



CHALMERS



GÖTEBORGS UNIVERSITET



Algoritmisk aktiehandel med marknadsanalys via Reddit och Yahoo Finance

DATX02-21-62

Kandidatarbete inom data- och informationsteknik

Emanuel Enberg
Anton Hagvall
William Sköld
Adam Söderholm
Viking Zandhoff Westerlund
William Öqvist

INSTITUTIONEN FÖR DATA- OCH INFORMATIONSTEKNIK

CHALMERS TEKNISKA HÖGSKOLA
Göteborg, Sverige 2021
www.chalmers.se

KANDIDATARBETE 2021

Algoritmisk aktiehandel med marknadsanalys via Reddit och Yahoo Finance

Emanuel Enberg
Anton Hagvall
William Sköld
Adam Söderholm
Viking Zandhoff Westerlund
William Öqvist



GÖTEBORGS
UNIVERSITET



CHALMERS

Institutionen för data- och informationsteknik

Avdelningen för datorteknik

CHALMERS TEKNISKA HÖGSKOLA

Göteborg, Sverige 2021

Algoritmisk aktiehandel med marknadsanalys via Reddit och Yahoo Finance

Emanuel Enberg
Anton Hagvall
William Sköld
Adam Söderholm
Viking Zandhoff Westerlund
William Öqvist

© Emanuel Enberg, Anton Hagvall, William Sköld, Adam Söderholm, Viking Zandhoff Westerlund, William Öqvist, 2021.

Handledare: Arne Linde, institutionen för data- och informationsteknik
Examinator: Carl-Johan Seger, institutionen för data- och informationsteknik

Kandidatarbete 2021
Institutionen för data- och informationsteknik
Avdelningen för datorteknik
Chalmers tekniska högskola
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Omslag: Aktiepris för Apple Inc. (AAPL) under perioden 2018-01-02 kl. 14.50-16.00 med intervall på en minut. Framtaget med hjälp av *mplfinance* för Python.

Typeset i L^AT_EX
Chalmers Reproservice
Göteborg, Sverige 2021

Abstract

Algorithmic trading is becoming increasingly common in today's digital society. We aim to investigate if algorithmic trading in combination with web scraping can yield a higher return than the market index, Dow Jones Industrial Average (DJIA). We have researched various inter- and intraday investment strategies, as well as analyzed the correlation between sentiment in Reddit comments on the one hand and news articles on the other hand, and stock prices. Here, we show that interday trading using the filter rule strategy performed better than DJIA, but also that no correlation was observed when sentiment analysis was performed on a small dataset utilizing our methods.

Keywords: investment, algorithmic trading, scraping, sentiment analysis, djia, reddit, yahoo

Sammanfattning

Algoritmisk handel är alltmer förekommande i dagens digitala samhälle. Detta projekt har som mål att undersöka om algoritmisk handel i kombination med marknadsanalys i form av webbskrapning kan visa sig prestera bättre än aktieindexet Dow Jones Industrial Average (DJIA). Projektet har undersökt diverse inter- och intradagshandelsstrategier, samt analyserat korrelationen mellan aktiepriser och sentimentet i Redditkommentarer och nyhetsartiklar publicerade av Yahoo Finance. Resultaten visar att intradagsstrategier bestående av filterregeln presterar bättre än DJIA under den undersökta perioden, men också att attitydsanalys av en relativt liten mängd data med listade metoder inte indikerar användbar korrelation. Sammanfattningsvis kan det inte sägas att algoritmisk handel med utförd attitydsanalys via listade metoder genererar högre avkastning än DJIA.

Nyckelord: investeringsalgoritm, algoritmisk handel, webbskrapning, attitydanalys, djia, investering, reddit, yahoo

Förord

Först och främst vill vi uttrycka vår tacksamhet till vår handledare Arne Linde för insiktsfull vägledning genom projektets gång. Vi vill även rikta ett tack till Lars Norén som hjälpte oss att sätta upp de fjärrstyrda datorerna på Chalmers tekniska högskola som använts under arbetet.

Er hjälp har varit av stor betydelse.

Ordlista

API - Application Programming Interface

DJIA - Dow Jones Industrial Average

PRAW - Python Reddit API Wrapper

VADER - Valence Aware Dictionary and sEntiment Reasoner

Innehåll

Figurer	xi
Tabeller	xi
1 Inledning	1
1.1 Bakgrund	1
1.1.1 Algoritmisk handel	2
1.1.2 Webbskrapning	3
1.2 Syfte	3
1.3 Arbetsätt	3
1.4 Avgränsningar	4
1.5 Relaterat arbete	4
1.6 Rapportens disposition	5
2 Teori	6
2.1 Aktier och aktiemarknaden	6
2.2 Investeringsalgoritmens uppbyggnad	7
2.2.1 Tekniska indikatorer	7
2.2.2 Handelsfrekvens	10
2.3 Tester mot historisk aktiedata	11
2.3.1 Fallgröpar vid historiska tester	11
2.4 Datainsamling via webbskrapning	11
2.4.1 Reddit och PRAW	12
2.4.2 Åsiktsmätning via sentimentanalys	12
2.4.3 Hemsidors byggstenar	13
2.5 Korrelationsanalys	14
3 Metod	16
3.1 Framställning av investeringsalgoritmen	16
3.1.1 Litteraturundersökning och implementation	16
3.1.2 Insamling av historiska aktiekurser	16
3.1.3 Testning av handelsstrategier	17
3.1.4 Fiktiv handel i realtid	17
3.2 Insamling och analys av media	17
3.2.1 Litteraturundersökning och datamängd	18
3.2.2 Insamling av artiklar och kommentarer	18
3.2.3 Databearbetning och sambandsanalys	18

4	Investeringsalgoritmerna	19
4.1	Hämtning av aktiedata	19
4.2	Tidiga simuleringsutfall	20
4.3	Implementerade handelsstrategier	21
4.3.1	Trendföljande strategier	21
4.3.2	Reversion- och trendföljande strategier	23
4.4	Slutgiltiga investeringsalgoritmerna	24
5	Webbskrapningen	27
5.1	Datainsamling och analys av Reddit	27
5.1.1	Inhämtning och sortering av data	27
5.1.2	Analys och statistisk signifikans	28
5.2	Artiklar från Yahoo News	28
5.2.1	Insamling av nyhetsartiklar	28
5.2.2	Sortering och analys av nyheter	29
6	Resultat	31
6.1	Investeringsalgoritmernas avkastning	31
6.1.1	Intradagshandel	31
6.1.2	Interdagshandel	32
6.2	Statistisk signifikans av insamlad media	33
6.2.1	Redditkommentarer	33
6.2.2	Yahoo News-artiklar	35
7	Diskussion	38
7.1	Investeringsalgoritmerna	38
7.1.1	Resultatens validitet	38
7.1.2	Konkurrensen om vinstmarginalerna	39
7.1.3	Kursutveckling på kort och lång sikt	39
7.1.4	Rivalitet på aktiemarknaden	40
7.1.5	Investeringsalgoritmernas förbättringsområden	41
7.1.6	Förbättringsmöjligheter i metodval	41
7.2	Webbskrapning	42
7.2.1	Sentiment i Redditkommentarer	42
7.2.2	Sentiment i nyhetsartiklar	43
7.2.3	Begränsningar och möjliga förbättringar	45
7.3	Framtida arbete	47
7.4	Etiska aspekter	47
7.4.1	Algoritmisk handels inverkan på samhället	48
7.4.2	Webbskrapningens etik och samhällspåverkan	49
7.5	Slutsats	49
	Litteratur	50
A	Aktiemarknadens mikrostruktur	I
A.1	Olika typer av ordrar	I
A.2	Orderboken	I

A.3	Transaktionskostnader	III
A.4	Hur mikrostrukturen hanterades i projektet	III
B	Investeringsalgoritmernas simuleringsresultat	V
B.1	Intradagshandel	V
B.2	Interdagshandel	VII
C	Dow Jones Industrial Average (DJIA)	IX

Figurer

4.1	Parameteruppsättningens påverkan på avkastning.	20
4.2	Köp- och säljsignal för glidande medelvärden.	21
4.3	Säljsignal baserad på lutningen av ett glidande medelvärde.	22
4.4	Köp- och säljsignal för filterregeln.	22
4.5	Köp- och säljsignal för Bollingerband kombinerat med filterregeln. . .	24
4.6	Köp- och säljsignal för den slutgiltiga intradagsalgoritmen.	25
6.1	Intradagsalgoritmen vs. DJIA - Testset.	32
6.2	Interdagsalgoritmen vs. DJIA - Testset.	33
A.1	Exempel på orderboken.	II
B.1	Intradagsalgoritmen vs. DJIA - Träningsset.	V
B.2	Intradagsalgoritmen vs. DJIA - Coronaset.	VI
B.3	Intradagsalgoritmen vs. DJIA - Testset.	VI
B.4	Interdagsalgoritmen vs. DJIA - Träningsset.	VII
B.5	Interdagsalgoritmen vs. DJIA - Coronaset.	VII
B.6	Interdagsalgoritmen vs. DJIA - Testset.	VIII

Tabeller

2.1	Exempelmeningar för VADER.	13
5.1	Aktier med flest och minst artiklar.	30
5.2	Tidsintervall för korrelation av artikelsentiment och aktiepris.	30
6.1	Investesteringsalgoritmens avkastning (intradagshandel).	31
6.2	Investesteringsalgoritmens avkastning (interdagshandel).	32
6.3	Korrelation (ρ) för DJIA samt utplockade aktier med högst korrelation.	34
6.4	Korrelation (ρ) för DJIA samt utplockade aktier med högst korrelation.	34
6.5	Korrelation mellan Redditkommentarers sentiment och volatilitet.	34
6.6	Korrelation mellan Redditkommentarers sentiment och avkastning	35
6.7	Artiklar publicerade vid tidpunkten t innan de största prisskillnaderna.	36
6.8	Artiklar publicerade vid tidpunkten t med de största sentimenten.	37
6.9	Korrelationskoefficienter för sambandet mellan artikelsentiment och aktiepris under specifika intervall runt om publiceringstid.	37
C.1	Lista på alla 30 företag i DJIA.	X

1

Inledning

När den amerikanska börsen öppnade 4 januari 2021 handlades aktier i Gamestop Corp. till en kurs på 19 dollar per aktie [1]. I slutet av januari stängde Gamestops aktiekurs på 325 dollar per aktie, vilket motsvarar en ökning på hela 1700 %.

Denna unika händelse rubricerades som slaget mellan Wall Street och privatinvestorare [2]. Dessa privatinvestorare koordinerade en köpråd på Gamestops aktie genom underforumet *wallstreetbets* på den sociala medieplattformen Reddit, med syfte att slå tillbaka mot de stora investeringsbankernas manipulation av aktiemarknaden [3].

Privata investorer kan kommunicera på Reddit och tillsammans agera i tusental- eller miljontal, och tillsammans kan de ha en betydande inverkan på en aktiekurs. Detta väcker intresse för frågan huruvida det går att kvantifiera data från exempelvis diskussioner på Reddit, eller annan media, för att använda det som en indikator i en investeringsalgoritm.

1.1 Bakgrund

Datavetenskap är en växande disciplin som har fått stor uppmärksamhet både som forsknings- och utbildningsområde såväl som inom företagsverksamhet [4]. Datavetenskap har i takt med sin egen utveckling funnit ett flertal olika tillämpningsområden, däribland på aktiemarknaden.

Aktiemarknaden är en av tre finansiella marknader som ligger till grund för den moderna samhällsekonomin [5]. Denna värdepappersmarknad gör det möjligt för företag såväl som privatpersoner att investera och bli delägare i företag världen över, samtidigt som företag kan använda aktiemarknaden för att ta in kapital.

För att mäta avkastning på aktiemarknaden mäts ofta den sammanlagda upp- eller nedgången hos en mängd aktier i ett aktieindex, varav ett exempel är Dow Jones Industrial Average (DJIA) [6]. DJIA är ett index på den amerikanska aktiemarknaden, och anses vara ett av de mest kända och viktigaste marknadsindexen i världen. Indexet innehåller 30 nationellt kända företag (se bilaga C) med stabil ekonomi som är noterade på aktiebörserna Nasdaq eller New York Stock Exchange. Indexets avkastning är en sammanställning över värdeutvecklingen i alla de aktier som ingår i indexet.

Aktiemarknaden har enligt Guo *et al.* [7, s. xiii] genom historien undergått ett flertal omstruktureringar tack vare den teknologiska utvecklingen. Från att ske i så kallade "trading pits" där människor samlades och handlade med aktier genom att använda sina röster och gestikulera med händerna, till att idag vara helt digitaliserat. Den finansiella handeln trädde in i en ny epok efter den världsomfattande finanskrisen 2007–2008. I denna nya epok menar Guo *et al.* att datavetenskap har en viktig roll, där kvantitativa handelsstrategier som bygger på avancerade algoritmer, statistisk modellering och dataanalys är av växande betydelse.

I anknytning till aktiemarknaden finns det mycket data att tillgå. De två centrala delarna i denna rapport, algoritmisk handel samt insamling och analys av media, fokuserar på data från två olika källor.

Gällande algoritmisk handel är fokusområdet att hämta och analysera data direkt från aktiemarknaden. Denna data ska användas för att implementera handelsstrategier som fattar beslut utifrån aktiedatan. Dessa handelsstrategier ska sedan användas i investeringsalgoritmen.

Under webbskrapningen ligger fokus på analys av data från nyheter och social media. Avsikten med analysen är att undersöka ifall diskussioner kring särskilda aktier kan korreleras med dessa aktiers prisutveckling och således användas som beslutsunderlag för att köpa och sälja dessa aktier.

Nedan följer en kort genomgång av begreppen algoritmisk handel och webbskrapning.

1.1.1 Algoritmisk handel

Algoritmisk handel innebär att datorstyrda algoritmer tillåts genomföra transaktioner på aktiemarknaden. Bergan och Devine [8] menar att algoritmisk handel som praktik inte är något nytt fenomen, då företag och organisationer har använt sig av algoritmisk handel i över två decennier. I deras artikel, publicerad redan 2005, diskuteras det att införande av algoritmisk handel kommer vara avgörande för att fondförvaltare och liknande institutionella aktörer ska kunna bevara sin konkurrenskraft på den finansiella marknaden.

Det finns ett flertal fördelar med algoritmisk handel gentemot traditionell handel. Dels är algoritmisk handel förprogrammerad, vilket innebär att investeringarna är konsekventa och ej styrda av känslor [9, s. 108]. Ett förekommande problem bland traditionella investerare är att de fattar emotionellt baserade investeringsbeslut som avviker ifrån deras tänkta handelsstrategi. Fortsättningsvis har datorstyrda algoritmer betydande fördelar i form av hastighet när det gäller att hämta och processera information [10].

Enligt hypotesen om effektiva marknader ska aktiepriset reflektera all information som gemene människa har tillgång till [11]. Syftet med att informationen ska vara transparent är således att ge alla intressenter samma utgångsläge, där ingen har

något försprång. När den offentliga informationen publiceras blir det därför centralt att på ett blixtsnabbt sätt kunna analysera denna information, varpå automatiserade system bör kunna ge en fördel på aktiemarknaden.

1.1.2 Webbskrapning

Insamling av data från hemsidor, så kallad webbskrapning [12], har existerat i stort sett lika länge som internet självt. Den första automatiserade webbskraparen skrevs redan 1993 med uppgift att undersöka hur stort internet var [13]. Webbskrapning möjliggör insamling av stora mängder data på kort tid, jämfört med arbetet och tiden som krävs för att en människa manuellt ska söka upp och spara relevant information. Webbskrapning innebär att ett program simulerar en användarinteraktion med en hemsida och programmatiskt plockar ut (skrapar) information från hemsidans innehåll.

Runt millenieskiftet släppte eBay ett av världens första publika applikationsprogrammeringsgränssnitt (API), för att underlätta för företag att göra transaktioner via eBays egna betalningssystem [14]. Likt webbskrapning möjliggör ett API insamling av stor mängd data på kort tid.

Det finns fördelar med att använda ett API framför egenutvecklade webbskrapningsprogram. En av dessa är att det är finansiellt fördelaktigt för den som tillhandahåller hemsidan då den överförda datamängden minskas, vilket är en prispunkt hos flertal företag som tillhandahåller servrar, exempelvis Amazon Web Services [15]. Idag har många hemsidor någon form av programmeringsgränssnitt för att underlätta för utvecklare att använda tjänster eller rentav bidra till vidareutveckling av tjänsterna.

1.2 Syfte

Projektets syfte är att undersöka huruvida det är möjligt att utveckla en investeringsalgoritm som tillsammans med marknadsanalys via webbskrapning kan generera högre avkastning än indexet Dow Jones Industrial Average.

1.3 Arbetsätt

I den inledande fasen av projektet ska en litteraturundersökning genomföras för att undersöka befintliga handelsstrategier. Med handelsstrategierna som utgångspunkt kommer resterande del av projektet ägnas åt att utvärdera, implementera, och kombinera dessa handelsstrategier för att framställa investeringsalgoritmen.

Parallellt med framställandet av investeringsalgoritmen ska data samlas in från Reddit och Yahoo Finance. Målet är att undersöka om uttalanden i social media och artiklar på nyhetssidor kan förutspå en akties framtida kursutveckling och därmed användas som en indikator i tidigare nämnt investeringsalgoritm.

1.4 Avgränsningar

Investeringsalgoritmen som utvecklas i detta arbete kommer att handla på den amerikanska aktiemarknaden. Detta beror dels på att det finns ett stort utbud av tjänster för att hämta aktiedata och simulera aktiehandel på denna marknad. Dessutom antas det att de företag som ingår i DJIA omnämns i större utsträckning på engelskspråkiga nyhetssidor och i forum, vilket tillgängliggör mer data för marknadsanalysen.

Utöver aktiemarknaden finns även bland annat marknader för valutor, räntor, kryptovalutor och finansiella derivat som exempelvis obligationer och optioner. Syftet med detta projekt är att utveckla en algoritm som endast handlar på aktiemarknaden. Den bakomliggande anledningen till detta är antagandet att nyheter och diskussioner på sociala medier har större inverkan på aktiekursen för den omnämnda aktien, än vad en nyhet skulle ha på till exempel en valutakurs.

I en studie gällande risk och avkastning för företag som engagerar sig i högfrekvent handel dras slutsatsen att den mest vitala konkurrensaspekten på aktiemarknaden är hastigheten [16]. Dessa hastigheter innebär att transaktioner kan ske inom loppet av mindre än en mikrosekund från det att en köpsignal aktiveras. Investeringsalgoritmen i detta projekt kommer ej att kunna uppnå dessa hastigheter, och avser ej att mäta sig med den typen av institutionell högfrekvent handel.

