



Förutse pris på kryptovalutor med maskininlärning

Rasmus Hellen

EXAMENSARBETE	
Arcada	
Utbildningsprogram:	Informationsteknik
Identifikationsnummer:	
Författare:	Rasmus Hellen
Arbetets namn:	Förutse pris på kryptovalutor med maskininlärning
Handledare (Arcada):	Fredrik Welander
Uppdragsgivare:	
<p>Sammandrag:</p> <p>Detta arbete behandlar forskningsfrågan om det med hjälp av maskininlärning är möjligt att förutse priset på en kryptovaluta. För att få svar på denna fråga har det utvecklats en maskinlärningsmodell. Arbetet redogör vilka metoder som finns tillgängliga när man gör en sådan utveckling, samt vilka metoder som används vid utvecklingen av detta arbete. I detta arbete går man också igenom vad en blockkedja är, blockkedjans historia, och vad man kan använda blockkedjor till. Även vad en kryptovaluta är, kryptovalutas historia samt vad man kan använda en kryptovaluta till. Samt vad maskininlärning är, maskininlärningens historia och vad man använder maskininlärning till. Målet med detta arbete var att uppnå ett resultat som påvisar att det inte är fullständigt omöjligt att förutse priset på en kryptovaluta med hjälp av maskinlärning. Resultaten indikerar att det är mycket svårt att utveckla en modell som kan förutse priset till punkt och pricka. Men att det är möjligt att förutse en uppåt- eller nedåtgående trend i priset. Dock rekommenderas mer utförliga tester och mer noggrann utveckling för att få ett mer konsekvent resultat.</p>	
Nyckelord:	Kryptovaluta, maskininlärning, artificiell intelligens,
Sidantal:	
Språk:	Svenska
Datum för godkännande:	

DEGREE THESIS	
Arcada	
Degree Programme:	Information Technology
Identification number:	
Author:	Rasmus Hellen
Title:	Predict prices on cryptocurrencies with machine learning
Supervisor (Arcada):	
Commissioned by:	
<p>Abstract:</p> <p>The thesis deals with the research question, with the help of machine learning, is it possible to predict the price of a cryptocurrency. To answer this question a machine learning model has been developed. In the thesis it is described which methods that are available when making such a development, as well as which methods that are used in the development. In this thesis you also go through what a blockchain is, the history of the blockchain and what you can use blockchains for. Also what a cryptocurrency is, the history of cryptocurrencies and what you can use cryptocurrencies for today. As well as what machine learning is, the history of machine learning, and what machine learning is used for. The goal for this thesis was to achieve a result that it is not impossible to predict the price of a cryptocurrency using machine learning. The results indicate that it is very difficult to develop a model that can predict the exact price of a cryptocurrency. But that it is possible to predict an upward or downward trend in the price. However, more detailed tests and more accurate development is recommended to get a more consistent result.</p>	
Keywords:	Cryptocurrency, machine learning, artificial intelligence,
Number of pages:	
Language:	Swedish
Date of acceptance:	

INNEHÅLL / CONTENTS

Figurer	5
TERMER	6
1 INLEDNING	8
1.1 Syfte, mål och forskningsfrågor	9
1.2 Avgränsningar	9
1.3 Blockkedjor	10
1.3.1 Vad är en Blockkedja?.....	10
1.3.2 Blockkedjans användningsområden.....	11
1.4 Kryptovalutor	11
1.4.1 Vad är en kryptovaluta?.....	11
1.4.2 Kryptovalutans användningsområden	12
1.5 AI och Maskininläring	13
1.5.1 Vad är AI och Maskininläring?.....	13
1.5.2 Maskininläringens historia	14
1.5.3 Maskininläringens användningsområden	14
2 Metod.....	15
2.1 Forskningsmetod	16
2.2 Utvecklingsprocessen	16
2.3 Tillgängliga teknologer	21
2.3.1 Teknologier för förutsägelse av pris	21
2.3.2 Prisdatabaser	22
2.3.3 Kodning	22
2.4 Valet av teknologier	22
2.4.1 Yahoo Finance	23
2.4.2 Jupyter Notebook	23
2.4.3 Python.....	23
2.4.4 Kodbibliotek	23
2.5 Krav för att utföra förutsägelse av pris	24
3 Resultatredovisning	24
3.1 Resultat	25
4 Slutsatser	29
Källor / References	30

FIGURER

Figur 3 Kodexempel på funktion som sparar data från Yahoo Finance.....	17
Figur 4 Kodexempel på funktion som hämtar csv-fil och konverterar till dataframe. ...	17
Figur 5 Exempel på de första fyra epokerna i utskriften för en körning med 150 epoker.	18
Figur 6 Del 1 av modellträningskoden. (Stavan, 2021).....	18
Figur 7 Del 2 av modellträningskoden. (Stavan, 2021).....	20
Figur 8 Kod hämtad från GitHub för plottning av grafer. (Stavan, 2021)	21
Figur 9 Modifierad kod för plottning av grafer till detta arbete.	21
Figur 10 Resultat av en förutsägelse på priset av Bitcoin under perioden 02.01.2022- 08.01.2022 (100 Epoker)	25
Figur 11 Resultat av en förutsägelse på priset av Bitcoin under perioden 02.01.2022- 06.01.2022 (150 Epoker)	26
Figur 12 Resultat av en förutsägelse på priset av Bitcoin under perioden 01.09.2022- 07.09.2022 (300 Epoker).	27
Figur 13 Förutsägelser på de första 7 dagarna varje månad från 01.04.22 (300 epoker).	28

TERMER

AI – Artificiell Intelligens är en vetenskap som används för att skapa intelligens med hjälp av hårdvara och mjukvara.

Trading – Köp och sälj med värdepapper och kryptovalutor.

Blockchain – En typ av databas som är uppbyggd i olika block av data, som i sin tur är ihopkedjade. Blockchain används till kryptovalutor för att hålla dem decentraliserade, så att ingen enskild person eller myndighet skall göra dem centraliserade.

Dapp – En typ av applikation uppbyggd med hjälp av blockkedjeteknologi

Kryptovaluta – Decentraliserad digital valuta som är uppbyggd på blockkedjeteknologi.

FÖRORD

Detta arbete har skapats av eget intresse för kryptovalutor, maskininlärning och blockkedje-teknologin. Under min studietid på Arcada har jag inte stött på kryptovalutor eller blockkedje-teknologin så mycket. Mitt intresse av kryptovalutor födde en idé att göra mer djupgående forskning inom just de områdena. Före detta arbete hade jag endast arbetat med maskininlärning i skolan, vilket födde ett intresse hos mig om hur det kan användas med kryptovalutor. I detta arbete kommer man att ta reda på om det går att förutse priset på olika kryptovalutor med hjälp av maskininlärning. Kryptovalutor har blivit en stor del av dagens samhälle, och det är något som var och varannan person vet vad är för något. I detta arbete kommer det att skapas en AI som med hjälp av maskininlärning skall kunna förutse vilket håll priset kommer att gå på olika kryptovalutor. Genom att analysera och lära sig mönstren från tidigare handlingsdata. Bakgrunden till att jag valt att skriva om just AI och kryptovalutor är för att jag haft ett stort intresse av kryptovalutor sedan 2011. Jag har själv hållit på mycket med trading och diverse andra saker om just kryptovalutor. Så jag anser mig kunna en del om det redan, vilket har hjälpt mig mycket i färdigställandet av detta arbete.

