els beneficis després d'impostos i els dividendes. Llavors, quan la gent pren beneficis i ven les seves accions, es produeix un efecte dominó que ensorra el preu i el sol situar al que seria el seu valor real. Tot aquest procés crea unes fluctuacions molt brusques que són difícils de predir. Això encara s'accentua més en un mercat completament emocional i especulatiu com és el del Bitcoin, ja que aquest encara no té una utilitat real i la gent en compra perquè preveu que d'aquí uns anys és possible que la tingui.

Apart de tot el mencionat fins ara, també li hem de sumar el component de Black Swan. El Black Swan en economia és un terme per referir-se a esdeveniments altament improbables però de gran impacte que sorprenen i tenen conseqüències significatives en els mercats financers i l'economia. Aquests esdeveniments són inesperats i es consideren imprevisibles segons els models i teories convencionals. Un cas de Black Swan podria ser la crisi del 2007 o la pandèmia del Covid. Són events que passen una o dues vegades cada dècada. En el cas del mercat de les criptomonedes, ens trobem Black Swans anuals. En els últims 2 anys hi ha hagut Black Swans com la prohibició de minar Bitcoins a la Xina, l'error sistèmic de la criptomoneda Luna, la caiguda del segon exchange més gran del mercat (FTX), diversos hackejos...

Tot això fa que el preu del Bitcoin sigui un dels valors més difícils de predir en el camp de l'inversió.

Considero que si volguéssim millorar el model hauriem de considerar recopilar més indicadors de sentiment i estipular un altre rang temporal de predicció (setmanal o mensual). També hauriem de tenir en compte que el mercat s'anticipa a les pujades o baixades de tipus d'interès basant-se en dades econòmiques com l'inflació o l'atur i que per tant, quan el moviment de tipus d'interès es fa efectiu, en el mercat ja s'ha tingut en compte dies abans.

Referències

<https://www.kaggle.com/datasets/lind18/full-moon-calendar-1900-2050>
<https://home.treasury.gov/interest-rates-data-csv-archive>
<https://finance.yahoo.com/>
<https://finance.yahoo.com/>
<https://alternative.me/crypto/fear-and-greed-index/>