I en annan studie, gällande förmåga att förutspå aktiemarknadens rörelser menar Chong *et al.* [17] att maskininlärning i form av artificiella neurala nätverk är mer lämpat för uppgiften än tekniska indikatorer. Tillämpningen av maskininlärning för att förutspå aktiekurser ligger utanför detta projektets omfång och kommer inte undersökas i arbetet. Investeringsalgoritmens handelsstrategier kommer därför enbart baseras på tekniska indikatorer, vilket beskrivs vidare i underavsnitt 2.2.1.

För webbskrapningen begränsades datainsamlingen till Yahoo Finance och Reddits underforum wallstreetbets, investing, och stocks. Reddit är världens 19:e mest besökta hemsida och underforumet wallstreetbets har nästintill 10 miljoner registrerade användare [18], [19].

Korrelationsanalysen kommer endast bearbeta data som publicerats på Reddit och Yahoo Finance under arbetets gång.

1.5 Relaterat arbete

För att undersöka om högre avkastning än DJIA kan genereras genom en investeringsalgoritm i kombination med en marknadsanalys via webbskrapning har närliggande studier granskats.

I en studie från 2020 undersöks hur tekniska indikatorers förmåga att förutspå aktiekurser kan utökas med hjälp av restriktioner som är ekonomiskt förankrade [20]. Resultaten från studien visar att prestationen förbättrades för samtliga tekniska

indikatorer då de användes i kombination med dessa restriktioner. På samma sätt ämnar detta arbete att kombinera tekniska indikatorer, däremot med sentimentanalys snarare än de ekonomiska restriktioner som beskrivs i studien.

Ett flertal studier har analyserat sentiment på social media gentemot aktier. I en artikel från 2015 analyserades attityden i foruminlägg från Yahoo Finance för att förutspå aktiekurser [21]. Artikelförfattarna lyckades bland annat generera 2,07 % högre avkastning jämfört med en modell som endast baserade investeringar utifrån historisk aktiedata. En annan artikel från 2016 använde sig utav Twitter och analyserade sentimentet hos 2,5 miljoner tweets som nämnde specifika aktier i indexet S&P500 [22]. Slutsatsen var att en investeringsstrategi som köpte och sålde aktier utifrån sentimentanalysen genererade en årlig avkastning på 11-15 %. Även år 2019 skrevs en uppsats som analyserade sentimentet hos Redditkommentarer och framtida prisfluktuationer och drog slutsatsen att kommentartexterna korrelerar med framtida aktiepriser [23].

Ovan nämnda artiklar visar att korrelation mellan aktiekurser och diskussioner i forum kan utlysas och kan således ha användningspotential inom utveckling av investeringsalgoritmer. Dessa sentimentstudier utgår från en stor mängd data. Frågan kvarstår om samma slutsats kan nås med en mindre mängd data, vilken är den urskiljande aspekten i detta projekt.

1.6 Rapportens disposition

Fortsättningen av denna rapport är strukturerad på följande vis: I rapportens andra kapitel beskrivs bakomliggande teori kring aktier och aktiemarknaden, algoritmisk handel samt webbskrapning och statistisk analys.

I det tredje kapitlet beskrivs de olika tillvägagångssätten som användes i projektets båda delar. Dels processen att framställa en investeringsalgoritm, och även korrelationsundersökningen kring Reddit, finansnyheter och aktiekurser.

I fjärde kapitlet presenteras i detalj de resultat och slutsatser som varit vägledande i arbetet och som motiverat de val som gjorts för att nå den slutgiltiga investeringsalgoritmen. Likaså redogörs för den detaljerade arbetsprocessen för webbskrapningen i kapitel fem.

I kapitel sex sammanställs resultat gällande både investeringsalgoritmen och webbskrapningen, vilket följs med avslutande diskussion i kapitel sju. Här tas även etiska aspekter och framtida arbete upp, samtidigt som möjliga förbättringar kring dels datainsamling och databearbetning diskuteras.

Rapporten innehåller även tre bilagor. I bilaga A beskrivs aktiemarknadens mikrostruktur, samt hur den hanterades i projektet. I bilaga B återfinns samtliga resultat från investeringsalgoritmerna, endast resultat från testsetet redovisas i resultatkapitlet. I bilaga C återges alla företag som ingår i DJIA.

2

Teori

Detta kapitel behandlar relevant teori kopplat till aktiemarknaden, algoritmisk handel, webbskrapning, och statistisk analys. Hur fungerar egentligen aktiemarknaden? Hur skrapas information från webben? Hur beräknas korrelation mellan två variabler? Kapitlet ämnar ge all information som krävs för att förstå arbetets tillvägagångssätt och resultat.

2.1 Aktier och aktiemarknaden

En aktie är ett värdepapper som motsvarar en ägandedel i ett företag [24]. Aktier köps och säljs på aktiemarknaden, som i stor utsträckning utgörs av aktiebörser. Exempel på sådana aktiebörser är Stockholmsbörsen och New York Stock Exchange. Aktier är en vanligt förekommande sparform, och har historiskt sett genererat hög avkastning [25].

Trots att aktier som sparform historiskt sett har varit lönsamt så är det ej riskfritt. En akties pris är direkt kopplat till marknadens uppskattning av vad en aktie i ett bolag är värd [26, s.61]. Detta innebär att om marknadens uppfattning av ett bolag sjunker, kommer aktiepriset bete sig likartat. De som äger aktier i detta bolag kommer då se värdet på deras aktieportfölj minska. Om marknadens uppfattning av ett bolag istället förbättras kommer aktiekursen troligtvis att stiga. För att kontinuerligt uppnå hög avkastning på aktiemarknaden måste man alltså försöka förutspå aktiekursers prisutvecklingar.

Att förutspå marknadens rörelser är något som praktionärer länge har försökt med, men det har visat sig vara en tuff uppgift. Aktiemarknadens rörelser är vad som inom matematiken kallas för en icke-stationär tidsserie [27]. Detta innebär att dess rörelser ej följer en specifik trend eller förhåller sig till ett visst medelvärde.

Något som ytterligare försvårar förutspåendet av aktiemarknaden är att den genomgår konstant förändring i form av större regimskift [9, s. 91]. Dessa regimskift kan bero på justeringar i relaterade regelverk eller andra makroekonomiska faktorer. Förändringar av detta slag kan i stor utsträckning påverka hur lönsam en viss investeringsstrategi är.

En extern effekt som nyligen hade kraftig påverkan på aktiemarknaden var Covid-19-

pandemin. Utbrottet skapade oro hos allmänheten kring hur ekonomin och företag skulle klara följderna av viruset [28]. Den 16:e mars 2020 rasade Dow Jones Industrial Average 12,9 % vilket rapporterades som indexets tredje största nedgång någonsin under ett dygn [29]. I sviterna av coronapandemis marknadsnedgångar började en försiktig optimism spridas som gjorde att investerare varsamt började investera igen. Denna marknadsoptimism växte snabbt och resulterade i en kraftig återhämtning av aktiemarknaden under de efterföljande månaderna av coronapandemins utbrott [28].

För den intresserade läsaren beskrivs i bilaga A några fundamentala delar kring aktiemarknadens mikrostruktur som en investerare på aktiemarknaden måste ta hänsyn till. I synnerhet vid algoritmisk handel där flera transaktioner genomförs dagligen. I bilagan beskrivs förenklat olika typer av ordrar, hur orderboken är uppbyggd och vad transaktionskostnader är.

2.2 Investeringsalgoritmens uppbyggnad

Investeringsalgoritmen i detta projekt bygger på handelsstrategier som grundar sig i tekniska indikatorer. En teknisk indikator är en matematisk formel vars indata är en akties historiska prisutveckling, och ibland även transaktionsvolym [20]. Syftet med en teknisk indikator är att identifiera trender eller mönster i en akties tidsserie för att kunna förutspå aktiens framtida pris. I underavsnitt 2.2.1 beskrivs fyra olika tekniska indikatorer som använts under projektet.

2.2.1 Tekniska indikatorer

Handelsstrategier kan bygga på en mängd olika tekniska indikatorer. I detta arbete har fokus legat på indikatorerna glidande medelvärde, Bollingerband, RSI, samt filterregeln.

Ett **glidande medelvärde** är ett oviktat medelvärde av en akties pris, beräknat på ett givet historiskt tidsintervall [30]. Det glidande medelvärdet fungerar som ett filter för att minska inverkan av extremvärden i aktiepriset. Detta medför att medelvärdet fluktuerar mindre än aktiekursen, och ju längre tidsintervall det glidande medelvärdet beräknas på, desto mindre fluktuerar det [30].

Eftersom att medelvärdet baseras på historiska aktiepriser är det tidsfördröjt jämfört med det aktuella aktiepriset. Då det glidande medelvärdet har egenskaperna att det filtrerar bort extremvärden i aktiepriset och är tidsfördröjt, ger medelvärdet en indikation på en uppåtgående eller nedåtgående utvecklingsriktning för aktien, även kallat trend [20].

Ett glidande medelvärde definieras enligt följande [31]:

$$MA = \frac{1}{t_s} \sum_{i=0}^{t_s-1} p_{t-i} \quad (2.1)$$

där p_t är aktiepriset vid tiden t och t_s är tidsperioden det glidande medelvärdet beräknas på.

Bollingerband utvecklades av en framstående investerare vid namn John Bollinger under 1980-talet [32], [33]. Indikatorn bygger på idén att oförutsägbara variabler tenderar att anta värden inom en till två standardavvikelser från sitt medelvärde [32]. Utifrån denna teori innesluter Bollingerbandindikatorn ett område som är två standardavvikelser högre respektive lägre än det glidande medelvärdet av aktiepriset. Standardavvikelsen definieras i ekvation 2.14.

Traditionellt sett ger indikatorn en köpsignal då aktiepriset är lägre än två standardavvikelser från sitt medelvärde [32]. En investering görs således i förväntan om att aktiepriset ska återgå till sitt medelvärde. Med samma resonemang ger indikatorn en säljsignal då aktiepriset är högre än två standardavvikelser från medelvärdet. En strategi som placerar investeringar i förväntan om att aktiepriset ska återfå sitt medelvärde kallas för reversionstrategi, (från engelskans *mean reversion strategy*) [27].

Bollingerband definieras enligt [33]:

$$\begin{aligned} \text{Bollinger}_{\text{mitt}} &= MA(p_t, t_s), \\ \text{Bollinger}_{\text{övre}} &= MA(p_t, t_s) + 2\sigma(p_t, t_s), \\ \text{Bollinger}_{\text{undre}} &= MA(p_t, t_s) - 2\sigma(p_t, t_s) \end{aligned} \quad (2.2)$$

där MA är det glidande medelvärdet på aktiepriset p_t vid tiden t , över tidsperioden t_s . σ är standardavvikelsen på aktiepriset.

RSI (från engelskans *relative strength index*) introducerades av J. Welles Wilder Jr. 1978, och är ett mått på hastigheten i prisförändringarna för en aktie [34]. RSI tillämpas för att utvärdera om en aktie är överköpt eller översåld. Indikatorns värde placeras mellan extrempunkterna 0 och 100. Traditionellt sett är ett värde på över 70 en indikation på att aktien är överköpt och ett värde lägre än 30 en indikation på att aktien är översåld [32].

RSI definieras enligt följande [35, s. 65]:

Låt U och N vara en uppåtgående respektive nedåtgående prisskillnad i aktiepriset p_t vid tiden t . En uppåtgående period karaktäriseras enligt:

$$\begin{aligned} U_t &= p_t - p_{t-1}, \\ N_t &= 0. \end{aligned} \quad (2.3)$$

En nedåtgående period karaktäriseras enligt:

$$\begin{aligned} U_t &= 0, \\ N_t &= p_{t-1} - p_t. \end{aligned} \quad (2.4)$$

Medelvärden över samtliga U respektive N beräknas enligt:

$$\begin{aligned} U_{medel} &= \frac{1}{t_s} \sum_{i=0}^{i=t_s-1} U_{t-i}, \\ N_{medel} &= \frac{1}{t_s} \sum_{i=0}^{i=t_s-1} N_{t-i} \end{aligned} \quad (2.5)$$

där t_s är tidsperioden det glidande medelvärdet beräknas på.

Notera att ekvation 2.5 endast beräknas en gång för att få två initiala medelvärden. Därefter beräknas U_{medel} respektive N_{medel} med hjälp av ett utjämnat glidande medelvärde som beskrivs i ekvation 2.8.

RS (från engelskans *relative strength*) beräknas enligt:

$$RS = \frac{U_{medel}}{N_{medel}}. \quad (2.6)$$

Värdet av RSI-indikatorn fås enligt:

$$RSI = 100 - \frac{100}{1 + RS}. \quad (2.7)$$

När tiden passerat perioden t_s beräknas nästkommande U_{medel} respektive N_{medel} enligt:

$$\begin{aligned} U_{medel,t} &= \frac{(t_s - 1)U_{medel,t-1} + U_t}{t_s}, \\ N_{medel,t} &= \frac{(t_s - 1)N_{medel,t-1} + N_t}{t_s}. \end{aligned} \quad (2.8)$$

Därefter beräknas RS- samt RSI-värde på samma sätt som i ekvation 2.6 respektive 2.7.

Filterregeln utvecklades under 1960-talet av flera aktörer och bygger på konceptet att en aktie som har en uppåtgående utvecklingskurva troligtvis kommer fortsätta sin trend [36]. Mot den bakgrunden placeras traditionellt en köporder då aktiepriset stigit en viss procentsats från sitt lägsta värde. Med samma resonemang placeras en säljorder då aktiepriset fallit en procentsats från aktieprisets högsta värde.

Filterregelns köpsignal definieras enligt följande [36]:

Låt d_t vara skillnaden mellan aktiepriset p_t vid tiden t och aktieprisets lägsta värde inom intervallet $[t_0, t]$.

$$d_t = p_t - \min_{t_0 \leq i \leq t} p_i. \quad (2.9)$$

Filterregeln genererar en köpsignal på det första t som uppfyller:

$$\frac{d_t}{\min_{t_0 \leq i \leq t} p_i} \geq x_{köp} \quad (2.10)$$

där $x_{köp}$ bestämmer prisskillnaden som måste ske innan en köpsignal genereras.

Filterregelns säljsignal definieras enligt följande:

Låt d_t vara skillnaden mellan aktiepriset p_t vid tiden t och aktieprisets högsta värde inom intervallet $[t_0, t]$.

$$d_t = \max_{t_0 \leq i \leq t} p_i - p_t. \quad (2.11)$$

Filterregeln genererar en säljsignal på det första t som uppfyller:

$$\frac{d_t}{\max_{t_0 \leq i \leq t} p_i} \geq x_{sälj} \quad (2.12)$$

där $x_{sälj}$ bestämmer prisskillnaden som måste ske innan en säljsignal genereras.

2.2.2 Handelsfrekvens

En faktor som beaktas när en investeringsalgoritm utvecklas är hur frekvent den ska handla. Ett förekommande sätt att handla är intradagshandel, där högupplöst data såsom sekunddata eller minutdata används. Detta innebär att investeringsalgoritmen kan genomföra flertalet köp och sälj under samma dag och inte håller någon position över natten [9, s. 12]. Alternativet till intradagshandel benämns i denna rapport som interdagshandel. Vid interdagshandel tillåts investeringsalgoritmen vara mer långsiktig genom att hålla positioner över flera dagar, veckor eller till och med månader.

Fördelen med intradagshandel är att investeringsalgoritmens avkastning potentiellt sett kan hålla en jämnare utveckling [9, s. 151]. Eftersom intradagshandel innebär att investeringsalgoritmen köper och säljer mer frekvent så påverkas inte avkastningen i lika stor utsträckning av varje enskild affär, vilket gör att dåliga och bra affärer snabbare jämnar ut sig. Detta medför att en jämnare avkastning kan erhållas med intradagshandel.

Eftersom aktiepriset på kort sikt ibland inte alls behöver spegla hur det faktiskt går för ett bolag [37] skulle en strategi som bygger på att köpa aktier i företag som prognostiseras att gå bra framöver dock kunna tjäna på att vara mer långsiktig genom interdagshandel.

2.3 Tester mot historisk aktiedata

Innan en färdigutvecklad investeringsalgoritm används i realtid för att genomföra investeringar på aktiemarknaden kan algoritmen utvärderas via tester mot historisk aktiedata [9, s. 31]. Denna typ av test medför att investeringsalgoritmen kan prövas under riskfria förhållanden där algoritmens funktionalitet kan säkerställas, samt att dess förväntade avkastning kan estimeras.

2.3.1 Fallgropar vid historiska tester

Vid testning mot historisk data finns det enligt Chan [9, ss. 50-59] ett flertal fallgropar som kan leda till att testningen inte blir verklighetstrogen. En av dessa är att algoritmen ges information som inte ska vara tillgänglig vid en specifik tidpunkt, vilket kan medföra att algoritmen baserar sina investeringsbeslut på framtida data. Om algoritmen använder sådan information finns risken att man överskattar algoritmen.

En annan fallgrop som tas upp är överoptimering av parametrar. Chan förklarar att den första aspekten kring överoptimering är att man försöker optimera för många parametrar med för lite data. Den risk som uppstår då är att parametrarna blir överanpassade för att hantera avvikande brus i datan, vilket är suboptimalt. Den andra aspekten kring överoptimering som Chan nämner är att man justerar sin strategi för att uppnå bästa möjliga resultat, allt eftersom man genomför upprepade tester på samma data. Man utnyttjar alltså att man kan utvärdera vad som fungerar och inte, men i för hög utsträckning.

För att undvika överoptimering av parametrar föreslår Chan att man använder ett stort dataset vilket minskar risken för överanpassade parametrar, samt att datasetet delas upp i träningsset och testset. Träningssetet ska då användas för att anpassa parametrar och justera strategin, medan testsetet ska användas som den slutgiltiga utvärderingen. Testsetet ska alltså förbli orört tills dess att investeringsalgoritmen anses färdigutvecklad [9, ss. 53-54].

Slutligen varnar Chan för testning mot aktiekurser som sträcker för långt bakåt i tiden. Detta med hänsyn till de regimskift som nämndes i avsnitt 2.1. Sådana förändringar kan nämligen påverka hur lönsam en specifik handelsstrategi är.

2.4 Datainsamling via webbskrapning

För datainsamling har webbskrapning använts, vilket är en direkt översättning från engelskans *web scraping*. Begreppet innebär en systematisk inhämtning av data från en webbplats [12]. Det kan genomföras med ett API, eller genom att, med hjälp av mjukvara, simulera en användarinteraktion med en hemsida och extrahera relevant data som presenteras.

2.4.1 Reddit och PRAW

Reddit är en av världens största hemsidor [18], och en social nyhetssida. Likt andra sociala medier har användare möjlighet att lägga upp inlägg som kan röstas upp eller ned, och innehåll som övervägande röstas upp får större synlighet på plattformen än innehåll som övervägande röstas ned. Inläggen har i sin tur kommentarer, vars synlighet styrs enligt samma principer.