1 INLEDNING

Blockkedjor skapades för att användas av kryptovalutor. Den första blockkedjan föreslogs i en teknisk rapport, ”Bitcoin: A Peer-to-Peer Electronic Cash System” av Satoshi Nakamoto, år 2008. Blockkedjor ger en bra lösning till förvaltning av digitala tillgångar tack vare deras oföränderliga natur och motståndskraft mot attacker. I början fanns endast Bitcoin, byggd på öppen källkod. Vilket gjorde att många fler kryptovalutor har skapats. Användningen av öppen källkod är viktig för att användarna skall kunna lista på systemet, och det tillåter i sin tur utvecklarna att bilda egna gemenskaper som jobbar på olika projekt. (Summers, 2022, Chapter 5 Cryptocurrencies section)

Ethereum skapades 2015 av Vitalik Buterin, finansierad av en publik försäljning. Då var redan 11,9 miljoner Ethereum myntade. Ethereum är en decentraliserad, blockkedjebaserad mjukvaruplattform som är byggd på öppen källkod. Precis som vilken annan blockkedja som helst, så behövs tusentals personer som kör mjukvara på sina enheter för att driva nätverket. (McDonald, 2021, 2. New cryptocurrencies and new developments section)

Ripple är en kryptovaluta som skapades av ett privatägt företag år 2012. Ripple baserar sig på RippleNet-blockkedjan som kallas för ”värdefullt internet” (eng. ”Internet of Value). Företagets mål är att kunna skapa ett globalt nätverk av institutioner och banker som i sin tur skall kunna reducera kostnaderna av internationella betalningar. På Ripples hemsida påstår dem sig vara det enda blockkedjeföretaget som har produkter i kommersiell användning. Många banker använder sig av Ripple, bland annat American Express, Santander och Siam Commercial Bank. När en användare av Ripple skickar mynt till en annan användare tar transaktionen endast 4 sekunder. Ripple kan hantera cirka 1500 transaktioner per sekund. (McDonald, 2021, 2. New cryptocurrencies and new developments section)

Tidigt 2018 kunde Bitcoin endast processa ungefär 3 transaktioner per sekund, i dagens läge kan Ethereum hantera ungefär 25 transaktioner per sekund. Det är nästan ingenting om man jämför med större centraliserade betalningssystem som till exempel Visa (25000

transaktioner per sekund) och Mastercard (44000 transaktioner per sekund). (Williams, 2019, IV. The Thrill of the New section)

1.1 Syfte, mål och forskningsfrågor

Syftet med detta arbete är att utreda om det med hjälp av maskininlärning, är möjligt att förutse åt vilket håll på en kryptovaluta kommer att gå. När man använder sig av en AI för att fatta beslut om att köpa, sälja eller hålla sina tillgångar, så finns det ingen risk för mänskliga misstag. Så vidare inte det sker något misstag vid utvecklingen eller träningen av den artificiella agenten. Varför jag väljer att göra just en AI med maskininlärning är för att en AI kan ta emot och behandla mycket mera data än vad än människa klarar av. Vilket innebär att min AI kommer att klara av att förutse kommande prisutveckling bättre, snabbare och mera korrekt än vad man skulle klara av på egen hand.

”Kan man med hjälp av AI och maskininlärning förutse åt vilket håll priset på kryptovalutor kommer att gå?” och ”Är det möjligt att analysera de tidigare data som finns för att kunna förutspå om priset på en kryptovaluta kommer att öka eller minska med hjälp av AI?” Är två frågor som detta arbete kommer att ge svar på.

Utmaningen har varit att hitta det bästa sättet att träna den artificiella intelligensen för att uppnå ett såpass bra resultat att man framöver kan använda den för att se mönster inom kryptomarknaden. Alla som håller på med köp och sälj av kryptovalutor vet att det är otroligt svårt att förutse åt vilket håll priset är på väg. Målet är att ta reda på om man med hjälp av maskininlärning kan ta reda på åt vilket håll priset på kryptovalutan Bitcoin kommer att gå.

1.2 Avgränsningar

Detta arbete är avgränsat till att endast analysera den största kryptovalutan sett till marknadsvärde, Bitcoin. Det kommer att göras för att det är den kryptovalutan som har störst värde sett till vad som just Bitcoin tillför kryptomarknaden. Det är även Bitcoin som var den första kryptovalutan så det finns mycket data som kan analyseras.

En annan avgränsning i detta arbete är att data endast analyseras som rå data.

1.3 Blockkedjor

I detta kapitel ingår information om vad en blockkedja består av, blockkedjans historia och även blockkedjans användningsområden.

1.3.1 Vad är en Blockkedja?

Blockkedjor (eng. Blockchain) är mjukvara. Det innebär att blockkedjor inte existerar som fysiska objekt, utan endast som olika strängar med kod på enheter som datorer, telefoner och andra enheter. Du kan använda dig av mjukvara eller internet för att ansluta dig till en blockkedja på din telefon, dator eller någon annan enhet som har möjlighet. Genom att ansluta dig till en blockkedja så blir du en del av ett system. Systemet kan bestå av flera hundra, tusen eller till och med miljoner olika människor och enheter. I ett blockkedjesystem finns det ingen som helst hierarkikontroll, alla människor och enheter befinner sig på samma plan. (Williams, 2019, I. Experiencing Blockchain section)

Tänk dig att du ligger på rygg och tittar upp på en klar natthimmel, och varje stjärna är kopplad med en ljusstråle till den stjärnan som är närmast. Varpå den stjärnan är kopplad till en annan stjärna. Sen en till, och en till. Så den ljusstrålen är kopplad till alla stjärnor i närheten, och via de i närheten till stjärnor längre bort. Ända in i oändligheten. Din fantasi gör natthimlen till en modell av ett distributionsnätverk. Där lever blockkedjan, stjärnorna är datorer, telefoner och enheter. Alla är kopplade till varandra, och alla har samma syfte. Att föra information till olika delar av blockkedjan med hjälp av sig själv. – Stephen P. Williams

Den första och mest kända blockkedjan skapades 2008 för att understöda bitcoin. Den digitala valutan som skapades av en väldigt mystisk man som aldrig har visat sitt ansikte i verkligheten. Först så visste inte någon att blockkedjor hade ett så pass brett användningsområde som det har. Det bestod bara av kod som bitcoin använde sig av. Men efter en stund så började folk förstå att plattformen möjligtvis kunde användas till häpnadsväckande innovationer. Det finns många olika blockkedjor nuförtiden, som används i många olika områden. Bland annat av andra kryptovalutor än bitcoin, som kallas för altcoins. Men även av modeföretag, Walmart, konstnärer, regeringen i Dubai, människor

som samlar på bilder av katter och folk som jobbar med köp och sälj av derivat. (Williams, 2019, II. How it Works section)