Inläggen på Reddit görs i så kallade *subreddits*, ett slags underforum. Ett sådant underforum är *wallstreetbets*, som i maj 2021 hade nästintill 10 miljoner registrerade användare och således utgör Reddits största underforum relaterat till aktiehandel [19]. Andra stora subreddits relaterade till aktiehandel är *stocks* och *investing* [38]-[39].

Reddits API möjliggör hämtning av all nödvändig information, till exempel information om en viss kommentar eller ett helt underforum. Det går således att läsa av alla inlägg och kommentarer som skrivs i en eller flera givna subreddits. Förutom titel och brödtext går det även att erhålla författare, poäng, med mera.

PRAW är ett programbibliotek till Python som ligger som ett hölje utanpå Reddits egna API [40]. Med hjälp av PRAW förenklas användningen av Reddits API. Exempelvis erbjuder PRAW en funktion som hämtar nyskrivna kommentarer och inlägg i utvalda underforum, såsom *wallstreetbets*.

2.4.2 Åsiktsmätning via sentimentanalys

Sentimentanalys är ett fält inom ‘Natural Language Processing’ (NLP) som syftar till att identifiera, extrahera och kvantifiera författarens attityd i en text via den subjektiva informationen i texten [41]. Det vill säga, är författaren till en kommentar eller artikel positivt eller negativt inställd gentemot ämnet och informationen som förmedlas?

Enligt en artikel från IBM, skriven år 2016, är det estimerat att ca 80 % av all världens data är icke-ordnad och ostrukturerad i form exempelvis social media och artiklar. [42]. Genom att använda sentimentanalys kan denna typ av data kvantifieras så att den lättare kan jämföras och utvärderas. Artikeln nämner även att under de senare åren har intresset för sentimentanalys ökat, särskilt hos företag för att förbättra till exempel kundupplevelser eller för att hitta ny affärsmöjligheter. Till exempel skulle ett företag kunna kvantifiera internetpublicerad text rörande dess verksamhet i avseende att undersöka samhällets syn på företaget. Detta kan sedan användas för exempelvis reklamkampanjer med syfte att svänga befolkningens åsikter, vilket skulle kunna leda till ett ökat intresse att investera i företagets aktie.

VADER är ett ordboks- och regelbaserat verktyg framtaget för sentimentanalys, speciellt anpassat för analys av text i sociala medier och är gratis att tillgå [43], [44].

Programmet beräknar fram ett sentiment för en given text. Beräkningen innehåller

andelen positiva och negativa samt neutrala ord i texten, och ett sammanlagt värde vilket betraktas som sentimentvärdet. Sentimentvärdet är normaliserat mellan -1 och +1, där +1 innebär 100% positiv attityd och vice versa. Enligt skaparna [44] anses sentimentvärdet vara det främsta verktyget inom attitydanalys.

I tabell 2.1 förekommer några exempelmeningar som visar att VADER tar hänsyn till negation, dubbel negation, värdeord och skiljetecken. Det förekommer även ett exempel som visar på en svaghet med VADER, nämligen ord med flera betydelser såsom engelskans *sick*.

Tabell 2.1: Exempelmeningar för VADER.

Exempelmening	Sentiment
AAPL is not a great company	-0,5096
AAPL is not not a great company	0,4014
AAPL is a solid company	0,1531
AAPL is a very solid company	0,2247
AAPL is a very solid company!	0,2926
AAPL is a sick company to own!	-0,5562

2.4.3 Hemsidors byggstenar

Moderna hemsidor är för det mesta skrivna med HTML, CSS, och JavaScript [45]. HTML kan förklaras som byggstenarna till hemsidan då den berättar för webbläsaren, oavsett vilken, vad som ska visas på skärmen. HTML använder sig av taggar, såsom `div`, `button`, `form` etc. som sedan fylls med informationen skaparen vill visa på hemsidan [46]. Exempelvis är en `div` en division på hemsidan vilken man kan fylla med andra taggar och annan information. För att hålla isär alla HTML-taggar på hemsidan brukar de tillhöra en klass eller ha ett ID. Nedan visas ett exempel med en `div` som tillhör klassen `exampleDiv` som innehåller en knapp med texten "Button title".

```
<div class="exampleDiv">
  <button>Button title</button>
</div>
```

CSS används därefter i syfte att orientera alla element och besmycka hemsidan, genom att exempelvis sätta färger, ändra font, specificera placering på skärmen och så vidare. Slutligen används programmeringsspråket JavaScript för att producera alla händelser på hemsidan, exempelvis en sökruta som visar sökresultat i realtid, formulärdata som ska skickas vidare, extra information som laddas och visas vid scrollning av hemsidan (även kallat asynkron laddning), med mera [47].

Selenium är ett open-source Python-verktyg som används primärt för applikationstester samtidigt som det även fungerar utmärkt inom webbskrapning [48]. Med hjälp av Selenium är det nämligen möjligt att imitera ett mänskligt besök på en hemsida, och därefter extrahera specificerad data direkt i Python, vilket tillåter en att

läsa och potentiellt spara information utan att manuellt behöva besöka hemsidorna. Genom att undersöka en hemsidas HTML-kod är det därav möjligt att inhämta relevant information automatiskt, genom att referera till de specifika klasserna som innehåller uppgifterna som efterfrågas.

2.5 Korrelationsanalys

Ett sätt att undersöka korrelationen, eller sambandet, mellan variabler är med Pearsons korrelationskoefficient [49]. Denna siffra är alltid mellan -1 och +1. -1 visar ett direkt, det vill säga starkt, negativt linjärt samband och +1 visar det motsatta, alltså ett starkt positivt och linjärt samband. Ju lägre absolutbeloppet av korrelationskoefficienten är, desto lägre samband har de två variablerna. En koefficient på noll visar att variablerna inte har någon som helst relation. Innan Pearsons korrelationskoefficient kan definieras behövs några grundläggande statistiska samband först definieras.

Kovariansen visar förhållandet mellan slumpmässiga variabler, det vill säga om de rör sig tillsammans i samma riktning. En positiv kovarians visar på positivt linjärt beteende, medan negativ kovarians visar inverterat samband. Med andra ord berättar kovariansen riktningen på ett linjärt samband. Kovarians $\text{Cov}(X, Y)$ definieras som [50]:

$$\text{Cov}(X, Y) = \frac{1}{N} \sum_{i=0}^N (X_i - \bar{X})(Y_i - \bar{Y}) \quad (2.13)$$

där \bar{X} och \bar{Y} är medelvärdet av X och Y och N är antalet sampel. Vidare berättar **standardavvikelsen** hur mycket alla värden i en specifik mängd avviker från deras gemensamma medelvärde. Detta är intressant för att se hur pass stor variation det är mellan de olika värdena och om det finns extremvärden. Standardavvikelse σ definieras som [51]:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2.14)$$

där

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.15)$$

Pearsons korrelationskoefficient ρ , kan nu, med hjälp av ovanstående information, definieras som kovariansen hos två variabler (X och Y) dividerat med produkten av deras respektive standardavvikelse:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (2.16)$$

Värt att notera här är att formeln för standardavvikelse, ekvation 2.14 och 2.15, endast gäller för kompletta datamängder. Om beräkningen enbart skulle använda ett urval av en datamängd istället så måste antalet sampel (N) minskas med en, det vill säga $N - 1$. Detta kallas för Bessels korrektion, som används för att ta med en eventuell variation som finns i hela datamängden men som kanske inte representeras fullständigt i urvalet [52].

Korrelationskoefficienten ρ är beroende av antalet sampel. En hög koefficient med endast två sampel visar inte ett lika starkt statistiskt samband som en hög koefficient beräknad med åtskilliga sampels. Detta symboliseras av **signifikansen**, p , där lägre värde indikerar en mer pålitlig korrelationskoefficient [53]. Signifikansen är sannolikheten att slumpmässigt placerade datapunkter skulle resultera i en lika stor korrelation eller större.

För att beräkna signifikansen kan t-test användas, enligt ekvation 2.17 [54], där n motsvarar antalet sampels som ingick i respektive test och ρ är sampelkorrelationskoefficienten som beräknas med ekvation 2.16. Från t-värdet erhålls ett p-värde genom en t-tabell [55, s. 472]. Signifikansnivån i denna studie är satt till 0,05. Detta bedömdes vara tillräckligt litet för att inte utesluta en faktisk korrelation, men samtidigt tillräckligt högt för att inte innebära för stor risk för missledande resultat.

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}} \quad (2.17)$$

Vid korrelationsanalys kan en akties **volatilitet** användas, liksom det faktiska aktiepriset. En akties volatilitet avser måttet på aktiens prisvariation. Ett välkänt volatilitetsmått är aktieprisets aktuella standardavvikelse från det nutida genomsnittspriset [56]. Med hänsyn till mängden data kan det därmed vara mer relevant att undersöka volatiliteten istället för aktiepriset. En akties pris varierar markant över en längre tidsperiod, varpå en jämförelse av två momentanvärden kan medföra en stor prisskillnad. Volatiliteten tar däremot denna variation i beaktning genom att enbart indikera de värden som avviker från aktiens normala beteende, det vill säga medelvärde. Följaktligen kan missvisande momentanvärden antas vara ett problem vid en ringa datamängd, eftersom en mindre mängd data riskerar att enbart återge upp- och nedgångar i aktiens pris.

3

Metod

Utförandet har skiljt mellan de olika delarna i projektet. Investeringsalgoritmen har fokuserat på försök och misstag medan webbskrapningen har undersökt den insamlade datan med ett flertal tillvägagångssätt för att hitta en möjlig korrelation. Nedan beskrivs den grundläggande metodik som de båda delarna utgick ifrån.

3.1 Framställning av investeringsalgoritmen

Framställningen av investeringsalgoritmen i detta projekt bestod huvudsakligen av tre delar: Datainsamling, implementation och testning mot historisk data. I slutet av projektet implementerades även investeringsalgoritmen för att utföra fiktiv handel i realtid.

3.1.1 Litteraturundersökning och implementation

Projektet inleddes med en litteraturundersökning för att studera befintliga handelsstrategier. Litteraturundersökningen utgjorde grunden för vilka strategier som senare implementerades. Undersökningen pågick under större delen av arbetet, parallellt med implementation av de handelsstrategier som undersöktes.

3.1.2 Insamling av historiska aktiekurser

För att kunna testa de handelsstrategier som implementeras behövdes högupplöst data med historiska aktiekurser. Tillgången till högupplöst data var begränsad, således behövde datan samlas in.

För att samla in datan användes en tjänst hos Alpaca Securities LLC [57]. Via denna tjänst kunde en anslutning sättas upp direkt till aktiemarknaden. Med hjälp av en fjärrstyrd dator lokaliserad på Chalmers tekniska högskola kunde datahämtningen kopplas upp och pågå kontinuerligt under arbetets gång. Relevant aktiedata erhöles i ett enda flöde, vilket innebar att datan behövde processeras innan den kunde användas i miljön där historiska tester skulle genomföras. I slutet av varje börsdag sorterades därför datan i mappar som tillhörde respektive aktie.

I ett senare skede dömdes den egeninsamlade datan otillräcklig, vilket innebar att

data istället köptes. Denna data innehöll historisk minutdata från 2004 till nutid för samtliga aktier som under projektet ingick i DJIA.

Med hänsyn till det som diskuteras i underavsnitt 2.3.1 delades den införskaffade datan upp i ett träningsset och ett testset. Datan delades även in i ett separat dataset för perioden där börsen föll tungt i samband med coronautbrottet, hädanefter coronasetet. Träningssetet sträckte sig mellan perioden 2018-01-02 till 2020-01-31, coronasetet mellan 2020-02-01 till 2020-04-30, och slutligen testsetet mellan 2020-05-01 till 2021-03-31.

3.1.3 Testning av handelsstrategier

För att kunna testa strategierna mot historiska aktiekurser skapades en simuleringsmiljö. Simuleringsmiljön skrevs i programmeringsspråket Python, där *Backtrader* användes. *Backtrader* är ett ramverk i Python som användes för att underlätta uppbyggnaden av infrastrukturen till simuleringsmiljön [58]. *Backtrader* är bland annat utrustat med en mängd förprogrammerade tekniska indikatorer, samt metoder för att simulera realtidshandel utifrån den historiska datan som tillförs.

Med *Backtrader* som grund byggdes en iterativ simuleringsmiljö där en mängd parameteruppsättningar kunde specificeras för den handelsstrategi som skulle testas. I simuleringarna försåg handelsstrategierna med en datapunkt per aktie i taget, vilket motsvarar hur realtidshandel skulle gått till.

Strategierna simulerades först på träningssetet. Resultaten från varje simulering på träningssetet sparades undan i en separat fil, som sedan kunde analyseras för att finna den optimala parameteruppsättningen för strategin som testades. Som en slutgiltig utvärdering av handelsstrategierna simulerades de på coronasetet och testsetet.

3.1.4 Fiktiv handel i realtid

Handelsstrategier som uppvisat potential vid historiska tester implementerades för fiktiv realtidshandel. För detta användes återigen Alpaca. Via Alpaca kunde de implementerade strategierna kopplas till ett konto som försetts med ett fiktivt kapital. Kontot anslöt direkt till aktiemarknaden, och på så vis kunde handel i realtid simuleras.

3.2 Insamling och analys av media

Framställningen av webbskrapningsarbetet inleddes, i likhet med investeringsalgoritmen, genom att undersöka litteratur av tidigare arbeten på ämnet. Likaså diskuterades vilka hemsidor som skulle användas för analysen. Kärnan i dataanalys grundar sig i datan som ska analyseras. ”Skräp in, skräp ut” är ett populärt uttryck gällande statistik och analys, då även korrekt utförda analyser inte kan rädda dålig data [59]. Vid webbskrapningsanalysen var det därför viktigt att välja vilken typ av

data som skulle samlas in.

3.2.1 Litteraturundersökning och datamängd

Tidigare studier har visat en klar korrelation mellan bland annat Redditchmentarer eller Twittermeddelanden och aktiepriser (se avsnitt 1.5). Dessa studier har använt sig av en stor mängd historisk data med miljontals datapunkter. För att inte upprepa tidigare studier beslutades det tidigt i projektets skede att analys endast skulle göras på data som samlats in under projektets gång. Analysen skulle därför utgå från ett mindre dataset vilket möjligtvis kunde rendera annorlunda resultat än tidigare undersökningar.

3.2.2 Insamling av artiklar och kommentarer

För att kunna undersöka eventuella samband mellan omnämningar av aktier i Redditchmentarer, kommentarens attityd, samt aktiens avkastning och volatilitet är det nödvändigt att samla in data som kan beskriva detta över tid. Med ovanstående i åtanke samlades Redditchmentarer in från följande underforum: wallstreetbets, investing, och stocks. Insamlandet fortskred under projektets gång.

Gällande nyhetsartiklar användes Yahoo Finance. Yahoo, som även innefattar Yahoo Finance, är världens 12:e mest besökta hemsida [18], och Yahoo Finance aggregerar nyheter från en mängd välkända, större tidningar och nyhetsbyråer. Detta innebär att Yahoo Finance erbjuder diversifiering i nyheterna jämfört med en undersökning hos endast enstaka tidningar eller nyhetsbyråer som exempelvis Bloomberg eller The Wall Street Journal.

Artiklarna som listas på Yahoo Finance innehåller dels artikeln i sig, men även tillhörande information, exempelvis publiceringsdatum och nyhetskälla. Denna data samlades in under mars och april månad.

3.2.3 Databearbetning och sambandsanalys

När datainsamlingen bedömdes vara komplett påbörjades den statistiska analysen. Inledningsvis sorterades all data in i olika grupper, såsom kommentarer och nyheter publicerade under börsens stängningstid, artiklar och kommentarer relaterade till aktier som ansågs ha för låg signifikans på grund av dess låga antal. Därefter tillämpades de analytiska ekvationer som omnämns i avsnitt 2.5, med syfte att integrera en potentiellt stark korrelation med investeringsalgoritmen.

4

Investeringsalgoritmerna

Som nämnades i metoden har arbetet med investeringsalgoritmerna följt en iterativ arbetsgång där handelsstrategier har testats, och antingen förkastats eller vidareutvecklats. I detta kapitel beskrivs dels tillvägagångssättet för att hämta data, samt vilka handelsstrategier som implementerades och hur de genererar köp- och säljsignaler. Avslutningsvis beskrivs i detalj de slutgiltiga investeringsalgoritmerna för intradagshandel och interdagshandel.

4.1 Hämtning av aktiedata

Intentionen med detta projekt var inledningsvis att handla aktier med hög frekvens, så kallad högfrequenshandel. Detta medförde att data med hög upplösning behövde införskaffas för att historiska tester skulle kunna genomföras. Kostnaden för att köpa denna typ av data täcktes inte av den budget som projektet tillförsågs med, vilket innebar att datan behövdes samlas in.

Vid insamling av den högfrekventa datan registrerades pris och tidpunkt för samtliga transaktioner som genomfördes i någon av aktierna som ingår i DJIA. Att transaktionerna registrerades i samma ögonblick som de genomfördes innebar att tidsintervallen mellan varje inhämtad datapunkt blev oregelbundna. Detta visade sig vara problematiskt, eftersom att flera av de tekniska indikatorerna skulle beräknas på ett visst tidsintervall och inte på ett antal transaktioner.

För att undvika problemet med oregelbundna intervall i tidsserierna så samplades datan om till sekundupplösning genom att endast registrera den sista transaktionen som genomfördes under varje enskild sekund. Vid undantaget att ingen transaktion genomfördes vid en given sekund skapades en datapunkt för sekunden med hjälp av interpolation. Den nya datapunkten tilldelades samma värde som föregående registrerade datapunkt, se ekvation 4.1 [60, s. 65].

$$\hat{T}_t = T_{t-1} \tag{4.1}$$

där \hat{T}_t är den interpolerade datapunkten.

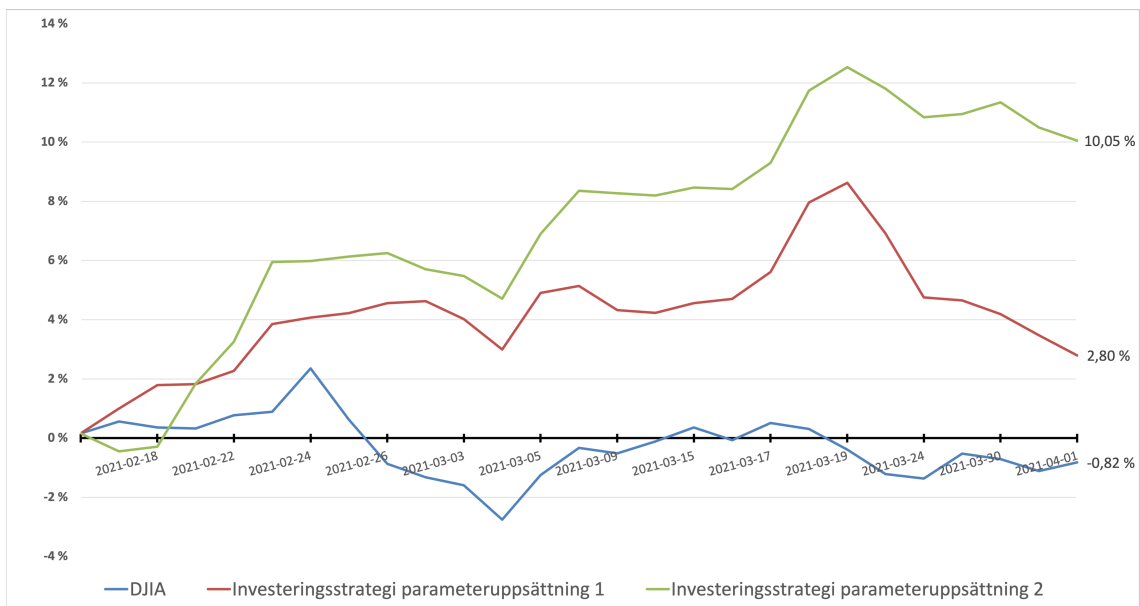
Datan samplades även om till upplösningar på 1, 5, 15 och 30 minuter utifall att

det skulle visa sig användbart senare i projektet.

4.2 Tidiga simuleringsutfall

Utifrån resultat av tidiga simuleringar konstaterades det att investeringsalgoritmen presterade bättre på minutupplösning än sekundupplösning, därav togs beslutet att övergå till minutupplöst data. Detta innebar att kostnaden för att köpa data minskade avsevärt vilket öppnade möjligheten att köpa data, vilket senare blev ett måste.

I samband med att olika investeringsalgoritmer simulerades visade flera av resultaten att en liten parameterskillnad gav stor skillnad i avkastningen. I figur 4.1 visas en simulering där samma investeringsalgoritm får två noterbart olika resultat. Investeringsalgoritmen i denna simulering bestod av sex parametrar där skillnaden mellan de två resultaten endast berodde på en parameter.



Figur 4.1: Parameteruppsättningens påverkan på avkastning för samma investeringsalgoritm.

Med resultatet som visas i figur 4.1 blev det tydligt att den data som samlats in under arbetets gång ej kommer kunna användas för att genomföra verklighetstroga simuleringar. Med hänsyn till risken för överanpassning, som beskrivs i underavsnitt 2.3.1, inköptes därför ett nytt dataset innehållande aktiekurser från 2004 till 2021 för samtliga företag som ingår i DJIA.

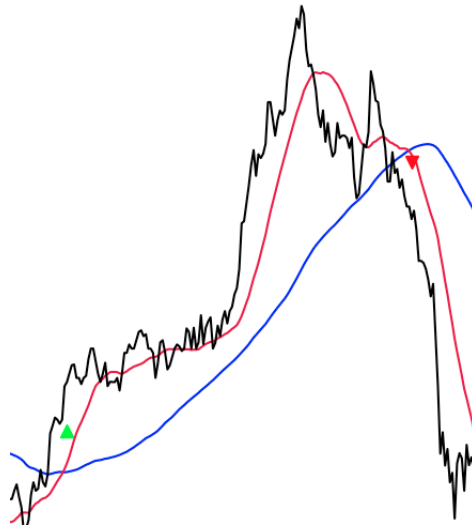
4.3 Implementerade handelsstrategier

Som nämndes i avsnitt 3.1 gjordes en litteraturundersökning för att undersöka befintliga handelsstrategier. Detta gjordes dels för att få inspiration, samt för att få en grundläggande förståelse för strategierna. Litteraturundersökningen inriktades på de tekniska indikatorer som beskrivs i underavsnitt 2.2.1. Med utgångspunkt i dessa tekniska indikationer inleddes implementeringen av de handelsstrategier som följer nedan.

4.3.1 Trendföljande strategier

Den första handelsstrategin som implementerades var en kombination av två glidande medelvärden. Genom att kombinera två glidande medelvärden som beräknas på ett kort respektive långt tidsintervall kan, enligt [31], köplägen identifieras då det korta medelvärdet skär det långa medelvärdet underifrån.

I figur 4.2 representerar den blå grafen det långa medelvärdet och den röda grafen det korta. Den svarta grafen representerar aktiekursen som dessa medelvärden beräknas på. Investeringsalgoritmen får en köpsignal när den röda grafen skär den blå grafen underifrån och placerar därefter en köporder, vilket representeras av den gröna pilen. Säljorden placeras på samma sätt efter att den röda grafen skär den blå grafen ovanifrån, vilket representeras av den röda pilen.

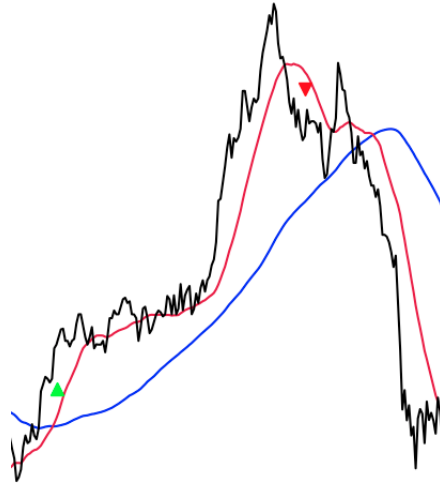


Figur 4.2: Köp- och säljsignal för glidande medelvärden, Apple Inc. den 12 januari 2018.

Vid simuleringar av strategin som beskrivs ovan observerades det vid flera tillfällen att aktiekursen hade hunnit falla påtagligt innan säljsignalen genererades. Ett exempel på en sådan säljsignal kan ses i figur 4.2, där aktiekursen har haft en nedåtgående trend under en längre tidsperiod innan det korta medelvärdet skär det långa. Utifrån dessa observationer implementerades en ny strategi med avsikt att minska

den tidsperiod som aktiepriset hinner falla innan en säljsignal genereras.

Den nya strategin bygger på samma grundteori som föregående strategi. Skillnaden är att säljsignalen genereras då lutningen på det korta medelvärdet har varit negativt under ett givet tidsintervall. Strategin åskådliggörs i figur 4.3.



Figur 4.3: Säljsignal baserad på lutningen av ett glidande medelvärde, Apple Inc. den 12 januari 2018.

Skillnaden mellan dessa två strategier tydliggörs vid en jämförelse av figur 4.2 och figur 4.3. I den senare figuren placeras säljorden i ett tidigare skede vilket resulterar i en större vinst vid aktieaffären.

Ytterligare en strategi som implementerades var filterregeln. Köpsignalen i figur 4.4 genereras då aktiepriset har stigit en given procentsats från sitt lägsta värde. Därefter placeras en köporder i förväntan om att aktiepriset ska fortsätta sin uppåtgående trend. När aktiepriset sedan har fallit en given procentsats från sitt högsta värde genereras en säljsignal. En säljorder placeras då i förväntan om att aktien ska fortsätta sin nedåtgående trend.



Figur 4.4: Köp- och säljsignal för filterregeln, Microsoft den 13 juni 2018.

4.3.2 Reversion- och trendföljande strategier

I projektet utvecklades även en handelsstrategi som kombinerar de tre tekniska indikatorerna Bollingerband, RSI och filterregeln. Som beskrivs i underavsnitt 2.2.1 är Bollingerband en indikator som används i reversionstrategier, medan filterregeln är en trendföljande indikator.

Att kombinera Bollingerband och RSI är redan en förekommande handelsstrategi [61]. RSI-indikatorn används då för att förstärka indikationen om att aktiepriset sannolikt kommer återvända till sitt medelvärde då aktiepriset ligger utanför något av Bollingerbandets värden. Bollingerbanden genererar en köpsignal då aktiepriset är lägre än Bollingerbandets undre värde och RSI-indikatorn kan då antingen bekräfta eller avslå köpsignalen. Det är först när RSI-indikatorn har bekräftat Bollingerbandets köpsignal som en köporder ska placeras. På samma sätt kan RSI-indikatorn bekräfta eller avslå en säljsignal från Bollingerbanden.

För att RSI-indikatorn ska bekräfta en köpsignal från Bollingerbandet måste indikatorn anta ett värde på 30 eller lägre, vilket indikerar att aktien är översåld. I en översåld aktie kommer fler aktiehandlare se en investeringsmöjlighet och aktiepriset kommer därför sannolikt att stiga. Då en köpsignal från Bollingerbandet istället avslås har RSI-indikatorn ett värde över 30. Aktien är således inte översåld och aktiepriset riskerar att fortsätta falla innan en reversion till aktieprisets medelvärde sker. Med liknande resonemang kan en säljsignal bekräftas eller avslås beroende på om aktien är överköpt eller inte.

Bollingerband kan även användas för att identifiera perioder med låg volatilitet genom att kvoten mellan det övre och undre värdet i indikatorn minskar [62]. Ett sådant tillfälle visualiseras i figur 4.5. Aktiekursen går från en period med hög volatilitet till en period med lägre volatilitet, vilket medför att avståndet mellan Bollingerbandets värden blir allt mindre.

Enligt [62] kommer aktier som haft låg volatilitet under en period att påbörja en positiv eller negativ trend då aktiekursen träder utanför det övre eller undre värdet på Bollingerbanden. I lågvolatila tillstånd genereras därför en köpsignal då aktiekursen stiger ovanför Bollingerbandets övre värde, i förväntan om att den påbörjade trenden kommer fortsätta uppåt. Kombinationen med Bollingerband och RSI-indikatorn inaktiveras då aktien infinder sig i ett tillstånd med låg volatilitet.

Efter implementation av handelsstrategin som beskrivs ovan genomfördes simuleringar för att undersöka hur hög avkastning strategin gav. Vid simuleringarna upptäcktes det att flera av de köpsignaler som genererats vid lågvolatila tillstånd gav upphov till förlustaffärer. Därför utökades handelsstrategins köpsignal med filterregeln vid dessa tillstånd.

Den utökade köpsignalen innebar att aktiekursen behövde överstiga Bollingerbandets övre värde. Samtidigt var aktiekursen tvungen att stiga med en viss procentsats från det lägsta värde som aktiepriset antog i sitt lågvolatila tillstånd. Detta medförde att indikationen på en uppåtgående trend behövde vara kraftigare för att en köp-

signal skulle genereras. Säljsignalen som genereras efter detta typ av köp genereras när aktiepriset fallit en viss procentsats från sitt högsta värde.

Sammantaget kunde den utökade handelsstrategin generera två köp- och säljsignaler med olika karaktär. I figur 4.5 visas två sekvenser, en för vardera köp- och säljsignal. De två röda graferna i figuren representerar Bollingerbandets övre respektive undre värde, och den blå-röda, randiga grafen representerar aktieprisets medelvärde. I den undre delen av figuren visas RSI-indikatorns värde, där de två streckade linjerna markerar tröskelvärdena 30 och 70.



Figur 4.5: Köp- och säljsignal för Bollingerband kombinerat med filterregel, JPMorgan Chase & Co den 18 juni 2018.

I den första sekvensen har aktiepriset fallit under Bollingerbandets lägre värde, vilket genererar en köpsignal. RSI-indikatorn har samtidigt antagit ett värde under 30 vilket bekräftar köpsignalen. Således placeras en köporder. Efter att denna typ av köporder har genomförts placeras sedan en säljorder då aktiepriset stigit utanför Bollingerbandets övre värde samtidigt som aktien är överköpt, vilket sker då RSI har antagit värdet på över 70.

I nästkommande sekvens placeras en köporder efter att aktiekursen ökat kraftigt ifrån ett lågvolatilt tillstånd. Säljorden placeras sedan när aktiepriset fallit en viss procentsats från sitt högsta värde.

4.4 Slutgiltiga investeringsalgoritmerna

Handelsstrategierna som beskrevs i avsnitt 4.3 simulerades för både intradags- såväl som interdagshandel. Resultaten från simuleringarna visade att olika handelsstrategier lämpar sig olika bra för olika handelsfrekvenser. Således skapades två slutgiltiga investeringsalgoritmer utifrån två olika handelsstrategier, en för intradagshandel och

en för interdagshandel. Parametrarna som används nedanför är de som gav upphov till bäst resultat vid simuleringar. Parametrarna presenteras löpande i texten med kursiv stil.

Den slutgiltiga investeringsalgoritmen för intradagshandel bygger på en modifikation av handelsstrategin som beskrivs i underavsnitt 4.3.2. Algoritmen identifierar aktier med låg volatilitet för att därefter investera då aktiepriset stiger markant. För att urskilja perioder med låg volatilitet används Bollingerband, vilket beräknas över ett 8-minuters intervall. För att en aktie ska kategoriseras som en aktie med låg volatilitet ska kvoten mellan Bollingerbandets över och undre värde vara mindre än 1,001 under minst 6 minuter.

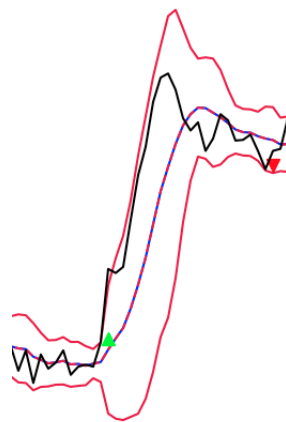
Låg volatilitet definieras därför enligt:

$$LV = \begin{cases} 1, & \text{om } \frac{Bollinger_{\text{övre},i}}{Bollinger_{\text{undre},i}} \leq 1.001 \quad \forall i \in [t-6 \leq i \leq t] \\ 0, & \text{annars,} \end{cases} \quad (4.2)$$

där t är tiden.

I samtliga aktier där $LV = 1$ i ekvation 4.2 kan investeringsalgoritmen placera en köporder. En köpsignal genereras då aktiepriset stiger mer än 0,5 % från sitt lägsta värde, samtidigt som aktiepriset överstiger Bollingerbandets övre värde. Ett sådant scenario åskådliggörs i figur 4.6. En investering görs i förväntan om att aktiepriset ska fortsätta sin uppåtgående trend efter ett lågvolatilit tillstånd. Vid en köpsignal placeras en order till marknadspris och investerar 25 % av investeringsalgoritmens totala kapital.

Vidare genereras en säljsignal då aktiepriset sjunker mer än 0,2 % av aktieprisets högsta värde. Vid en säljsignal placeras en säljorder till marknadspris innehållande samtliga aktier algoritmen innehar i företaget. En köp- och säljsignal illustreras i figur 4.6



Figur 4.6: Köp- och säljsignal för den slutgiltiga intradagsalgoritmen, JPMorgan Chase & Co den 23 april 2018.

Vid situationer där investeringsalgoritmen innehar en position i en aktie, utan att en säljsignal genereras innan aktiemarknadens stängning, placerar algoritmen en säljorder oavsett aktiens utveckling till marknadspris två minuter före stängning. Denna säljorder placeras i syfte om att investeringsalgoritmen inte ska inneha positioner i aktier under tiden då marknaden är stängd. Detta eftersom att aktiekurser, enligt [63], i regel är volatila vid marknaden öppning. Således kan investeringsalgoritmen riskera en förlustaffär.

Den slutgiltiga investeringsstrategin för interdagshandel bygger på filterregeln som beskrivs detaljerat i underavsnitt 2.2.1. Algoritmen placerar en köporder till marknadspris då aktiepriset stigit $0,5$ % av sitt lägsta värde. Vid en köpsignal placerar investeringsalgoritmen en köporder motsvarande 10 % av sitt totala kapital. En säljorder placeras då aktiepriset fallit $0,1$ % av sitt högsta värde.

5

Webbskrapningen

Att träna algoritmer och analysera data, något som tidigare omnämnts i underavsnitt 2.3.1, har sina risker. Algoritmen eller analysmetoden skulle exempelvis kunna ges tillträde till information som egentligen är otillgänglig vid den specifika tidpunkten. Detta kan i sin tur leda till missvisande resultat. Om en investeringsalgoritm ska agera med kommentarer och nyheter som underlag kan den endast göra det efter att kommentaren eller nyheten i fråga har publicerats. Således är det huvudsakligen intressant att undersöka korrelationerna framåt i tiden, det vill säga vad som har hänt med aktiekursen efter publicering.

5.1 Datainsamling och analys av Reddit

Nedan motiveras val som underlättat analysen av Redditkommentarer. Vidare presenteras även arbetsprocessen som lett till resultaten som presenteras i kapitel 6.2.

5.1.1 Inhämtning och sortering av data

Genom PRAW samlades kommentarer in löpande, allt eftersom de skrevs. Varje kommentar som levererades av PRAW sparades i en Postgresdatabas tillsammans med metadata, såsom kommentarens ID (en unik kod som identifierar en kommentar på Reddit) och tidsstämpel för publiceringen. Om kommentaren gjorts på ett inlägg som inte tidigare sparats, sparades även det tillhörande inlägget i en separat tabell.

Postgres, en databashanterare, valdes som datalagringsmetod eftersom den ansågs vara ett mer användbart alternativ jämfört med vanlig datalagring, då bland annat användarskrivna procedurer tillåts för att enklare arbeta med data. Dessutom underlättade Postgres analysen genom SQL (Structured Query Language), ett programmeringsspråk som används för att aggregera och transformera strukturerad data.

Den insamlade datan sorterades upp i olika kategorier, kommentarer och inlägg. Vidare så sparades kommentarer och inlägg som omnämnde aktier i DJIA i separata vyer i databasen för att förenkla analysen.

Med hjälp av funktionaliteten i Postgres skapades en hjälpfunktion som automatisk räknade fram nästa börsdag utifrån kommentars eller inläggs tidsstämpel. Exem-

pelvis kunde tabeller med innehållet kommentar, aktie och nästa börsdag hämtas direkt ur databasen istället för att bearbetas fram i Python.

5.1.2 Analys och statistisk signifikans

För att fördjupa undersökningen av korrelationen mellan kommentarers sentiment och aktiekursen utvidgades experimentet till att inte bara undersöka korrelation för efterföljande dag utan även undersöka korrelationen för flera tidsspänn omedelbart efter och innan att kommentarer skrivits. För varje kommentar som skrevs medan marknaden var öppen beräknades volatilitet och avkastning för aktien från publiceringspunkten och två minuter, tio minuter, en timme, respektive ett dygn framåt och bakåt i tiden, samt en vecka framåt i tiden.

Samtidigt är det intressant att också undersöka om det uppträder en korrelation för kommentarer som skrivits efter en marknadsrörelse. En sådan korrelation skulle kunna indikera att det existerar något samband även för kommentarer som skrivits innan.