1.3.2 Blockkedjans användningsområden

Dapps, eller distribuerade applikationer öppnar blockkedjan till alla typer av användningsområden. De är i grund och botten mjukvara som är byggd på toppen av blockkedjan i ett mer lättillgängligt lager. Dapps kan utvecklas till att göra väldigt många olika uppgifter. Bland annat kan man utveckla Dapps till att räkna hur mycket som skall betalas ut till någon som genererar el till elnätet. Det finns i dagens läge Dapps som är utvecklade till att följa alpakaull från farmen den kommer från, till det att den tillverkas till en tröja. Det är för att se att ullen inte tar något extrasteg som gör att ullen inte är organisk. Det används även Dapps i andra sammanhang, bland annat som när befolkningen i Estland röstar i sina val, och Dapps som hjälper journalister ta betalt för sina verk av personer som läser dem. Det i sin tur tar bort alla mellanhänder som en journalist har, så som ägare och redaktörer. (Williams, 2019, IV. The Thrill of the New section)

Blockkedjeteknologi är något som verkligen behöver användas inom hälsovården för att skapa en mer teknisk förbättring. Det finns många nystartade företag som jobbar med att utveckla blockkedjeteknologi som skall användas inom hälsovården. Där det i första hand skall användas för patientjournaler och samtycke för att dela med sig informationen som finns i patientjournalen till andra. (Williams, 2019, IV. The Thrill of the New section)

1.4 Kryptovalutor

I denna underrubrik så ingår information om vad en kryptovaluta är för något, historien bakom hur kryptovalutor blev till och vad man kan använda kryptovalutor till.

1.4.1 Vad är en kryptovaluta?

En kryptovaluta är i grund och botten en digital kedja av signaturer. Vilket användarna för över till varandra genom att digitalt signera en hash. Vilket är en typ av kod som håller kryptovalutans blockkedja säker. Hashen är en kort, fixerad sträng av data som i sin tur motsvarar en längre sträng av data med varierande längd. Den skapas av en datoralgoritm

som heter SHA-256. SHA-256 är en variation av den nationella säkerhetsmyndigheten i USA:s algoritim som heter SHA-2, som skapades år 2001. (McDonald, 2021, 1. Introduction: Bitcoin beginnings section)

Hashen av ett blocks huvud måste vara 256 bitar (binära nummer, antingen 1:or eller 0:or). På så sätt är det lätt att se om data som skickas har blivit ändrad på något sätt. Det är en snabb process att verifiera att den data som skickas är korrekt. Om det är två strängar som jämförs mellan varandra och dom inte är likadana så har den data som skickas blivit ändrad någonstans. Hash används på flera olika ställen, bland annat för att se om den data som skickas har blivit ändrad. Till exempel när data skickas från en hårddisk eller i ett nätverk så sparas en oftast hash av den data som skickas eller sparats. Så när data sen laddas från hårddisken eller nätverket kollar man med hjälp av algoritmen om den data som laddas är samma data som en gång sparats. Det är ett mycket snabbt och pålitligt sätt att kontrollera data. (McDonald, 2021, 1. Introduction: Bitcoin beginnings section)

Den första kryptovalutan, Bitcoin skapades när Satoshi Nakamoto publicerade en teknisk rapport som hette ”The Peer-to-Peer Electronic Cash System” i oktober 2008. Under 2009 skapade han den första bitcoinen med hjälp av Bitcoins blockkedja. Som i nutid kallas för ”The Genesis Block” (Det ursprungliga blocket). Bitcoins mål var att byta ut bankerna med en ny valuta som inte hade någon finansiell mellanhand. Bitcoin skapades under en tid när det traditionsenliga banksystemet hade blivit rejält testat efter den ekonomiska kris som skedde år 2009. (McDonald, 2021, 1. Introduction: Bitcoin beginnings section)

1.4.2 Kryptovalutans användningsområden

För att använda sig av kryptovalutor så måste du ha en så kallad plånbok för kryptovalutor. Din kryptoplånbok förvarar din privata nyckel, för att kunna spendera dina kryptovalutor så behöver du ha en privat nyckel. Dock fungerar inte en kryptoplånbok på samma sätt som en vanlig plånbok. En kryptoplånbok innehåller din privata nyckel, som är sammankopplad med din publika nyckel. Den publika nyckeln är den som används vid transaktioner, och den privata nyckeln används för att validera transaktionen. Transaktionen kan inte bli validerad till och med det att ägaren av kryptovalutan har verifierat att kryptovalutan tillhör användarens privata nyckel. Alla transaktioner sänds ut till

blockkedjenätverket för att valideras av andra blockkedjeanvändare. Transaktionstiden kan ta uppemot 20 minuter. (McDonald, 2021, 1. Introduction: Bitcoin beginnings section)

Den mörka sidan av kryptovalutor är att den används mycket inom olagliga transaktioner på 'Dark Web'. Anledningen till det är för att transaktioner som görs med kryptovalutor är anonyma. Mer än en fjärdedel av alla transaktioner som görs med kryptovalutor är länkat till kriminella aktiviteter. (Matharu, 2018, Is This Not Used for Anonymous Transactions for Illegal Purposes? section)

1.5 AI och Maskininläring

I denna underrubrik ingår information om vad maskininläring är för något, historien bakom maskininläring och vad man använder maskininläring till i riktiga världen.

1.5.1 Vad är AI och Maskininläring?

Artificiell Intelligens är en vetenskap som används för att skapa intelligens med hjälp av hårdvara och mjukvara. Maskininläring är ett studieområde som handlar om att ge datorer möjligheten att lära sig utan att vara programmerade på ett visst sätt. (Nagy, 2018, Introduction section)

Maskininläring är ett sätt att åstadkomma artificiell intelligens, dock är det möjligt att uppnå artificiell intelligens utan att använda sig av maskininläring. Till exempel är det möjligt att koda en artificiell agent med hjälp av hårdkodade regler och beslutsträd, även fast det tillvägagångssättet har i princip ingenting att göra med maskininläring. (Nagy, 2018, Introduction section)

Både maskininläring och artificiell intelligens har starka ingenjörskomponenter. Det innebär att man kan använda både maskininläring och artificiell intelligens för att kvantifiera något baserat på teori. I stället för att helt enkelt använda sig av hypoteser. Maskininläring och artificiell intelligens har starka band till vetenskapliga komponenter, där det

testas koncept och skapas nya idéer om hur man kan uttrycka tankeprocessen. (Mueller och Massaron, 2021, s. 22)

1.5.2 Maskininlärningens historia

Artificiell Intelligens uppstod genom att man blev inspirerad till att dekonstruera hur människans hjärna fungerar, till exempel hur neuroner fungerar. Vår hjärna består av små enheter som kallas neuroner. Nätverken som neuroner ingår i kallas neuronnätverk. Bakom neuronnätverken så finns det många modeller som kan användas för att skapa modeller som kan lösa problem i riktiga världen med hjälp av artificiell intelligens. (Nagy, 2018, Introduction section)

Maskininlärningen kommer från början av Alan Turings rapport från år 1950, där Turing forskade i idén om att maskiner kan tänka. Turings rapport ledde till att ett spel skapades mellan tre spelare, spelare A, spelare B och spelare C. Där spelare A är en dator och spelare B är en människa. Spelare A och spelare B skall försöka övertyga spelare C, som också är en människa som inte kan se spelare A eller spelare B att de är människor. Om spelare C inte kan välja vem som är en människa så vinner spelare A, datorn. (Mueller och Massaron, 2021, s. 25)