I likhet med Wolf och Bergdorf studie på Twittermeddelanden [64] som inte tog hänsyn till dess gillamarkeringar eller delningar valdes att inte ta hänsyn till Redditkommentarernas poäng. Detta eftersom poängen varierar efter att kommentaren skrivits och således inte kan användas för att väga dess relevans mot andra kommentarer när den publiceras. Vidare så filtrerades borttagna kommentarer och inlägg bort. Endast ett fåtal förekom och de var lätta att hitta då Reddit automatiskt ändrar innehållet till *[Removed]*.

För att beräkna fram korrelationskoefficienterna så användes Pythonpaketet *NumPy* som underlättar utvärdering och automatiserar beräkningen. Paketet innehåller färdiga funktioner som tillåter beräkning av bland annat Pearsons korrelationskoefficient för datamängder. Vidare användes även Pythonpaketet *SciPy* för att beräkna fram p-värdet. Med hjälp av inbyggda funktioner kan t-tester utföras och ett p-värde erhållas automatiskt.

5.2 Artiklar från Yahoo News

Som nämnades i avsnitt 3.2.2 uppvisar Yahoo Finance en mångfald bland sina källor. Det var således mer effektivt att fokusera på Yahoo Finance jämfört med att skrapa åtskilliga nyhetshemsidor. Nedan följer tillvägagångssättet som lett till resultaten som presenteras i kapitel 6.2.

5.2.1 Insamling av nyhetsartiklar

Ett första steg i webbskrapningsprocessen var att undersöka hemsidans struktur. Lyckligtvis användes specifika HTML-taggar för artiklarnas uppbyggnad, vilket skulle underlätta skrapningen. Det visade sig även att Yahoo Finance använde sig av asynkron laddning av artiklar, det vill säga att hemsidans innehåll ändrades och

fylldes på i och med att användaren scrollade längre ner på sidan. Det krävdes således en scrollsimulering, vilket Pythonpaketet Selenium kan bidra med. Artiklarna visades då i en tabelliknande form tillsammans med källa, textbaserad tidsstämpel (ex. "12 hours ago"), rubrik, och artikelabstrakt.

När hemsidan laddades vid det första tillfället sparades all HTML-kod som visades. Därefter scrollade programmet långsamt hemsidan och sparade den uppdaterade koden, som innehöll all den nya informationen som hemsidan läst in. Förfarandet upprepades tills de senast laddade artiklarna saknade tidsmarkörerna "hours" eller "minutes", eftersom de därav publicerats för mer än ett dygn sedan. Den första versionen av programmet avslutade skrapningen i detta skede och sparade artikelinformationen i en kommaseparerad fil. Denna version av programmet startades manuellt vid ungefär samma tid varje morgon.

Efter en inledande analys av nyheterna visade det sig att det var svårt att beräkna det exakta publiceringsdatumet utifrån den sparade, textbaserade tidsstämpeln. Eftersom programmet inte startades vid *exakt* samma klockslag varje dag och hade varierande körtid kunde den sparade textbaserade tidsstämpeln, som endast existerade i programmets temporära minne, skilja sig från tidsstämpeln då filen sparades. Exempelvis kunde en artikel ha publicerats för fyra minuter sedan, men om det tog programmet sex minuter att köra och spara datan till en fil, skulle det innebära att artikeln egentligen publicerades tio minuter innan filens spartid. För att åtgärda detta problem skrevs programmet om med funktionen att spara den exakta tiden och fullständiga artikeln, genom att automatiskt klicka sig in på varje artikel. Följaktligen automatiserades körningen så programmet startades varje morgon kl 06:00.

5.2.2 Sortering och analys av nyheter

För att kunna tillämpa den statistiska analysen behövde artiklarna sorteras. Alla artiklar vars publiceringstid inte överensstämde med börsens öppettider sorterades bort, vilket gjordes i syfte att kunna observera prisreaktionen direkt efter publicering. Artiklar som publicerades under stängningstid har troligtvis även de en påverkan på priset, men eftersom rapporten ämnar att undersöka hur sentimentet i artiklar påverkar aktiepriset vore det omöjligt att utröna *vilken* artikel i en mängd artiklar som bidragit till specifik prISRÖRELSE. Att endast se en prisskillnad efter ett antal artikelpubliceringar var irrelevant då rapporten ämnar undersöka hur artiklar och dess sentiment påverkar aktien, inte *om* de gör det. I detta skede var det även viktigt att ta sommartid i beaktning, då USA och Sverige gör skiftet vid olika tidpunkter på året. Artiklarnas tidsstämpel modifierades därefter.

Det visade sig tidigt att det var stor skillnad i antalet artiklar per aktie (se tabell 5.1). Detta skulle medföra att en nyhetsanalys per aktie skulle ge upphov till lågt statistisk värde, då det var skevt antal sampel. Det visade sig alltså redan i det här skedet att inkorporera nyhetsinsamling och analys av dessa artiklar, på en per-aktie basis, inte skulle vara givande för alla aktier i DJIA.

Tabell 5.1: Aktier med flest och minst artiklar publicerade under projektets tidsperiod samt antal av dessa som publicerats under börsens öppettider.

Aktie	Antal artiklar	Under öppningstid
AAPL	565	96
MSFT	318	71
MMM	12	3
TRV	4	2

Sentimentanalys med VADER applicerades på de artiklar som kvarstod efter sorteringen och det sammanlagda sentimentvärdet per artikel sparades ned. Därefter beräknades prisskillnaden i specifika intervall, det vill säga innan och efter publiceringstid, för aktie relaterad till artikeln.

Korrelationsundersökning tillämpades slutligen på det beräknade artikelsentimentet och prisskillnader i tidsintervallen i tabell 5.2. För analysen användes NumPy.

Tabell 5.2: Tidsintervall för korrelation av artikelsentiment och aktiepris.

Undersökt tid relativ publiceringstid
-2 tim
-1 tim
-10 min
-1 min
+1 min
+10 min
+1 tim

6

Resultat

Detta kapitel består av en sammanställning av resultaten som uppnåddes för investeringsalgoritmerna samt korrelationsanalysen som utfördes i samband med webbskrapningen.

6.1 Investeringsalgoritmernas avkastning

Nedanför presenteras de slutgiltiga resultaten för intradags- och interdagsalgoritmen. Algoritmernas avkastning på vid simuleringar på samtliga dataset sammanställs i tabeller, medan algoritmernas avkastning på testsetet även presenteras grafiskt. Endast intradagsalgoritmen programmerades för fiktiv realtidshandel, dess resultat presenteras i tabell 6.1.

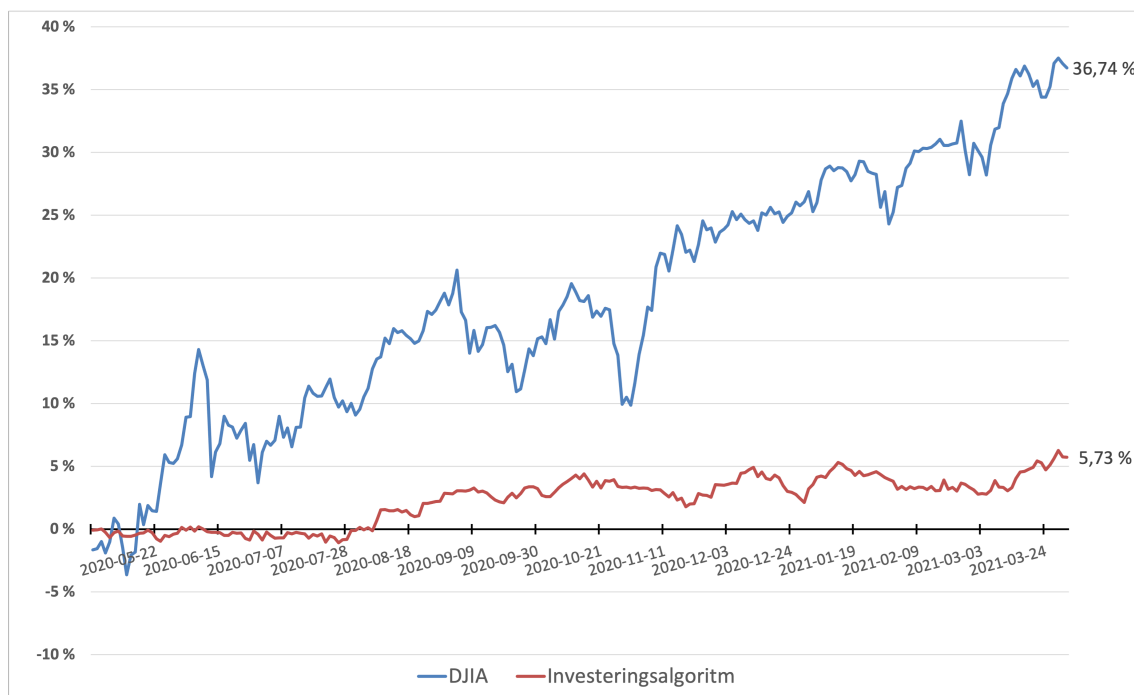
6.1.1 Intradagshandel

Avkastningen för investeringsalgoritmen som utvecklades för intradagshandel presenteras i tabell 6.1. Algoritmen lyckades inte generera högre avkastning än DJIA under varken träningssetet eller testsetet. Däremot gav investeringsalgoritmen en liten positiv avkastning under coronasetet medan aktiemarknaden föll kraftigt. Investeringsalgoritmen gav även vid realtidshandel högre avkastning än DJIA.

Tabell 6.1: Avkastning för investeringsalgoritmen (med intradagshandel) i jämförelse med DJIA under de olika perioderna.

Set	Period	DJIA	Investeringsalgoritmen
Träningsset	2018-01-02 - 2020-01-31	13,89 %	9,80 %
Coronaset	2020-02-01 - 2020-04-30	-14,03 %	0,76%
Testset	2020-05-01 - 2021-03-31	36,74 %	5,73 %
Realtidshandel	2020-04-27 - 2020-05-12	-1,17 %	0,53 %

I figur 6.1 åskådliggörs att investeringsalgoritmen inte lyckades följa samma positiva värdeutveckling som DJIA under testsetet. För att se grafer med investeringsalgoritmens avkastning jämfört med DJIA över alla perioder, se bilaga B.



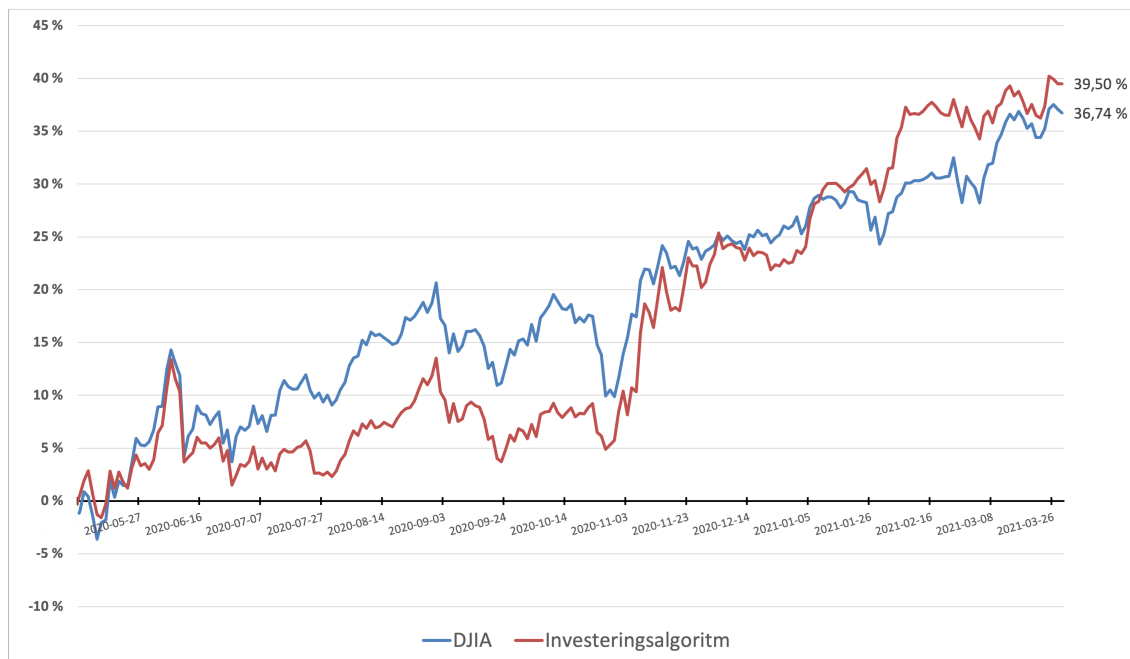
Figur 6.1: Intradagsalgoritmen vs. DJIA - Testset.

6.1.2 Interdagshandel

Investeringsalgoritmen anpassad för interdagshandel genererade avkastning visas i tabell 6.2. Algoritmen lyckades generera högre avkastning än DJIA under samtliga perioder som undersöktes. Även under testsetet presterade alltså algoritmen bättre än DJIA som kan ses i figur 6.2. För att se grafer med investeringsalgoritmens avkastning jämfört med DJIA över alla perioder, se bilaga B.

Tabell 6.2: Historisk simulerad avkastning från investeringsalgoritmen (med interdagshandel) i jämförelse med DJIA under de olika perioderna.

Set	Period	DJIA	Investeringsalgoritmen
Träningsset	2018-01-02 - 2020-01-31	13,89 %	28,55 %
Coronaset	2020-02-01 - 2020-04-30	-14,03 %	-8,52 %
Testset	2020-05-01 - 2021-03-31	36,74 %	39,50 %



Figur 6.2: Interdagsalgoritmen vs. DJIA - Testset.

6.2 Statistisk signifikans av insamlad media

Ned till presenteras resultaten av korrelationsanalysen utförd på den insamlade datan från Reddit samt Yahoo Finance inom tidsperioden 2021-03-11 till 2021-04-26.

6.2.1 Redditkommentarer

I tabell 6.3 presenteras ett urval av de aktier som visade högst korrelation mellan sentiment i redditkommentarer samt omnämningar och volatilitet dagen efter publiceringstid, dessutom presenteras korrelationen för DJIA. Tabellen visar även antalet datapunkter som användes, och p-värdet för korrelationen.

Fortsättningsvis i tabell 6.5 så uppvisas korrelationen mellan sentiment och volatilitet för en aktie i DJIA efter omnämning. Detta över ett brett tidsintervall, från dagen innan till dagen efter publicering, såväl så några minuter innan och efter. Exempelvis visas förändringen i volatilitet tio minuter efter en publicerad Redditkommentar omnämner en aktie i DJIA. Slutligen i tabell 6.6 presenteras korrelationen mellan sentiment och avkastning efter omnämning i Redditkommentar, på samma tidsintervall som tidigare tabell.

Tabell 6.3: Korrelation (ρ) för DJIA samt utplockade aktier med högst korrelation.

Nyckelord	X	Y	ρ	p -värde	Antal
DJIA	Volatilitet	Sentiment	0,015	0,0	24608
AMGN	Volatilitet	Sentiment	0,223	0,0135	8
CSCO	Volatilitet	Sentiment	0,174	0,0064	56
AXP	Volatilitet	Sentiment	-0,194	0,5269	32
UNH	Volatilitet	Sentiment	-0.246	<0,00001	61

Tabell 6.4: Korrelation (ρ) för DJIA samt utplockade aktier med högst korrelation.

Nyckelord	X	Y	ρ	p -värde	Antal
DJIA	Volatilitet	Omnämningar	-0,073	<0,00001	24608
NKE	Volatilitet	Omnämningar	0,505	<0,00001	143
CAT	Volatilitet	Omnämningar	0,477	<0,00001	582
BA	Volatilitet	Omnämningar	0,465	<0,00001	1300
INTC	Volatilitet	Omnämningar	0,429	<0,00001	1140

Tabell 6.5: Korrelationskoefficienter för kommentarer på Reddits sentiment och tillhörande aktiekurs volatilitet, samt statistisk signifikans av korrelationen. Ungefär 4000-7000 kommentarer nämnde en aktiekurs i Dow Jones och skrevs under tider som möjliggjorde beräkning av volatilitet för tidsperioden.

Relativ tid	ρ för sentiment och pris	p	ρ för absolutvärde av sentiment och pris	p
-24 tim	-0,025	0,073056	-0,016	0,232129
-1 tim	-0,045	0,000265	-0,039	0,001575
-10 min	-0,054	< 0,00001	-0,030	0,012027
-2 min	-0,025	0,037136	-0,022	0,079997
+2 min	-0,045	0,000245	-0,006	0,677007
+10 min	-0,037	0,002661	-0,014	0,278085
+1 tim	-0,035	0,004548	-0,009	0,452724
+24 tim	-0,016	0,181232	-0,007	0,558618
+1 v	-0,011	0,357405	-0,02	0,023864

Tabell 6.6: Korrelationskoefficienter för kommentarer på Reddits sentiment och tillhörande aktiekurs avkastning, samt statistisk signifikans av korrelationen. Ungefär 4000-7000 kommentarer nämnde en aktiekurs i Dow Jones och skrevs under tider som möjliggjorde beräkning av avkastning för tidsperioden.

Avkastning	ρ för sentiment och pris	p	ρ för absolutvärde av sentiment och pris	p
-24 tim	-0,009	0,521519	0,044	0,001802
-1 tim	0,002	0,903789	0,058	0,00001
-10 min	-0,003	0,868856	0,054	< 0,00001
-2 min	0,003	0,802308	0,030	0,011162
+2 min	0,005	0,658382	0,017	0,14659
+10 min	-0,016	0,215362	0,013	0,286656
+1 tim	0,003	0,841238	0,037	0,003947
+24 tim	-0,021	0,157611	-0,006	0,723897
+1 v	-0,012	0,456233	-0,001	0,946

6.2.2 Yahoo News-artiklar

Sammanlagt analyserades 278 artiklar som publicerats under amerikanska börsens öppettider. Tabell 6.7 visar de största prishöjningarna och prissänkningarna före och efter publicering av artiklar. Som synes varierar sentimentet (kolumn S) mellan starkt negativt och starkt positivt, oavsett prisupp- eller prisnedgång. Både artiklar med största prissänkningen och största prishöjningen hade sentiment som enligt VADER ansågs vara i stort sett fullständigt positiva. Detta är återkommande i den kompletta datamängden. Prisskillnaden Δ är från en timme innan publicering till en timme efter.

Tabell 6.7: Artiklar publicerade vid tidpunkten t innan de största prisskillnaderna.

Aktie	Artikel	Tid $t=0$	S	Pris $t-1$ tim	Pris $t=0$	Pris $t+1$ tim	Δ
BA	Brazil's Covid-19 Crisis Could Pose Trouble for a Global Travel Recovery. Here's Why. [65]	2021-03-11 18:58	-0,65	254,09	251,16	250,05	-4,04
BA	Boeing (BA) Expects Big 737 MAX Orders From Southwest Airlines [66]	2021-03-11 17:57	1,00	254,89	254,21	250,95	-3,94
BA	S&P 500 Snaps 2-Day Losing Streak as Tech U-Turn Triggers Rally [67]	2021-03-25 17:30	0,96	236,39	244,18	244,59	8,20
BA	Southwest Airlines to Reintroduce the Boeing 737 MAX [68]	2021-03-10 17:58	0,00	235,46	241,14	241,36	5,90

Tabell 6.8 visar artiklar med de högsta och lägsta sentimenten som publicerats under projektets gång. Dessa leder till både prisupp- och nedgång.

Tabell 6.8: Artiklar publicerade vid tidpunkten t med de största sentimenten.

Aktie	Artikel	Tid $t=0$	S	Pris $t-1$ tim	Pris $t=0$	Pris $t+1$ tim	Δ
MRK	Merck & Co's Debt Overview [69]	2021-03-16 17:45	-0,98	76,59	76,37	76,59	0
BA	Boeing's Debt Overview [70]	2021-03-22 17:22	-0,96	252,75	251,99	252,81	0,06
AAPL	Influencers with Andy Serwer: Ruth Porat [71]	2021-03-18 18:26	1,00	122,33	121,87	121,60	-0,73
AAPL	iPhone vs Android: How to decide which smartphone is best for you [72]	2021-03-16 18:02	1,00	126,21	125,96	125,66	-0,55

Tabell 6.9 visar beräknade korrelationskoefficienter för artikelsentiment och aktiepris runtom publiceringstidpunkten. Som synes är alla negativa vilket tyder på negativt samband, det vill säga att positivt sentiment hade större samband med sänkning av aktiepris än negativt sentiment. Koefficienterna är även väldigt låga vilket tyder på ett mycket svagt samband. Enligt tillhörande signifikansvärde är endast korrelationen två timmar innan och tio minuter innan att lita på, då dessa är under gränsvärdet 0,05.

Tabell 6.9: Korrelationskoefficienter för sambandet mellan artikelsentiment och aktiepris under specifika intervall runtom publiceringstid.

Pearson ρ	Signifikans p	Tid relativ publicering
-0,12	0,047453	-2 tim
-0,06	0,318866	-1 tim
-0,21	0,000452	-10 min
-0,05	0,435069	-1 min
-0,02	0,690325	+1 min
-0,09	0,143335	+10 min
-0,11	0,069583	+1 tim

7

Diskussion

Nedan följer en diskussion av de erhållna resultaten för investeringsalgoritmerna och webbskrapningen. Vidare diskuteras möjliga förbättringsvägar för de båda delarna, samt förslag till framtida arbeten som detta arbete har lagt grund för. Därefter beaktas de etiska aspekter som kan kopplas till detta arbete, och avslutningsvis formuleras en slutsats för att sammanfatta projektet.

7.1 Investeringsalgoritmerna

För att återkoppla till projektets syfte, vilket delvis var att undersöka huruvida det går att utforma en investeringsalgoritm som genererar högre avkastning än DJIA, talar resultaten som presenteras i underavsnitt 6.1.2 för att detta är möjligt. Interdagsalgoritmens slutgiltiga avkastning nådde 39,5 %, jämfört med DJIA:s 36,74 %.

Som visas i underavsnitt 6.1.1 genererade inte intradagsalgoritmen några utstickande resultat på den historiska simuleringen. Däremot uppnåddes högre avkastning än DJIA vid den fiktiva realtidshandeln, men då detta var under en väldigt kort period kan inga slutsatser dras utifrån detta resultat.

Trots att interdagsalgoritmen gav bättre avkastning än intradagsalgoritmen var intradagshandeln projektets huvudfokus. På grund av detta kommer diskussionen som följer därför fokusera på hur högre handelsfrekvens ställer högre krav på investeringsalgoritmerna. Diskussionen kring hur avkastningen för de båda investeringsalgoritmerna kan förbättras beskrivs vidare i underavsnitt 7.1.5.

7.1.1 Resultatens validitet

Sett till projektets syfte och omfattning är resultatet som visas i figur 6.2 ett välkommet resultat, men aktsamhet bör iaktas innan generella slutsatser dras utifrån detta. Att en investeringsalgoritm under en simulering på elva månader lyckas generera högre avkastning än DJIA är inte underlag nog för att döma algoritmen mer lönsam än indexet för kontinuerligt framtida bruk.

En aspekt som troligtvis förstärkte interdagsalgoritmens avkastning är att testsetet präglades av en väldigt starkt uppåtgående trend, se figur 6.2. Som beskrivs i

underavsnitt 2.2.1 är filterregeln, den handelsstrategin som användes vid interdags-handel, en trendföljande strategi som bygger på att köpa en aktie med uppåtgående trend i hopp om att den ska fortsätta sin trend. Filterregeln hade således ultimata förhållanden för att generera hög avkastning under testsetet.

Filterregeln har dokumenterats sedan 1960-talet, vilket innebär att den är välkänd. Vore det så att filterregeln sedan 1960-talet konsekvent genererat högre avkastning än en investering i DJIA så borde den, även med hänsyn till sin simplicitet, användas av en stor andel aktörer på aktiemarknaden. I ett fall där flera investerare använder samma handelsstrategi minskar vinstmarginalerna som varje investerare kan förväntas erhålla, då de tvingas dela på vinsten från varje lyckat köp. Enligt detta resonemang bör filterregelns vinstmarginaler vara begränsade.

7.1.2 Konkurrenten om vinstmarginalerna

Vid tester på sekund- och minutdata lyckades inte filterregeln gå med vinst, men på dagsdata lyckades den uppnå högre avkastning än DJIA, se figur 6.2. Avkastningen blev alltså högre då algoritmen handlade på lägre frekvens. Det är emellertid komplicerat att identifiera exakt vad detta beror på. Nedanför följer ett resonemang kring en potentiell orsak till dessa skilda resultat.

Skillnaden mellan dessa resultat kan dels bero på konkurrenten om vinstmarginalerna. Vid handel på högre frekvenser hålls positioner i aktier under korta perioder, vilket medför att aktiepriset inte hinner fluktuera i samma utsträckning som vid interdagshandel [73]. Således kan vinstmarginalerna antas vara mindre vid handel på högre frekvenser.

Om vinstmarginalerna är lägre innebär det att ifall flera aktörer identifierar samma köpläge med liknande handelsstrategier, så bör den aktör som snabbast genomför köpet erhålla högst vinst. Efterföljande aktörer kommer handla aktien till ett högre pris vilket minskar vinstmarginalen ytterligare. För den långsammaste aktören medför detta att priset kan ha ökat påtagligt från det att köpläget identifierades tills dess att ett köp genomförs. En korrekt identifierad köpsignal kan således resultera i en förlustaffär.

Interdagsalgoritmen i detta projekt kan hålla positioner i aktier under en längre tidshorisont än intradagsalgoritmen, vilket i enlighet med föregående resonemang medför att vinstmarginalerna är högre. Om en aktie under exempelvis en veckas tid följer en positiv trend blir inte den exakta tiden och priset för köpet lika vitalt för att det ska generera en vinst. Vid en trendande marknad, så som i testsetet, så blir resultatet därför inte lika känsligt för konkurrenten om vinstmarginalerna.

7.1.3 Kursutveckling på kort och lång sikt

Som nämndes i underavsnittet ovan var resultaten från simuleringar av filterregeln på dagupplöst data betydligt bättre än de för sekund- och minutupplöst data. Utöver konkurrenten om vinstmarginalerna kan en bakomliggande faktor till simuleringensre-

sultaten vara skillnaden i de faktorer som påverkar ett aktiepris rörelse på kort och lång sikt.

I avsnitt 2.1 beskrivs att den allmänna oro som spreds i samband med coronapandemins utbrott fick aktiemarknaden att falla kraftigt. Efter den kraftiga nedgången började en marknadsoptimism träda fram och resulterade senare i att aktiemarknaden återhämtade sig. Resonemanget som kan utvinnas från detta är att en akties rörelser på längre sikt styrs av allmänhetens inställning till aktiemarknaden. Om allmänheten är optimistisk har aktien en positiv värdeutveckling och vice versa.

Som nämns i 2.2.2 så behöver en akties kortsiktiga rörelser inte bero på hur företaget går. Vid högre frekvenser är aktiernas rörelser inte på samma sätt förankrade i marknadens inställning, då det snarare baseras på utbud och efterfrågan i orderboken (se bilaga A).

Resonemanget som följer utifrån detta är att trender på längre sikt, som beror på marknadens optimism, tenderar att hålla i sig längre än trender på kort sikt som beror på utbud och efterfrågan. Detta resonemang grundar sig i antagandet att utbud och efterfrågan är mer flyktigt än vad marknadens inställning är.

Slutsatsen från detta resonemang är att vid syfte av att utveckla en investering algoritm som bygger på trendföljande strategier bör en lägre handelsfrekvens användas. Detta då lägre handelsfrekvens ger upphov till stabilare trender vilket borde vara gynnsamt för en trendföljande handelsstrategi.

7.1.4 Rivalitet på aktiemarknaden

En avslutande aspekt som kan förklara varför intradagsalgoritmen inte uppnådde högre avkastning är rivaliteten på aktiemarknaden. I en dokumentär från 2020 gällande investeringsalgoritmer, och framför allt högfrekvent handel, talas det om hur fysiker och matematiker är efterfrågade på finansmarknaderna för att försöka förutspå aktiemarknaden [37].

En av forskarna som intervjuas i dokumentären, Doyne Farmer, förklarar hur det tog honom och ett team med yngre forskare och affärsmän fem år att utveckla en modell som fungerade stabilt. Det nämns även att den amerikanska investmentbanken Goldman Sachs vid millenieskiftet köpte ett algoritmiskt handelsföretag för en halv miljard dollar för att hålla sig konkurrenskraftiga.

De högfrekventa algoritmerna som diskuteras i dokumentären är för avancerade och storskaliga för att kunna jämföras med intradagsalgoritmen i detta projekt. Trots detta stävar de mot samma mål och konkurrerar på samma marknad.

Anledningen till att intradagsalgoritmen inte lyckades uppnå lika hög avkastning som interdagsalgoritmen kan således bero på att intradagsalgoritmen, som handlar på högre frekvens, kommer närmre att konkurrera med dessa insitutionella aktörer. Med tanke på detta projekts omfattning och den enkla logik som intradagsalgoritmen bygger på är det enligt detta resonemang inte förvånande att algoritmen inte

genererar mer avkastning än vad den gjorde.

7.1.5 Investeringsalgoritmernas förbättringsområden

En första aspekt som eventuellt kan höja avkastningen vid intradagshandel på minutupplöst data är att samla transaktionsvolymerna som sker under varje given minut. Då information om aktiepriset registreras vid en transaktion kan även antalet aktier som affären innehöll samlas. Den ackumulerade transaktionsvolymen för varje minut innehåller information som potentiellt kan utnyttjas för att göra bättre investeringar. Att ej utnyttja denna information innebär att konkurrerande aktörer har mer information att basera sina investeringsbeslut på, vilket borde ge konkurrenterna en fördel.

En annan aspekt som potentiellt kan öka investeringsalgoritmernas avkastning är möjligheten att blanka en aktie (se bilaga A). Att inte kunna blanka en aktie innebär praktiskt taget att algoritmen går miste om halva vinstmöjligheten. Eftersom att varje säljsignal som genereras av handelsstrategierna är en signal på att aktiepriset troligtvis kommer falla skulle algoritmen kunna ha blankat en aktie varje gång den säljs. På så vis har algoritmen alltid en position på marknaden. Det ska förtydligas att om algoritmen från början gick med förlust hade den mest troligt gått med ännu mer förlust om möjligheten att blanka fanns.

Investeringsalgoritmen för intradagshandel fann endast ett fåtal aktier med lågvolatila tillstånd per dag. Detta medförde att investeringsalgoritmen endast investerade i några enstaka aktie per dag. Detta kan vara en av anledningarna till varför investeringsalgoritmen inte var i närheten av att generera lika hög avkastning som DJIA i testsetet (se figur 6.1). Då investeringsalgoritmen inte innehöll positioner i särskilt många aktier under denna perioden gick algoritmen miste om en stor del av den positiva marknadsutvecklingen.

Med hänsyn till ovanstående resonemang borde investeringsalgoritmen generera högre avkastning då algoritmen letar efter lågvolatila tillstånd i fler aktier än de som ingår i DJIA. Ett resultat av detta blir således att investeringsalgoritmen utnyttjar marknadsutvecklingen mer. Noterbart är att fler aktier även innebär en större risk vid negativa marknadsutvecklingar.

7.1.6 Förbättringsmöjligheter i metodval

En faktor som kan ha haft stor påverkan på projektets resultat är metoden för att dela upp den historiska datan som algoritmerna simulerades på. Med hänsyn till risken för eventuella regimskift som beskrivs i avsnitt 2.1 användes inte data från tidigare än 2018. Vid tidpunkten då detta beslut togs var huvudfokuset i arbetet fortfarande intradagshandel, vilket innebär att beslutet om hur datan skulle delas upp baserades på att algoritmen skulle genomföra flera transaktioner per dag.

För att utvärdera interdagsalgoritmen, som bygger på filterregeln, borde mer data ha använts. Samma antal dagar för att utvärdera de båda algoritmerna resulterar

i enorm skillnad i antal datapunkter i de dataset som algoritmerna testas mot. Här skulle alltså en noggran avvägning mellan mängd historisk data och potentiella regimskift i datan behöva göras.

Sett till arbetsprocessen som helhet borde en tydligare prioritetsordning följts. Implementeringen av simuleringsmiljön pågick fortlöpande under arbetet parallellt med implementeringen av nya handelsstrategier. Fokus borde istället varit att färdigställa simuleringsmiljön i första hand, för att sedan implementera och testa strategierna. Allt eftersom att strategier testades upptäcktes nya felaktigheter i simuleringsmiljön.

En konsekvens av att denna miljö ej färdigställdes innan en del av handelsstrategierna testades är att tidiga, vägledande resultat kan ha varit felaktiga. Då en stor del av tiden under arbetet gick åt till att bygga upp simuleringsmiljön hann inte någon större mängd strategier undersökas. I detta projekt kunde därför tidiga resultat bekräftas i efterhand, vilket minskade konsekvenserna av bristerna i simuleringsmiljön.

I mer omfattande projekt kan liknande metodtekniska misstag vara mer kostsamma och eventuellt vilseledande. Om felaktiga resultat erhålls tidigt i ett projekt, och dessa ej kan korrigeras i efterhand, föreligger det en risk att projektet ej uppnår optimala slutresultat.

7.2 Webbskrapning

Detta avsnitt behandlar utförligare diskussion och analys av resultaten benämnda i kapitel 6. Vidare så tas begränsningar och förbättringar för webbskrapningen och korrelationsanalysen upp.

I kapitel 2.5 definieras korrelation som ett värde mellan -1 och $+1$, där 0 indikerar ingen korrelation medan -1 och $+1$ en perfekt negativ respektive positiv korrelation. Vidare definieras p-värdet inom intervallet 0 till $+1$, där lägre värden uppvisar en starkare validitet av korrelationen. Signifikansnivån i projektet är satt till $0,05$ vilket betyder att p-värden större än $0,05$ inte visar på statistisk signifikans.

7.2.1 Sentiment i Redditkommentarer

Då projektets syfte med webbskrapning var att tillsammans med investeringsalgoritmerna generera en högre avkastning visar resultaten i kapitel 6.2 att syftet ej uppnås för webbskrapning av Reddit.

Resultaten i tabell 6.3 visar på en låg korrelation mellan sentimentet i Redditkommentarer och volatilitet dagen efter publicering för de utplockade aktierna. Vidare är p-värdena större än $0,05$ vilket försvagar validiteten i korrelationen och leder till att resultaten inte når statistisk signifikans.

Vidare uppvisar resultaten i tabell 6.4 en medel korrelation mellan antalet omnämningar per dag och volatilitet dagen efter publicering bland de utvalda aktierna. Dessutom förekommer låga p-värden som stärker validiteten i resultaten och medför

statistisk signifikans. Men endast ett fåtal aktier i DJIA uppvisade låg till medel korrelation, majoriteten visade en försumbar korrelation.

När det gäller korrelation mellan Reddittkommentarers sentiment och volatilitet indikerar resultaten i tabell 6.5 att det kan finnas en viss korrelation en timme före och efter att kommentaren skrevs. Likaså pekar resultaten i tabell 6.6 på att det kan finnas en viss korrelation mellan sentiment och avkastning föregående dygnet och en timme efter att kommentaren skrevs.

Sammanfattningsvis anses det inte möjligt utifrån de kommentarer som samlats in att förutse framtida avkastning hos aktier i DJIA i syfte att öka avkastningen hos investeringsalgoritmerna. Även om delar av resultaten uppvisar en korrelation så är det en överväldigande majoritet som endast påvisar en försumbar korrelation.

I undersökningen existerade en tendens där större antal omnämningar ledde till lägre korrelation, förmodligen då det leder till betydligt mer brus. Vilket ytterligare ifrågasätter resultatens validitet, även de med statistisk signifikans. Ifall alla aktier hade ungefär samma mängd omnämningar är det möjligt att fler haft en negligerbar korrelation.

En anledning till varför färre antal omnämningar gav en högre korrelation kan bero på att de användare som diskuterar de mindre populära företagen är mer insatta, det vill säga att de själva gjort en mer konkret undersökning och hittat ett undervärderat företag. Det kan förklara varför aktier med endast ett fåtal omnämningar fick högre korrelation i tabell 6.3, då sentimentet i kommentarerna är mer informerat. Dock leder fåtalet omnämningar till låg statistisk relevans och är därför svårare att använda i projektets syfte.

Ytterligare en anledning till att majoriteten av korrelationkoefficienterna blev låga eller försumbara kan vara att kommentarer förmodligen är reaktiva, med andra ord skrivs de efter att något hänt, inte innan. Därför, när kommentarer reagerar på händelser kan det vara så att den underliggande aktien återgår mot medelvärdet dagen efter enligt teorin i 2.2.1 vilket motskrider uttryckt sentiment.

Korrelationen mellan absolutvärdet av sentimentet och avkastning innan/efter publicering av kommentar efterliknar känsla. Korrelationen mellan känsla och avkastning som föregår en kommentar är större än den som följer. Detta kan bero på att investerare som gått med vinst eller förlust uttrycker starka känslor i forum efteråt.

Det kan även visa sig att webbskrapningen också misslyckades på grund av att marknaden inte alltid följer det generella sentimentet på kort sikt men oftast under en längre period. Webbskrapning kanske lämpar sig mer för en längre investeringshorisont.