Ett problem på 1950- och 1960-talet med artificiell intelligens, är att man är för optimistisk med vad den artificiella intelligensen kan åstadkomma. Vetenskapsmän försöker lösa problem med artificiell intelligens som är alldeles för komplexa. Dock gjorde den tidiga optimismen så att vetenskapsmännen tänkte att man inom 20 år skulle kunna skapa intelligenta maskiner. I och med att maskinerna redan gjorde krävande uppgifter så som att spela komplexa spel. Den artificiella intelligensen har i dagens läge haft sin största framgång i områden som logistik, datautvinning, och medicinska diagnoser. (Mueller och Massaron, 2021, s. 25)

1.5.3 Maskininlärningens användningsområden

När man löser ett problem, eller interagerar med människor så går man igenom en process. Om man limiterar omfattningen av ett problem eller en interaktion så kan man ofta

modellera och automatisera processen. (Nagy, 2018, How does AI Solve Real World Problems? section)

Ibland känns det som att den artificiella intelligensen vet precis vad vi vill ha, som till exempel personliga kuponger som man får när man köper något på internet. Det är ett effektivt sätt att få oss människor att tro att den artificiella intelligensen vet exakt vad vi behöver och vill ha. Vilket maximerar försäljningen. (Nagy, 2018, How does AI Solve Real World Problems? section)

En självkörande bil är livsfarlig om den inte skulle kunna känna av att det finns bilar i körväg som kör på samma väg som bilen är. Den artificiella agenten behöver bearbeta och känna vad som finns runt omkring den för att ha möjligheten att köra bilen på ett korrekt sätt. Men det i sig räcker inte för att skapa en artificiell agent som kan hantera en bil, agenten behöver information om fysiken av objekt som rör på sig. Att köra bilen i en normal miljö skulle vara helt livsfarlig utan den informationen. (Nagy, 2018, Diversity of Disciplines section)

För att skapa en artificiell agent som klarar av att köra en bil så behöver man förse den med mycket information. Så som fysik för att modellera den riktiga världen, statistik för att hämta värdefull information, förutse framtiden och analysera det förflutna, men även psykologi för att modellera hur den mänskliga hjärnan fungerar. (Nagy, 2018, Diversity of Disciplines section)

Man brukar säga att en intelligent artificiell agent klarar av att observera omgivningen med hjälp av sensorer, riktar sina aktiviteter mot att uppnå de mål som behövs, och att den gör det utan någon som helst inmatning från en människa. (Nagy, 2018, AI Tools and Learning Models section)

2 METOD

I metodkapitlet så beskrivs vilka metoder som är tillgängliga och möjliga att använda sig av, samt vilken forskningsmetod som använts. Därefter så väljs den metod som ses som

mest passande för arbetet som skall utföras. Efter att metoden är vald så planeras kraven för att utföra förutsägelser av pris på kryptovalutor, samt vilka olika teknologier som behövs för att uppnå de krav som är ställda.

2.1 Forskningsmetod

Forskningen påbörjades med läsning av material och tidigare forskning. För att efter det få en bra grund att stå på vid utförandet av detta arbete. Då det inte fanns mycket forskning inom förutsägelser av kryptovalutor, men mycket forskning gällande förutsägelser på aktier. Så grundar sig koden på förutsägelser av aktier, modifierad för att kunna användas vid förutsägelser av kryptovalutor.

Därefter påbörjades utvecklingen av den maskininlärningsmodell som skall förutse priset på kryptovalutor. Efter att koden var klar påbörjades testningen av parametrarna för maskininlärningsmodellen, det gjordes för att testa sig fram till vilka parametrar som ger ett så konsekvent resultat som möjligt. Då dessa var noggrant analyserade och utvalda påbörjades analysen av resultatet.

2.2 Utvecklingsprocessen

Utvecklingsprocessen började med att läsa sig in på området, samt att läsa tidigare forskning och resultat inom samma område. Sedan användes till största del kod som fanns tillgänglig på internet. (Banas, 2021) Med diverse förändringar för att utföra förutsägelse av pris på kryptovalutor. Då det som fanns tillgängligt var kod som utför förutsägelser av aktier och diverse andra instrument. Derek Banas har mycket användbar kod som är tillgänglig för gemene man. Banas kod har mycket att göra med att använda sig av python för att ladda ned och processa data från Yahoo Finance. Banas kod har varit mycket användbar under arbetet.

Steg 2 i utvecklingen av arbetet var att testa sig fram till det bästa möjliga resultatet med så lite fluktuation som möjligt från det verkliga priset. Det var också det som tog mest tid av hela utvecklingen och arbetet.

Funktionerna som används i arbetet är.

- Save_to_csv_from_yahoo

Vilket är en funktion som sparar data från Yahoo Finance, och konverterar datan till en maskinläsbar csv-fil. Denna funktion är till en början skriven av Derek Banas, med vissa förändringar i koden för att passa in i detta arbete.

```
def save_to_csv_from_yahoo(ticker, syear, smonth, sday, eyear, emonth, eday):
    start = dt.datetime(syear, smonth, sday)
    end = dt.datetime(eyear, emonth, eday)

    df = web.DataReader(ticker, 'yahoo', start, end)

    df.to_csv("C:/Users/razze/Desktop/Examensarbete/csv/" + ticker + '.csv')
    return df
```

Figur 3 Kodexempel på funktion som sparar data från Yahoo Finance.

- Save_to_csv_from_yahoo_test_data

För att göra datahanteringen så simpel som möjligt används en likadan funktion som föregående för att spara testdata som modellen sedan testas på.

- Get_df_from_csv

Vilket är en funktion som hämtar csv-filen och sedan konverterar den till en dataframe (sv. dataram) med hjälp av ett bibliotek i python som heter Pandas. Även denna funktion är till en början skriven av Derek Banas, med vissa förändringar i koden för att passa in i detta arbete.

```
def get_df_from_csv(ticker):
    try:
        df = pd.read_csv("C:/Users/razze/Desktop/Examensarbete/csv/" + ticker + '.csv')
    except FileNotFoundError:
        print("File doesn't exist")
    else:
        return df
```

Figur 4 Kodexempel på funktion som hämtar csv-fil och konverterar till dataframe.

- Modellträning

Detta är koden som sköter träningen av modellen som utför prisförutsägelsen på kryptovalutan. Det börjar med att *sklearn.preprocessing* används för att skala datan. Sedan väljer man hur många dagar framåt som modellen skall förutse priset. Efter det delas den skalade datan upp i två olika delar, för att göra träningen så optimal som möjligt. Därefter används Sequential, vilket är en del av Tensorflow-biblioteket. Sequential grupperar en linjär stapel av data till olika lager.