7.2.2 Sentiment i nyhetsartiklar

Gällande nyhetsartiklar har det inte påvisats någon stark korrelation mellan aktiepriser och sentimentet hos publicerade nyhetsartiklar under den undersökta tidspe-

rioden. Detta kan delvis ses i tabell 6.7 som visar att både relativt lågt sentiment och maximalt sentiment båda bidrog till de markanta prisminskningar som upptäcktes en timme efter publiceringstid. Det vill säga att priset sjönk oavsett artikelns attityd.

Likaså visar tabell 6.8 att nästintill minimalt sentiment bidrog till antingen ingen prisskillnad alls eller till och med en smärre ökning av priset. Två av de artiklar med maximalt positivt sentiment bidrog till prissänkning snarare än förväntad prishöjning.

Tabell 6.9 visar låga korrelationskoefficienter vilket tyder på litet samband, om ens något, mellan den publicerade artikelns sentiment och aktiepriset innan eller efter publiceringstid. Dock visar signifikansvärdet att den högst uppmätta korrelationen, tio minuter innan publicering, är av statistisk signifikans.

Resultatet indikerar inga konkreta bevis för att en attitydsanalys med VADER av de nyhetsartiklar som Yahoo Finance publicerat under börsens öppningstid skulle kunna prognostisera aktieprisets framtida rörelser. Värt att notera är dock tidsaspekten, eftersom datan enbart samlades in under en kort period.

En förklaring till varför idén om att förutse aktierörelser efter en nyhet eller kommentar misslyckades kan vara att de investerare som har störst inflytande på marknaden också är snabba på att reagera på bra och dåliga nyheter.

Att mätningen av aktieprisets korrelation var som störst tio minuter innan informationen publicerades är en intressant aspekt. Det kan tyda på att vissa investerare har fått tillgång till artikelinformationen strax innan publiceringstid, något som ger dem ett reellt övertag gentemot andra intressenter. Att vissa investerare på förhand ges åtkomst till viktig information står i stark kontrast till hypotesen gällande effektiva marknader [11], där all information som återges i aktiepriset ska offentliggöras för samtliga investerare vid samma tidpunkt. Ovanstående fynd kan dock indikera att spelplanen är skev redan från början.

Att ingen större korrelation mellan nyheter och pris har hittats kan anses vara förvånande. Detta kan bero på att inga oförutsägbara globala nyheter har publicerats under projektets gång. Som [74] nämner köper och säljer större firmor och investerare i förebyggande syfte på nyheter, snarare än som reaktion. Via deras företagsbesiktning har de möjligtvis redan förutspått resultatet i en årsredovisning. Förespråkare av ”köp och håll”-strategin är även de oberoende av tillfälliga prisfluktuationer då de investerar med avsikten att hålla aktien under lång tid, oavsett överraskande globala nyheter och händelser [75].

Möjligtvis kan en stor mängd publiceringar utmatta investeraren om denne reagerar på nyheter, vilket kan leda till beslutsutmattning [76]. Personerna kan därför efter ett tag sluta fatta beslut eller rentav fatta mer emotionellt reaktiva beslut. En aktie som istället har väldigt få nyheter gör att investeraren behöver fatta mindre beslut samt får utrymme att överväga informationen under en längre period, vilket därigenom ökar chanserna att fatta bättre och mer välgrundade beslut. Med detta

sagt är det möjligt att aktier med få relaterade nyheter har högre korrelation till aktiepriset, trots att dess låga antal bidrar till låg statistisk relevans och pålitlighet.

7.2.3 Begränsningar och möjliga förbättringar

Tidigare arbeten har använt sig av en betydligt **större mängd data** för korrelationsanalys än vad som samlades in under kandidatarbetets tidsperiod. Exempelvis refererar Gui [23] till en databas som sparar alla kommentarer som publicerats på hela Reddit under ett flertal år. Exakt hur många han har använt sig av i sin tes framgår inte i texten. Under insamlingsperioden samlades cirka tre miljoner Reddittkommentarer in i databasen. Det kan anses mycket, men detta är ofiltrerad data. Eftersom detta projekt fokuserar på aktier inom DJIA har endast kommentarer som nämnt dessa 30 aktier använts, vilket sorterar bort och begränsar antalet kommentarer avsevärt.

VADER är skrivet i syfte att analysera sociala medier, vilket nyhetsartiklar inte klassificeras som. En möjlig förbättring kan därmed vara, vid användning av VADER, att uteslutande fokusera på artiklar som närmare liknar textuformningen på social media, exempelvis opinionsartiklar. En mer gynnsam lösning kan vara att använda sig av ett annat analysverktyg än VADER, alternativt konstruera en någon egen klassificeringsmetod med önskat fokus.

Som det det går att se i tabell 2.1 är ett annat problem med VADER att **ord med dubbelmeningar kan misstolkas**. Då Reddit och särskilt underforum såsom wallstreetbets är fyllt med slang och informella texter är felberäkning av sentiment mycket möjlig. Meningar som bör ha varit positiva kan ha betraktas som negativa av VADER. En möjlig förbättring med möjligtvis markant påverkan hade därför varit att definiera en egen ordlista med vanligt förekommande slang och termer.

Artiklar och kommentarer publicerade under börsens stängningstid medförde även det problem. Eventuella prisfluktuationer vid nästa öppning kan bero på summan av tidigare publiceringar, eller på grund av en enstaka och separat orsak. Vilket av dessa är svårt att veta. Likaså är det svårt att över huvud taget veta orsaken till höjning alternativt sänkning av pris, trots eventuell positiv eller negativ publicering av artikel eller kommentar strax innan. Därmed blir det utmanande att forma en algoritm som bestämmer om en enskild artikel ska generera köp- eller säljrespons, baserat på artiklar publicerade under stängningstid. Om dessa artiklar ska användas i eventuell algoritm bör en kontinuerlig summering göras av dess sentiment under stängningstider.

Detta projekt har använt sig utav aggregeringssidan Yahoo Finance då det underlättade insamlingen av nyhetsartiklar. Eftersom **Yahoo är ett vinstdrivande och privatägt företag** finns det ingen garanti att de uppvisar objektivitet i valet av artikelpubliceringar. De kan exempelvis välja att inte publicera artiklar som kan generera förlust för företaget i sig eller dess ägare. De kan vidare själva använda sig av någon form av attitydanalys som bearbetar vilka artiklar som ska publiceras.

Ett problem med programmet som skrevs gällande insamling av artiklar var att det **endast fokuserade på de artiklar som publicerats de senaste 24 timmarna**. För att inte missa någon artikel behövdes programmet startas exakt samtidigt varje dag och ha en statisk körtid, så det inte uteslöt artiklar som exempelvis publicerades 24 timmar och 30 sekunder sedan. En lösning på problemet var ett skript som skrevs med syftet att starta programmet varje dag kl 6 på morgonen. Detta visade sig dock vara oberäkneligt på datorn där skriptet installerades, då programmet som imiterade webbläsaren inte stängdes av korrekt och bibehöll gammal data i sitt minne. Det finns således en risk för att artiklar har missats. En potentiell lösning skulle kunna vara att omkonstruera programmet till att hämta alla artiklar tills programmet stötte på artiklar som redan var sparade. På så vis blir publiceringstiden orelevant.

Reddit har en stor mängd botar, det vill säga datorprogram som agerar som en människa, och skriver egna kommentarer och inlägg som är förutbestämda av programmeraren. Dessa har olika ändamål, allt från att automatiskt länka till nämnd Wikipedia-artikel [77], till att automatiskt moderera en subreddit [78]. Botar kan även skrivas för att automatiskt förespråka till exempel en särskild produkt eller hemsida i Reddittrådar som är av relevans. Med detta sagt är det möjligt att botar kan ha skrivit en del av kommentarerna som samlats in, med uppsåt att påverka andra medlemmar. I och med att flera studier (se avsnitt 1.5) visat att det går att koppla sentiment i sociala medier till aktiekursen är det inte otroligt att vissa investerare använder sig av detta underlag för att fatta investeringsbeslut. Således finns det incitament för den som vill manipulera marknaden att systematiskt lägga ut missledande kommentarer i syfte att öka ett specifikt aktiepris för att sedan kunna sälja den till överpris. Detta skulle kunna ske genom att undersöka vilka typer av kommentarer ett sentimentanalyssystem fångar upp som ”positiva” för en aktiekurs, för att därefter skriva liknande kommentarer rörande den aktie som personen önskar öka köpvolum hos.

Den använda **modellen för att undersöka korrelationen kan också anses som något bristfällig**. En annan metod som skulle kunnat lämpa sig bättre för denna typ av studie är ”Granger causality test”. Då undersöks om en tidsserie (till exempel genomsnittligt sentiment över den senaste timmen) kan förutse en annan tidsserie, vilket i detta fall är aktiekurs. Denna typ av metod är vad Bergdorf & Wolf [64] använde i deras studie om korrelation mellan sentiment på Twitter och aktiekurser. Att sammankoppla varje kommentars sentiment till aktiekursen kanske inte lämpar sig lika bra som att istället beräkna ett genomsnittligt sentiment under en tidsperiod och försöka koppla det till aktiekursen.

Liksom på andra sociala plattformar finns det möjlighet att **köpa användare och uppröstningar**. En internetsökning på ”*buy reddit accounts*” genererar en mängd hemsidor som säljer konton med bra rykte. Genom att köpa ett stort antal konton är det möjligt att skapa och styra egna konversationer, vilket i sin tur kan påverka läsarens uppfattningen av diskussionsämne. På Reddit fästs stor betydelse vid uppröstningar; ju fler röster desto mer läsvärt anses något vara. Risken som detta kan medföra är att systemet vilsledds utifrån att olika aktörer betalar för uppröstningar. En ren lögn kan således fyllas på med oärliga uppröstningar och därigenom ge en

skev bild av verkligheten hos en oinsatt läsare. Exempelvis kan en kommentar som manar till att sälja en specifik aktie styrkas med stort antal köpta uppröstningar, med mål att originalförfattaren då kan köpa aktien till lägre pris vid ett senare skede.

Avslutningsvis hade arbetet kunnat utformas på liknande sätt som Bergdorf & Wolf [64], det vill säga **fokusera på de 20-30 största företagen** istället för ett specifikt index, då företag som Tesla (TSLA) inte är inkluderat i DJIA men som av TheStreet blev utsedd till årets aktie år 2020 [79]. I den insamlade datan omnämns alla aktier i DJIA sammanlagt 24 608 gånger medan aktier såsom TSLA och GameStop (GME) omnämns 15 957 gånger respektive 145 270 gånger. Förmodligen finns det fler företag som inte tillhör DJIA som är mer populära i nyheter och sociala medier.

7.3 Framtida arbete

Som nämndes i 1.5 har tidigare studier visat på relativt stark korrelation mellan internetsentiment och aktiepriser. Detta har inte lyckats genomföras i detta projekt, och den största skillnaden mellan utförande är mängden data som har använts för analysen. För att kunna utveckla en investeringsalgoritm som interagerar med marknadsanalys via webbskrapning kan det behövas en förstudie som undersöker om det finns någon uppskattad gräns för hur mycket data som behövs för att dra liknande slutsatser som tidigare sentimentstudier.

Framtida arbete skulle kunna syfta till att fortsätta utformningen av intradagsalgoritmen, då den inte lyckades prestera bättre än DJIA. Om medieanalys skulle generera givande resultat skulle det även vara intressant att undersöka om nuvarande intradagsalgoritm tillsammans med sentimentanalys skulle prestera bättre än index, eller om fortfarande samma resultat skulle uppnås.

Som diskuterades i underavsnitt 7.1.1 är resultatet från projektets interdagsalgoritm knappast generaliserbart. Däremot talar resultatet för att filterregeln har potential att generera hög avkastning i en uppåttrendande marknad. En intressant vidareutveckling av detta arbete vore således att kombinera filterregeln, eller liknande trendstrategier, med sentimentanalys som fokuserar på hela marknadens optimism snarare än individuella aktier. Detta skulle även kunna sammankopplas med eventuell maskininlärning som analyserar och fattar beslut utifrån historisk data, både gällande självständiga aktiepriser men även historiska sentiment.

7.4 Etiska aspekter

Algoritmisk aktiehandel och webbskrapning för med sig fördelar hos användaren. Dock existerar även en mängd negativa, etiska, aspekter rörande området, vilka några av dem nämns nedan. Till en början behandlar detta avsnitt ett antal negativa aspekter inom algoritmisk handel, där bland annat den egentliga innebörden av att äga en aktie diskuteras. Avslutningsvis ifrågasätts webbskrapning ur ett etiskt

perspektiv, däribland användandet av andras kommentarer för att påverka aktiekurser. Notera att ämnet aktiehandel som helhet inte avhandlas i nedan diskussion då arbetet endast rör algoritmisk handel.

7.4.1 Algoritmisk handels inverkan på samhället

Den verkliga meningen med en aktie berör å ena sidan företagsfinansiering, å andra sidan att erbjuda ett delägarskap i företaget, där även rösträtt ofta ingår på bolagsstämmor [24]. En etisk aspekt som bör diskuteras är därför att algoritmisk handel, beroende på utförande, skulle kunna anses förstöra syftet med aktier. Att automatiskt köpa och sälja olika aktier flera gånger inom en och samma dag utan att förstå vad det är för bolag det handlar om skulle kunna anses vara oetiskt eftersom det trots allt handlar om delägarskap i verkliga företag. Det finns exempel då algoritmisk handel orsakat snabba börskrascher [37]. Med det menas att börserna snabbt faller mycket mer än vanligt, men kan ofta snabbt återhämta sig. Även om börserna ofta återhämtar sig kan det bli en väldigt volatil börsmarknad som kan påverka samhället och dess ekonomi.

Att leva på algoritmisk handel och ha det som sitt yrke bidrar inte till samhället på samma sätt som många andra yrken gör. Det går endast ut på att exploatera andra personer som säljer aktier vid fel tidpunkt för att tjäna pengar på dem. Eftersom det krävs en viss kompetens inom matematik för att arbeta med algoritmisk handel lockas en del matematiker och fysiker [37], som med sin kompetens skulle ha kunnat bidra till samhället på andra sätt om de hade valt en annan yrkesinriktning. Det rör sig därmed om en miljardindustri som eventuellt inte medför någon samhällsnytta.

En annan aspekt är frågan om att stödja oetiska bolag genom att köpa aktier i dessa. Eftersom algoritmisk handel automatiskt köper och säljer utan att ta hänsyn till bolagen kan detta bli ett problem. Dock finns alltid lösningen att från början programmera en investeringsalgoritm som endast köper några utvalda aktier. Den metoden underlättar att utesluta bolag som till exempel tillverkar vapen och droger eller bolag som har en negativ inverkan på miljö och klimat.

Ett sista exempel på en etisk aspekt gällande algoritmisk handel avser hur förhållandena egentligen är på börsmarknaden. Det krävs en stor mängd datorkraft och en snabb uppkoppling för att algoritmen ska kunna erhålla informationen och ta beslut så snabbt som möjligt [37]. Därför väljer högpresterande investerare som satsar på algoritmisk handel att placera sina datorer i så nära anslutning till börsdatorerna som möjligt. All denna datorkraft som krävs för att lyckas bäst inom algoritmisk handel gör att de största inom branschen, med mest resurser, nästan alltid kommer ligga i framkant och ha konkurrensfördelar mot mindre aktörer. Detta gör att småsparare eller mindre aktörer som börjar med algoritmisk handel inte har en rättvis chans mot de större aktörerna. För småsparare kommer det nästan alltid vara någon annan som har tillgång till börsinformationen snabbare.

7.4.2 Webbskrapningens etik och samhällspåverkan

Det kan anses vara omoraliskt att använda sig av Redditanvändares kommentarer utan deras vetskap. Användarna är medvetna om att deras kommentarer och inlägg är öppna för alla som besöker hemsidan, men det tas eventuellt inte i beaktning av varje användare innan de publicerar en kommentar. De kan exempelvis vara inne i en djup diskussion med sina vänner, utan en tanke på att det som skrivs kan användas till någon annans fördel.

En annan aspekt att ta i beaktning gällande skrapning är ägarna till hemsidorna och serverna. Högfrekvent skrapning kan förväxlas med DoS-attack (Denial of Service), vilket innebär att en stor mängd förfrågningar skickas till hemsidan vilket saktar ner, eller till och med kraschar, hemsidan. Beroende på hemsida kan detta leda till förluster hos företaget, eller ökade serverkostnader och löner för utvecklarna som kanske måste arbeta övertid.

Det är upp till webbskraparen att använda datan på ett etiskt försvarbart sätt och göra sitt yttersta för att inte överbelasta leverantörer, det vill säga serverar och hemsidor.

7.5 Slutsats

Projektets syfte var att undersöka huruvida det var möjligt att utveckla en investeringsalgoritm som tillsammans med marknadsanalys via webbskrapning kunde generera högre avkastning än indexet Dow Jones Industrial Average.

Resultatet från sambandsundersökningen mellan aktiepris och nyheter visade endast relativt låg statistisk signifikans. Ett resultat som stod ut var sambandet mellan aktiepris tio minuter innan en artikel publicerades. Redditundersökningen lyckades inte fastställa något konkret samband mellan kommentarer och aktiepris.

Investeringsalgoritmen lyckades överträffa DJIA med specificerade handelsstrategier, specifikt filterregeln, på interdagsnivå, dock ej på intradagsnivå.

Då marknadsanalysen inte lyckades påvisa ett konkret samband mellan sentiment och aktiepriser integrerades inte marknadsanalysen med investeringsalgoritmen. Således krävs ytterligare arbete för att formulera en handelsstrategi utifrån sentimentanalys.

Litteratur

- [1] Yahoo Finance, "GameStop Corp. (GME)", 2021. [Online]. Tillgänglig: <https://finance.yahoo.com/quote/GME/history?period1=1609545600&period2=1612051200&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true> (hämtad: 2021-04-17).
- [2] E. Helmore, "How GameStop found itself at the center of a groundbreaking battle between Wall Street and small investors," *The Guardian*, [Online], Jan. 27, 2021. Tillgänglig: <https://www.theguardian.com/business/2021/jan/27/gamestop-stock-market-retail-wall-street> (hämtad: 2021-04-17).
- [3] M. Brignall, "How GameStop traders fired the first shots in millennials' war on Wall Street," *The Guardian*, [Online], Jan. 30, 2021. Tillgänglig: <https://www.theguardian.com/business/2021/jan/30/how-gamestop-traders-fired-the-first-shots-in-millennials-war-on-wall-street> (hämtad: 2021-04-17).
- [4] C. Longbing, "Data science: A comprehensive overview," *ACM Comput. Surv.*, årg. 50, nr 3, juni 2017. DOI: 10.1145/3076253.
- [5] P. Wissén och A. Wallgren, "De svenska finansmarknaderna," i *Marknad och politik*, L. Hultkrantz och P. Österholm, utg., Lund: Studentlitteratur, 2016, kap. 4, s. 129–166.
- [6] J. Young, "Market Index," *Investopedia*, [Online], April 29, 2021. Tillgänglig: <https://www.investopedia.com/terms/m/marketindex.asp> (hämtad: 2021-05-08).
- [7] X. Guo, T. Lai, H. Shek och S. Wong, *Quantitative Trading : Algorithms, Analytics, Data, Models, Optimization*. Boca Raton, Florida: CRC Press LLC, 2016.
- [8] P. Bergan och C. Devine, "Algorithmic Trading: What Should You Be Doing?" I *Investment Guides Series: Algorithmic Trading, Precision, Control, Execution*, London: Euromoney Trading Limited, 2005, s. 14–19.
- [9] E. P. Chan, *Quantitative Trading*. Hoboken, New Jersey: John Wiley & Sons, 2009.