Därefter väljs det hur mycket man vill träna sin modell. För detta arbete har de optimala värdena sett ut att vara mellan 100 och 150 epoker med en omgångsstorlek på 64. När den körningen i sin tur är klar kan man göra en förutsägelse av priset på den data som matats in tidigare i programmet.

```
Epoch 1/150
29/29 [=====] - 4s 28ms/step - loss: 0.0227
Epoch 2/150
29/29 [=====] - 1s 28ms/step - loss: 0.0033
Epoch 3/150
29/29 [=====] - 1s 27ms/step - loss: 0.0033
Epoch 4/150
29/29 [=====] - 1s 27ms/step - loss: 0.0028
```

Figur 5 Exempel på de första fyra epokerna i utskriften för en körning med 150 epoker.

```
scaler = MinMaxScaler(feature_range=(0,1))
scaled_data = scaler.fit_transform(BTC['Close'].values.reshape(-1,1))

prediction_days = 30

x_train = []
y_train = []

for x in range(prediction_days, len(scaled_data)):
    x_train.append(scaled_data[x-prediction_days:x, 0])
    y_train.append(scaled_data[x, 0])

x_train, y_train = np.array(x_train), np.array(y_train)
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))

model = Sequential()

model.add(LSTM(units=50, input_shape = (x_train.shape[1], 1), return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(units=50, return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(units=50))
model.add(Dropout(0.2))
model.add(Dense(units=1))

model.compile(optimizer='adam', loss='mean_squared_error')
model.fit(x_train, y_train, epochs=150, batch_size=64)
```

Figur 6 Del 1 av modellträningskoden. (Stavan, 2021)

Modellträningskoden börjar med en variabel för minimum- och maximumskalning, importerad från Scikit-learn-biblioteket. Min- och maxskalaren krymper data så att den blir

mer lätthanterlig för modellen att läsa och träna sig in på. Där den läser det högre värdet av två som en etta, och det lägre värdet av de två som en nolla. `Fit_transform` är även det en inbyggd funktion i Scikit-learn. Det är en funktion som är sammanslagen av två funktioner, `fit` och `transform`. `Fit` används för att beräkna medelvärdet samt standardavvikelsen för en funktion, för användning vid skalning av data. `Transform` är funktionen som skalar data, och använder sig då av de värden som räknades ut av `fit`-funktionen. Det vill säga, medelvärde och standardavvikelse. (GeeksForGeeks, 2022)

Sedan skapas två tomma variabler som skall användas som datasamlingar för när data som skall användas vid inträningen av modellen delas upp i två olika delar. Samt en variabel för hur många dagar modellen skall förutse priset. Därefter fylls variablerna med skalade data.

Efter datan har blivit skalad och ifylld i variablerna används en array-funktion från det importerade `numpy`-biblioteket. Det är en funktion som skapar en datasamling av den data som ges till den. Därefter används en till funktion från `numpy`-biblioteket, som heter `reshape`. Den formaterar om datasamlingen från en lång datasamling till fler kortare datasamlingar.

Därefter skapas en variabel som används för själva modellen. Som är en modell importerad från `tensorflow`-biblioteket. Modellen heter `Sequential`. Som parametrar tar den in `LSTM`, vilket står för long short-term memory (sv. långt korttidsminne). Vilket används för tidsserien när man förutser något. Samt vilken typ av datasamling som den data som används är skalad till.

Därefter läggs `dropout` till, `dropout` är en funktion som används för att slumpmässigt välja en del data till inträningen, och inte träna på de delar som inte används. Det används för att förhindra överexponering till den data som tränas in. (Klein, 2022)

Sedan kompileras modellen med två parametrar, `'adam'` och `'mean-squared-error'`. `Adam` är en optimeringsalgoritm, som används av modellen. I den modellen ingår vissa värden, så som lärningsgrad och `epsilon`. `Mean-squared-error` är en algoritm som beräknar skillnaden mellan värdet på den data som har förutsetts, och värdet på den data som modellen tränats mot. Därefter används `fit` från Scikit-learn, vilket beräknar medelvärdet och standardavvikelsen.

```

test_data = BTC_test
actual_prices = test_data['Close'].values

total_dataset = pd.concat((BTC['Close'], test_data['Close']), axis=0)

model_inputs = total_dataset[len(total_dataset) - len(test_data) - prediction_days:].values
model_inputs = model_inputs.reshape(-1,1)
model_inputs = scaler.transform(model_inputs)

# Make predictions on test data

x_test = []

for x in range(prediction_days, len(model_inputs)):
    x_test.append(model_inputs[x-prediction_days:x, 0])

x_test = np.array(x_test)
x_test = np.reshape(x_test, (x_test.shape[0], x_test.shape[1], 1))

predicted_prices = model.predict(x_test)
predicted_prices = scaler.inverse_transform(predicted_prices)

```

Figur 7 Del 2 av modellträningskoden. (Stavan, 2021)

Sedan skapas tre variabler, en för den data som modellen testas mot, en för det verkliga priset på den data som testas mot, samt en för den totala datasamlingen av värden. Modellen får sedan den totala datasamlingen, minus längden på den data som testas och minus hur många dagar som skall förutses. Därefter formas den data som modellen skall få med hjälp av reshape-funktionen. Sedan används transform från Scikit-learn på samma data.

Därefter skapas en tom datasamlingsvariabel för den data som skall förutses. Även där tar man in samma data. Sedan används två funktioner från numpy-biblioteket array och reshape. Därefter görs själva förutsägelsen på den data som modellen har fått.

- Prisplottning

I detta arbete användes matplotlib.pyplot-biblioteket, för att göra en så simpel och läsbar graf som möjligt vid plottning av förutsägelsen. För plottningen av graferna i resultatet användes kod från en utomstående källa. (Stavan, 2021) Dock modifierades koden för att passa detta arbete, med språk och liknande.

```

81 # Plot The Test Predictions
82 plt.plot(actual_prices,color="black",label=f"Actucal {stock} Price")
83 plt.plot(predicted_prices, color="green", label=f"Predicted {stock} Price")
84 plt.title(f"{stock} Share Price")
85 plt.xlabel('Time')
86 plt.ylabel(f'{stock} Share Price')
87 plt.legend()
88 plt.show()

```

Figur 8 Kod hämtad från GitHub för plotning av grafer. (Stavan, 2021)

```

plt.plot(actual_prices, color='black', label='ActualPrice')
plt.plot(predicted_prices, color='green', label='PredictedPrice')

form1 = {'family': 'serif', 'color': 'blue', 'size': 20}

form2 = {'family': 'serif', 'color': 'darkred', 'size': 15}
plt.legend(["Verklig Slutkurs", "Förutsägelse"], loc = "lower left")
plt.title("BTC Prisförutsägelse", fontdict=form2)
plt.xlabel('Dagar', fontdict=form1)
plt.ylabel("Pris", fontdict=form1)
plt.show()

```

Figur 9 Modifierad kod för plotning av grafer till detta arbete.

2.3 Tillgängliga teknologer

Det finns många tillgängliga metoder för att utföra förutsägelse av pris. I detta kapitel går man igenom vilka typer av metoder och modeller man kan använda sig av för att utföra förutsägelse av priser på kryptovalutor.

2.3.1 Teknologier för förutsägelse av pris

- Tensorflow.keras.Sequential

Tensorflow har många olika inbyggda funktioner i sitt bibliotek för att kunna utföra förutsägelser av pris. Bland annat Sequential. För att kunna använda sig av Sequential så måste din data bestå av en typ av inmatning och endast en typ av utmatning. Det innebär

att man inte kan använda sig av Sequential för att förutse priset på flera olika typer av instrument eller kryptovalutor på samma gång. (Tensorflow, 2022)

- Scikit-learn

Scikit-learn är ett mycket avancerat bibliotek, som man kan använda sig av för att skapa många olika typer av modeller. Bland annat lineära modeller, och många typer av regressionsmodeller. Scikit-learn är ett mycket bra bibliotek när man vill ha ut så mycket specifika data som möjligt. (Scikit-learn, 2022)

2.3.2 Prisdata

Det finns många ställen på internet där man kan få tag på prisdata för diverse kryptovalutor. De alternativen som var mest gångbara för detta arbete är:

- Yahoo Finance

Fördelarna med att använda sig av Yahoo Finance i detta arbete, är att man där finner den mest kompletta data. Även att det lätt går att ladda ner all data man vill ha tillgång till. Utan att behöva lämna ifrån sig några som helst personliga uppgifter. Samt användbarheten tillsammans med Jupyter Notebook, där du kan ladda ned data och spara den som en CSV-fil direkt i kodningsmiljön.

- Cryptodatadownload.com

Fördelen är att man får välja precis den data man vill ladda ned. Nackdelen är att man måste lämna ifrån sig personliga uppgifter och skapa ett konto på hemsidan för att kunna ladda ner ned data som behövs.

2.3.3 Kodning

Det finns många kodspråk att välja mellan. Fördelarna med att använda sig av Python är att det finns mycket tillgänglig information i böcker och dylikt. Jupyter Notebook är ett väldigt bra verktyg när man programmerar i Python tack vare Jupyter Notebooks möjlighet att visualisera data i en notebook-fil som olika datauppsättningar av en bild.

2.4 Valet av teknologier

I detta kapitel framgår det vilken metod och vilka teknologier som används i detta arbete.

2.4.1 Yahoo Finance

I detta arbete används Yahoo Finance till prisdata för Bitcoin. Valet är tack vare enkelheten att kunna ladda ner en .csv-fil och sedan processa den enkelt med hjälp av Pandas och Numpy.

2.4.2 Jupyter Notebook

Valet av editor är Jupyter Notebook, i och med att man i Jupyter Notebook har möjlighet att installera och importera de bibliotek man behöver. Samt att ha möjligheten att visualisera data på ett lättförståeligt sätt med hjälp av de bibliotek som finns tillgängliga.

2.4.3 Python

Programmeringsspråket som används i detta arbete är Python. Valet gjordes tack vare Pythons möjligheter att importera de kodbibliotek som är användningsbara när man skall utveckla en agent som använder sig av maskininlärning för att förutse pris på kryptovalutor eller andra instrument. Samt att Python har många bibliotek som är mycket användbara när det gäller visualisering av resultat.

2.4.4 Kodbibliotek

- `Tensorflow.keras.Sequential`

I detta arbete används `Tensorflow.keras.Sequential`. Tack vare den mängd information som finns att hitta på internet. Även på grund av att den data som används endast har en typ av inmatning, och en typ av utmatning. Det vill säga, den prisdata som man tillför modellen för att den skall lära sig.

- `Sklearn.preprocessing`

Sklearn har många olika bibliotek som är användbara för att göra förutsägelser av pris. I detta arbete används ett bibliotek som kallas `Sklearn.preprocessing`. Det används till att normalisera den data som skall användas för inträningen av modellen som skall förutsäga priset på kryptovalutan.

- Pandas

Pandas används för den största delen av all datahantering i arbetet. Bland annat för att konvertera den data som hämtas från Yahoo Finance till en maskinläsbar csv-fil. Pandas används dessutom till att konkatenera de data som delas för att utföra test och träning på modellen.

- Matplotlib.pyplot

Matplotlib är ett bibliotek som innehåller många olika funktioner som har med plottning av grafer och liknande att göra. I detta arbete används ett bibliotek i matplotlib som heter pyplot. Det är för att kunna visa ett så lättläst resultat som möjligt.

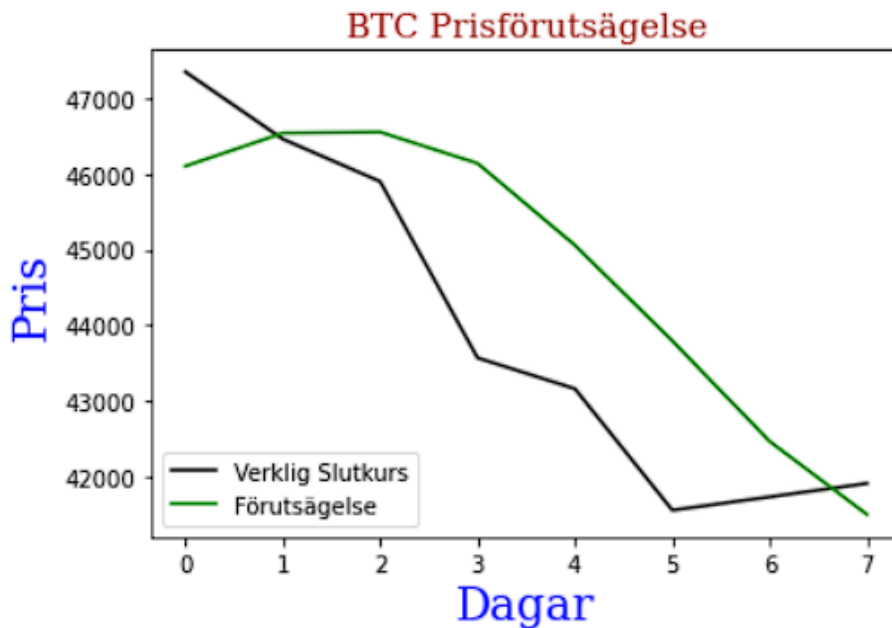
2.5 Krav för att utföra förutsägelse av pris

Kraven för att kunna utföra förutsägelser av pris på kryptovalutor är att man har en bra grund, det vill säga, bra data att utgå från. Samt att man gör mycket tester med olika metoder för att utesluta de metoder som ger ett mindre konsekvent resultat.

3 RESULTATREDOVISNING

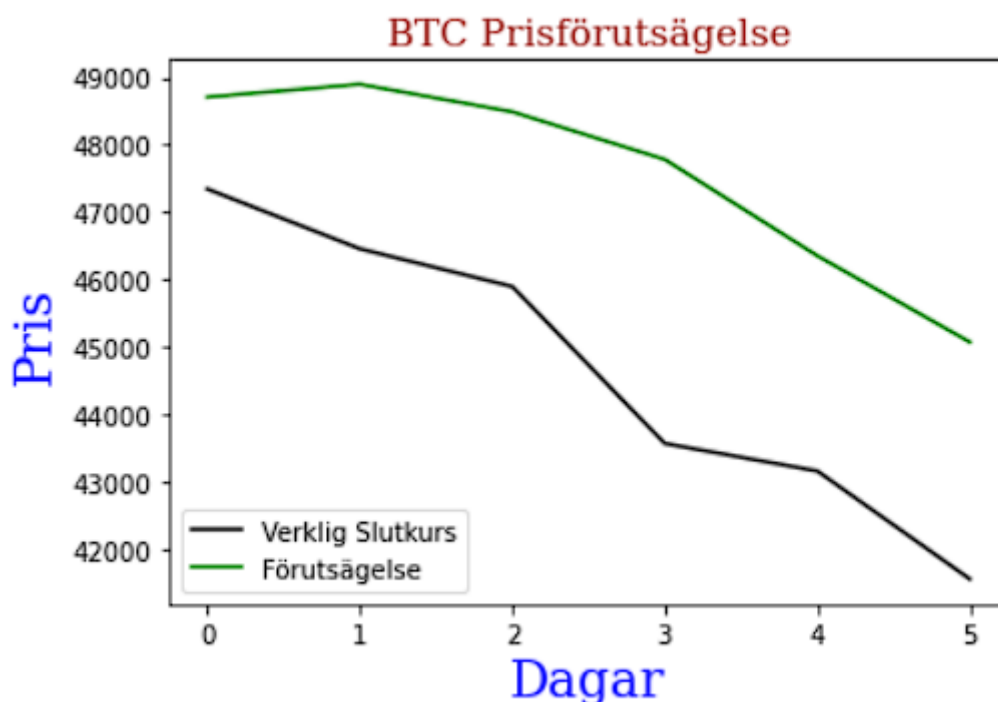
I kapitlet presenteras resultatet av utvecklingen, där resultatet och utvecklingsprocessen analyseras.

3.1 Resultat



Figur 10 Resultat av en förutsägelse på priset av Bitcoin under perioden 02.01.2022-08.01.2022 (100 Epoker)

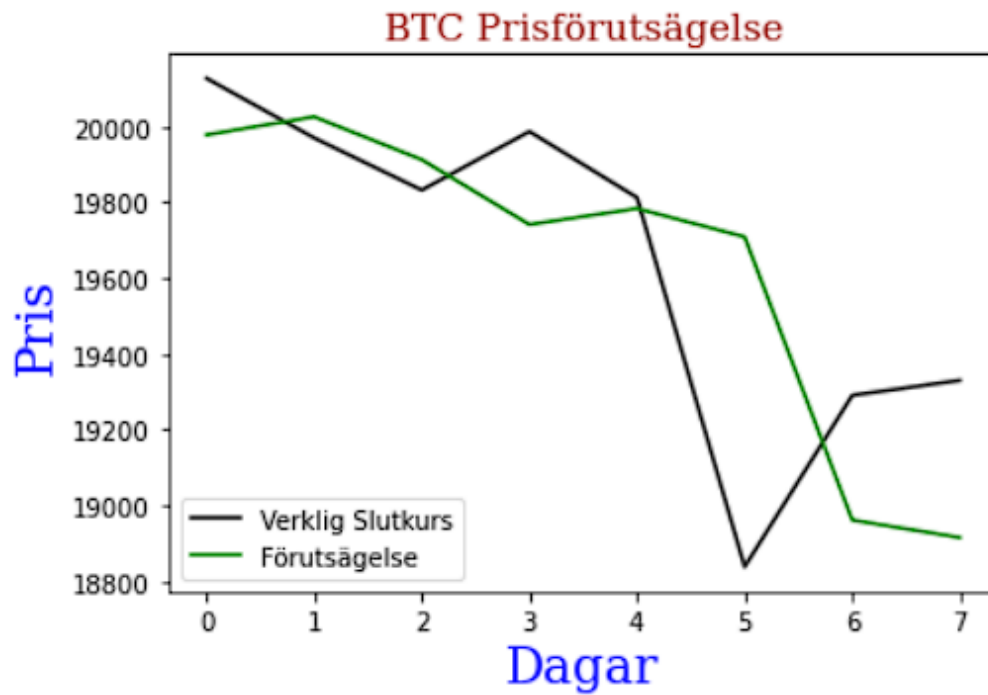
Detta är resultatet av en förutsägelse på Bitcoin med hjälp av modellen som utvecklades i detta arbete. Det här är en körning på 100 epoker och en förutsägelse på 7 dagar. Som man kan se så skiljer sig det förutsagda startpriset med cirka 2,8% mot det verkliga priset. Det är på grund av att modellen inte vet det exakta startpriset, då den inte sett det priset tidigare. Man kan också se att modellen i sin tur inte var helt korrekt med sin förutsägelse, dock kan man se att modellen kunde se en nedåtgående trend under de kommande dagarna då förutsägelsen gjordes.



Figur 11 Resultat av en förutsägelse på priset av Bitcoin under perioden 02.01.2022-06.01.2022 (150 Epoker)

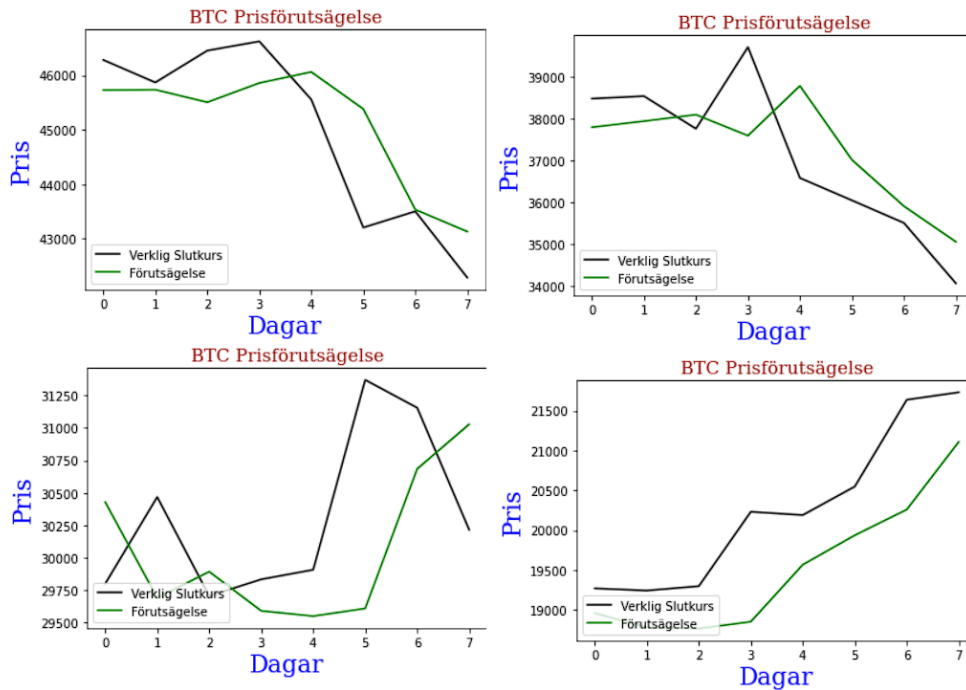
Detta är ett resultat av en kortare förutsägelse på fem dagar, med mera inträning. Där kan man se samma sak som i föregående resultat, att modellen börjar med ett högre öppningspris på dag 1. Men sedan rör sig modellens förutsagda pris parallellt med det verkliga priset neråt, vilket är ett mycket bra resultat för något som inte har ett referenspris att gå efter.

I sin helhet så påvisar resultatet att det är mycket svårt att göra förutsägelser av pris på något som styrs av gemene man. Då det finns så många faktorer som spelar in. Bland annat styrs priset av kryptovalutor mycket starkt av vad som sägs på sociala medier av inflytelserika personer. Det finns många exempel på sådana händelser inom kryptovalutsvärlden. Till exempel kan ett inlägg på sociala medier av en inflytelserik person få en kryptovaluta att antingen stiga i pris eller sjunka i pris väldigt drastiskt och mycket. Sådana saker är otroligt svåra att ta i beaktande när man utför förutsägelser av pris med hjälp av maskininlärning.



Figur 12 Resultat av en förutsägelse på priset av Bitcoin under perioden 01.09.2022-07.09.2022 (300 Epoker).

Som man kan se på detta resultat så förutsåg maskininlärningsmodellen en kraftig nedgång i priset, samt en relativt kraftig uppgång. Det verkliga priset består av en nästan lika kraftig nedgång som det förutsedda priset. Dock har det verkliga priset ingen kraftig uppgång. Vilket innebär i detta fall att man skall ta resultatet med en nypa salt. Det är mycket svårt att förutse ett korrekt pris på en kryptovaluta som handlas av gemene man. Som inte har någon som helst uppbackning av något annat än de personer som använder kryptovalutan i fråga.



Figur 13 Förutsägelser på de första 7 dagarna varje månad från 01.04.22 (300 epoker).

Som man kan se i det slutliga resultatet, så kan man argumentera om att resultatet från de fyra oberoende körningarna är för bra för att vara sanna. Där kan man se att förutsägelserna av priset visar på både uppåt- och nedåtgående trender som även det, har hänt med det verkliga priset. Man skulle även kunna argumentera om att modellen som är skapad i detta arbete har lyckats utföra ett väl utfört arbete med sina förutsägelser. Den sista förutsägelsen av de fyra är gjord då priset inte har fluktuerat så mycket. Troligtvis på grund av att ingen tredje part har påverkat priset. Där kan man se en stadigt uppåtgående trend i priset, både på det förutsagda priset och det verkliga priset. Förutsägelsen i fråga är den förutsägelsen som är mest pålitlig i arbetet.

4 SLUTSATSER

Som man kan se i resultatet av de sista förutsägelseerna, skulle jag ta resultatet av detta arbete med en nypa salt. Det är mycket svårt att förutse priset på ett mycket volatilt instrument. Det är väldigt många saker som man måste ta i beaktande när man skapar en modell som kan förutse priset på just en volatil kryptovaluta. Bland annat sådant som man inte har någon kontroll över. Till exempel, regler som ställs av myndigheter gällande kryptovalutor, eller inlägg på sociala medier.

Jag anser att jag har uppnått de mål som jag ställde mig vid början av denna forskning. I den teoretiska delen gick jag igenom vilka metoder som finns för utvecklingen, vilka metoder som användes, och slutligen vilka resultat jag fick och vad jag försökte uppnå med min utveckling. Forskningen har varit en mycket lärorik upplevelse för mig, och jag tror att jag i framtiden kommer att fortsätta jobba med detta och utveckla detta projekt till att bli bättre.

Personligen är jag mycket nöjd med att ha fått ett resultat av detta arbete som kunde visa på samma trender som det verkliga priset. När jag först började med detta arbete fick jag många kommentarer från klasskamrater, kollegor och vänner. Att det är ett projekt som kommer att vara helt omöjligt att utföra. Det fick mig att verkligen vilja få ut ett resultat som kunde påvisa att det inte alls är omöjligt. Det kanske är omöjligt att förutse priset på kryptovalutor till punkt och pricka. Men att förutse en uppåtgående eller nedåtgående trend är enligt mitt resultat mycket möjligt, om man skulle ge modellen mycket fler parametrar att jobba med.

Även om jag kände mig mycket bekväm med att jobba med kryptovalutor, i och med att det har varit ett av mina största intressen under en lång tid. Så var detta ett arbete som verkligen satte mina kunskaper på prov. Men att ha haft ett stort intresse av kryptovalutor en lång tid hjälpte verkligen mig i utförandet av detta arbete.

KÄLLOR / REFERENCES

Banas, D. (2021) Python for Finance. [Online]. Available at: <https://github.com/derekbanas/Python4Finance> (Accessed 10 September 2022).

GeeksForGeeks. (2022) Data Pre-Processing with Sklearn using Standard and Minmax scaler. Available at: <https://www.geeksforgeeks.org/data-pre-processing-wit-sklearn-using-standard-and-minmax-scaler/> (Accessed: 08 October 2022).

Holcombe, M. (2008) Running an Agile Software Development Project. 1st edn. Wiley. Available at: <https://www.perlego.com/book/2767285/running-an-agile-software-development-project-pdf> (Accessed: 21 July 2022).

Klein, B. (2022) Dropout Neural Networks in Python Available at: <https://python-course.eu/machine-learning/dropout-neural-networks-in-python.php> (Accessed: 08 October 2022).

Lazy, L. (2014) Software Development Process. [edition unavailable]. World Technologies. Available at: <https://www.perlego.com/book/1241202/software-development-process-pdf> (Accessed: 21 July 2022).

Matharu, A. (2018) Understanding Cryptocurrencies. [edition unavailable]. Business Expert Press. Available at: <https://www.perlego.com/book/859252/understanding-cryptocurrencies-pdf> (Accessed: 19 July 2022).

McDonald, O. (2021) Cryptocurrencies. 1st edn. Agenda Publishing. Available at: <https://www.perlego.com/book/2855332/cryptocurrencies-pdf> (Accessed: 19 July 2022).

Mueller, J. P. and Massaron, L. (2021) Machine Learning For Dummies. 2nd edn. Wiley. Available at: <https://www.perlego.com/book/2770772/machine-learning-for-dummies-pdf> (Accessed: 20 July 2022).

Nagy, Z. (2018) Artificial Intelligence and Machine Learning Fundamentals. 1st edn. Packt Publishing. Available at: <https://www.perlego.com/book/863703/artificial-intelligence-and-machine-learning-fundamentals-pdf> (Accessed: 20 July 2022).

Outsystems. (2022). What is Rapid Application Development? [Online]. Glossary. Available at: <https://www.outsystems.com/glossary/what-is-rapid-applicationdevelopment/> (Accessed 01 September 2022).

Scikit-learn, (2022). Linear Models. [Online]. Available at: https://scikit-learn.org/stable/modules/linear_model.html (Accessed, 10 September 2022).

Stavan, S. (2021) Predicting Stock Price with help of Machine Learning in Python. [Online]. Available at: <https://gist.github.com/ShahStavan/d0a440c4f1d15ae890d59429057a13af> (Accessed, 13 September 2022).

Summers, A. (2022) *Understanding Blockchain and Cryptocurrencies*. 1st edn. CRC Press. Available at: <https://www.perlego.com/book/3288917/understanding-blockchain-and-cryptocurrencies-pdf> (Accessed: 19 July 2022).

Tensorflow, (2022). *The Sequential Model*. [Online]. Available at: https://www.tensorflow.org/guide/keras/sequential_model (Accessed, 10 September 2022).

Williams, S. (2019) *Blockchain*. [edition unavailable]. Scribner UK. Available at: <https://www.perlego.com/book/1425597/blockchain-pdf> (Accessed: 19 July 2022).