-
- [10] S. R. Syamala och K. Wadhwa, "Trading performance and market efficiency: Evidence from algorithmic trading," *Research in International Business and Finance*, årg. 54, dec. 2020. DOI: 10.1016/j.ribaf.2020.101283.
- [11] L. Downey, "Efficient-market hypothesis," *Investopedia*, [Online], Mars 25, 2021. Tillgänglig: <https://www.investopedia.com/terms/e/efficient-markethypothesis.asp> (hämtad: 2021-05-08).
- [12] Techopedia, "Web Scraping," [Online]. Tillgänglig: <https://www.techopedia.com/definition/5212/web-scraping> (hämtad: 2021-05-09).
- [13] M. Gray, "Credits and Background," *Internet Statistics: Web Growth, Internet Growth*, [Online], 1996. Tillgänglig: <https://www.mit.edu/~mkgray/net/background.html> (hämtad: 2021-04-18).
- [14] eBay, "eBay Developers Program," 2021. [Online]. Tillgänglig: <https://developer.ebay.com/> (hämtad: 2021-04-18).
- [15] Amazon, "Amazon EC2 On-Demand Pricing," 2021. [Online]. Tillgänglig: https://aws.amazon.com/ec2/pricing/on-demand/#Data_Transfer (hämtad: 2021-05-13).
- [16] B. H. M. Baron J. Brogaard och A. Kirilenko, "Risk and Return in High-Frequency Trading," *Journal of Financial and Quantitative Analysis*, 2019. DOI: 10.1017/S0022109018001096.
- [17] L. S. Chong, K. M. Lim och C. P. Lee, "Stock Market Prediction using Ensemble of Deep Neural Networks," i *IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 2020. DOI: 10.1109/IICAIET49801.2020.9257864.
- [18] Alexa, "Alexa - Top sites," 2021. [Online]. Tillgänglig: <https://www.alexa.com/topsites> (hämtad: 2021-04-18).
- [19] Reddit, "wallstreetbets," 2021. [Online]. Tillgänglig: <https://www.reddit.com/r/wallstreetbets/> (hämtad: 2021-04-20).
- [20] Z. Dai, X. Dong, J. Kang och L. Hong, "Forecasting stock market returns: New technical indicators and two-step economic constraint method," *North American Journal of Economics and Finance*, årg. 53, juli 2020. DOI: 10.1016/j.najef.2020.101216.
- [21] T. H. Nguyen, K. Shirai och J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications*, årg. 42, nr 24, s. 9603–9611, 2015, ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2015.07.052>.
- [22] H. K. Sul, A. R. Dennis och L. I. Yuan, "Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns," *Decision Sciences*, årg. 48, nr 3, s. 454–488, 2017. DOI: <https://doi.org/10.1111/dec.12229>.
- [23] H. Gui, "Stock Prediction Based on Social Media Data via Sentiment Analysis: a Study on Reddit," mastersuppsats, Faculty of Information Technology

- and Communication Sciences, Tampere University, Tampere, Finland, 2019. [Online]. Tillgänglig: <http://urn.fi/URN:NBN:fi:tuni-201911246223>.
- [24] Avanza, "Vad är en aktie," Avanza. [Online]. Tillgänglig: <https://www.avanza.se/lar-dig-mer/avanza-akademien/aktier/vad-ar-aktier.html> (hämtad: 2021-04-25).
- [25] J. Ross, "The Pyramid of Equity Returns: Almost 200 Years of U.S. Stock Performance," *Advisor Channel*, [Online], Juni 25, 2020. Tillgänglig: <https://advisor.visualcapitalist.com/historical-stock-market-returns/> (hämtad: 2021-04-29).
- [26] J. Berk och P. DeMarzo, *Corporate Finance*. London: Pearson, 2017.
- [27] E. P. Chan, *Algorithmic Trading*. Hoboken, New Jersey: John Wiley & Sons, 2013.
- [28] L. Frazier, "The Coronavirus Crash Of 2020, And The Investing Lesson It Taught Us," *Forbes*, [Online], Feb. 11 2021. Tillgänglig: <https://www.forbes.com/sites/lizfrazierpeck/2021/02/11/the-coronavirus-crash-of-2020-and-the-investing-lesson-it-taught-us/> (hämtad: 2021-05-08).
- [29] F. Imbert, "Dow drops nearly 3,000 points, as coronavirus collapse continues; worst day since '87," *CNBC*, [Online]. Mars 16. 2020. Tillgänglig: <https://www.cnbc.com/2020/03/15/traders-await-futures-open-after-fed-cuts-rates-launches-easing-program.html>, Hämtad: 2021-04-28.
- [30] J. Fernando, "Moving Average (MA)," *Investopedia*, [Online], Jan. 17, 2021. Tillgänglig: <https://www.investopedia.com/terms/m/movingaverage.asp> (hämtad: 2021-04-26).
- [31] C. J. Neely, D. E. Rapach, J. Tu och G. Zhou, "Forecasting the equity risk premium: the role of technical indicators," *Management science*, årg. 60, nr 7, s. 1772–1791, 2014.
- [32] M. F. Dicle och J. D. Leventis, "Technical Financial Analysis Tools for Stata," *The Stata Journal*, årg. 17, nr 3, s. 736–747, 2017. DOI: 10.1177/1536867X1701700312.
- [33] A. Hayes, "Bollinger Band® Definition," *Investopedia*, [Online], April 30, 2021. Tillgänglig: <https://www.investopedia.com/terms/b/bollingerbands.asp> (hämtad: 2021-05-09).
- [34] I. Gurrib och F. Kamalov, "The implementation of an adjusted relative strength index model in foreign currency and energy markets of emerging and developed economies," *Macroeconomics and Finance in Emerging Market Economies*, årg. 12, nr 2, 2019. DOI: 10.1080/17520843.2019.1574852.
- [35] J. J. W. Wilder, *New concepts in technical trading systems*. Winston-Salem, New Jersey: Hunter Publishing Company, 1978.

-
- [36] K. Lam och H. C. Yam, "Cusum Techniques for Technical Trading in Financial Markets," *Asia-Pacific Financial Markets*, årg. 4, nr 3, jan. 1997. DOI: 10.1023/A:1009604804110.
- [37] SVT, "Bottarna som hotar börsen," *SVT Play*, April 18, 2021. [Video]. Tillgänglig: <https://www.svtplay.se/video/28815586/bottarna-som-hotar-borsen>.
- [38] Reddit, "stocks," 2021. [Online]. Tillgänglig: <https://www.reddit.com/r/stocks/> (hämtad: 2021-05-09).
- [39] Reddit, "investing," 2021. [Online]. Tillgänglig: <https://www.reddit.com/r/investing/> (hämtad: 2021-05-09).
- [40] Github, "PRAW," 2021. [Online]. Tillgänglig: <https://github.com/praw-dev/praw> (hämtad: 2021-02-05).
- [41] MonkeyLearn, "Everything There Is to Know about Sentiment Analysis," 2021. [Online]. Tillgänglig: <https://monkeylearn.com/sentiment-analysis/> (hämtad: 2021-05-06).
- [42] C. Schneider, "The biggest data challenges that you might not even know you have," *Watson blog*, [Online], Maj 25, 2016. Tillgänglig: <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/> (hämtad: 2021-04-22).
- [43] C. Hutto och E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media*, årg. 8, nr 1, maj 2014. URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>.
- [44] Github, "vaderSentiment," 2021. [Online]. Tillgänglig: <https://github.com/cjhutto/vaderSentiment> (hämtad: 2021-02-05).
- [45] K. Krol, "How to Code a Website," *WebsiteSetup*, Dec. 28, 2020. [Online]. Tillgänglig: <https://websitesetup.org/website-coding-html-css/> (hämtad: 2021-05-06).
- [46] MDN, "HTML: HyperText Markup Language," 2021. [Online]. Tillgänglig: <https://developer.mozilla.org/en-US/docs/Web/HTML> (hämtad: 2021-05-06).
- [47] MDN, "JavaScript," 2021. [Online]. Tillgänglig: <https://developer.mozilla.org/en-US/docs/Web/JavaScript> (hämtad: 2021-05-06).
- [48] Github, "Selenium Python Bindings Documentation," 2020. [Online]. Tillgänglig: <https://github.com/baijum/selenium-python> (hämtad: 2021-04-29).
- [49] Kent State University, "Pearson Correlation," 2021. [Online]. Tillgänglig: <https://libguides.library.kent.edu/SPSS/PearsonCorr> (hämtad: 2021-05-09).

-
- [50] E.W. Weisstein, "Covariance," *MathWorld*, [Online], Tillgänglig: <https://mathworld.wolfram.com/Covariance.html> (hämtad: 2021-05-09).
- [51] E.W. Weisstein, "Standard Deviation," *MathWorld*, [Online], Tillgänglig: <https://mathworld.wolfram.com/StandardDeviation.html> (hämtad: 2021-05-09).
- [52] S. Glen, "Bessel's Correction: Why Use N-1 For Variance/Standard Deviation?," *StatisticsHowTo.com*, [Online], Juli 25, 2018. Tillgänglig: <https://www.statisticshowto.com/bessels-correction/> (hämtad: 2021-05-09).
- [53] S. Glen, "P-Value in Statistical Hypothesis Tests: What is it?," *StatisticsHowTo.com*, [Online], Tillgänglig: <https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/p-value/> (hämtad: 2021-05-09).
- [54] Lumen Learning, "Testing the Significance of the Correlation Coefficient". [Online]. Tillgänglig: <https://courses.lumenlearning.com/introstats1/chapter/testing-the-significance-of-the-correlation-coefficient/> (hämtad: 2021-05-14).
- [55] L. Råde, B. Westergren och F. Wikström, *Mathematics Handbook For Science and Engineering*, 6. utg. Lund, Sweden: Studentlitteratur AB, 2019.
- [56] J. Kuepper, "Volatility," *Investopedia*, [Online], April 16, 2021. Tillgänglig: <https://www.investopedia.com/terms/v/volatility.asp> (hämtad: 2021-04-21).
- [57] Alpaca, "Market data streaming," 2021. [Online]. Tillgänglig: <https://alpaca.markets/docs/api-documentation/api-v2/market-data/streaming/> (hämtad: 2021-02-05).
- [58] Backtrader, "Welcome to backtrader!," 2020. [Online]. Tillgänglig: <https://www.backtrader.com> (hämtad: 2021-05-14).
- [59] SCB, "Statistikguiden - Frågeteknik." [Online]. Tillgänglig: <https://www.scb.se/dokumentation/statistikguiden/undersokning-och-urval/frageteknik/> (hämtad: 2021-05-13).
- [60] I. Aldridge, *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. Hoboken, New Jersey: John Wiley & Sons, Incorporated, 2013.
- [61] Investopedia, "How Do I Create a Trading Strategy With Bollinger Bands and the RSI?," 2021. [Online]. Tillgänglig: <https://www.investopedia.com/ask/answers/121014/how-do-i-create-trading-strategy-bollinger-bands-and-relative-strength-indicator-rsi.asp> (hämtad: 2021-05-11).
- [62] S. Gold, "Stock Market Algorithmic Trading: A Test of Bollinger Bands Incorporating the Squeeze Effect and MACD Conditions," *Journal of Applied Financial Research*, årg. 1, jan. 2018.

-
- [63] C. Mitchell, "Common Intra-Day Stock Market Patterns," *The Balance*, [Online], Aug. 24. 2020. Tillgänglig: <https://www.thebalance.com/common-intra-day-stock-market-patterns-1031456> (hämtad: 2021-04-28).
- [64] F. Wolf och O. Bergdorf, *Twitter Sentiment and Stock Returns*, Lund, Sweden: Department of Economics, Lund University, 2019. URL: <http://lup.lub.lu.se/student-papers/record/8984394>.
- [65] D. Fonda, "Brazil's Covid-19 Crisis Could Pose Trouble for a Global Travel Recovery. Here's Why.," *Barron's*, [Online], Mars 11, 2021. Tillgänglig: <https://www.barrons.com/articles/brazil-covid-19-travel-recovery-airline-stocks-51615485440> (hämtad: 2021-03-11).
- [66] Yahoo Finance, "Boeing (BA) Expects Big 737 MAX Orders From Southwest Airlines," 2021. [Online]. Tillgänglig: <https://finance.yahoo.com/news/boeing-ba-expects-big-737-165704167.html> (hämtad: 2021-03-11).
- [67] Y. Ebrahim, "S&P 500 Snaps 2-Day Losing Streak as Tech U-Turn Triggers Rally," *Investing.com*, [Online], Mars 25, 2021. Tillgänglig: <https://www.investing.com/news/stock-market-news/sp-500-snaps-2day-losing-streak-as-tech-uturn-triggers-rally-2457741> (hämtad: 2021-03-25).
- [68] F. Agomuoh, "Southwest Airlines to Reintroduce the Boeing 737 MAX," *TheStreet*, [Online], Mars 10, 2021. Tillgänglig: <https://www.thestreet.com/investing/southwest-air-to-resume-flying-the-boeing-737-max> (hämtad: 2021-03-10).
- [69] Benzinga, "Merck & Co's Debt Overview," 2021. [Online]. Tillgänglig: https://www.benzinga.com/news/21/03/20196195/merck-cos-debt-overview?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+benzinga+%28Benzinga+News+Feed%29 (hämtad: 2021-03-16).
- [70] Benzinga, "Boeing's Debt Overview," 2021. [Online]. Tillgänglig: <https://www.benzinga.com/news/21/03/20282230/boeings-debt-overview> (hämtad: 2021-03-22).
- [71] Yahoo Finance, "Influencers with Andy Serwer: Ruth Porat," 2021. [Online]. Tillgänglig: <https://www.yahoo.com/gma/video/influencers-andy-serwer-ruth-porat-172633641.html> (hämtad: 2021-03-18).
- [72] A. Martin, "iPhone vs Android: How to decide which smartphone is best for you," *The Independent*, [Online], Mars 16, 2021. Tillgänglig: <https://www.independent.co.uk/extras/indybest/gadgets-tech/phones-accessories/iphone-vs-android-apple-b1817403.html> (hämtad: 2021-03-16).
- [73] L. C. Martinez, D. N. da Hora, J. R. de M. Palotti, W. Meira och G. L. Pappa, "From an artificial neural network to a stock market day-trading system: A case study on the BM F BOVESPA," i *2009 International Joint Conference on Neural Networks*, 2009, s. 2006–2013. DOI: 10.1109/IJCNN.2009.5179050.

-
- [74] B. Beers, "How the News Affects Stock Prices," *Investopedia*, [Online], Jan. 9, 2020. Tillgänglig: <https://www.investopedia.com/ask/answers/155.asp> (hämtad: 2021-05-08).
- [75] B. Beers, "Buy and Hold Definition," *Investopedia*, [Online], Maj 13, 2020. Tillgänglig: <https://www.investopedia.com/terms/b/buyandhold.asp> (hämtad: 2021-05-08).
- [76] J. Tierney, "Do You Suffer From Decision Fatigue?," *The New York Times Magazine*, [Online], Aug. 17, 2011. Tillgänglig: <https://www.nytimes.com/2011/08/21/magazine/do-you-suffer-from-decision-fatigue.html> (hämtad: 2021-05-09).
- [77] Github, "AutoWikibot," 2014. [Online]. Tillgänglig: <https://github.com/acini/autowikibot-py> (hämtad: 2021-04-26).
- [78] Reddit, "AutoModerator," 2020. [Online]. Tillgänglig: <https://www.reddit.com/wiki/automoderator> (hämtad: 2021-04-26).
- [79] R. Maurer, "Best Stocks of the Year: Tesla Is Number 1," *TheStreet*, [Online], Dec. 21, 2020. Tillgänglig: <https://www.thestreet.com/investing/best-stocks-of-2020-tesla-number-one> (hämtad: 2021-04-29).
- [80] R. Dhir, "Market Order Definition," *Investopedia*, [Online], April 18, 2021. Tillgänglig: <https://www.investopedia.com/terms/m/marketorder.asp> (hämtad: 2021-04-29).
- [81] M. J. Kramer, "Limit Order," *Investopedia*, [Online], Jan. 28, 2021. Tillgänglig: <https://www.investopedia.com/terms/l/limitorder.asp> (hämtad: 2021-04-29).
- [82] A. Hayes, "Short Selling," *Investopedia*, [Online], Mars 13, 2021. Tillgänglig: <https://www.investopedia.com/terms/s/shortselling.asp> (hämtad: 2021-04-29).
- [83] W. Kenton, "Order Book Definition," *Investopedia*, [Online], April 4, 2021. Tillgänglig: <https://www.investopedia.com/terms/o/order-book.asp> (hämtad: 2021-04-29).
- [84] Avanza, "Vad kostar det att handla aktier?," [Online]. Tillgänglig: <https://www.avanza.se/lar-dig-mer/avanza-akademin/aktier/vad-kostar-det-att-handla-med-aktier.html> (hämtad: 2021-04-27).
- [85] *Alpaca Securities Brokerage Fee Schedule*: Alpaca Securities LLC. [Online]. Tillgänglig: https://files.alpaca.markets/disclosures/4.17.2020+Alpaca_Fee_Schedule_V2.pdf, Hämtad: 2021-04-27.
- [86] A. Ganti, "Dow Jones Industrial Average (DJIA)," *Investopedia*, [Online], Mars 8, 2021. Tillgänglig: <https://www.investopedia.com/terms/d/djia.asp> (hämtad: 2021-02-05).

A

Aktiemarknadens mikrostruktur

A.1 Olika typer av ordrar

En **marknadsorder** kan vara av typen köp eller sälj och innebär att en aktie handlas till det bästa pris som finns tillgängligt på marknaden [80]. Detta innebär att köparen inte specificerar något pris, men däremot är ordern av en bestämd kvantitet.

En **limitorder** innebär att köparen anger en bestämd kvantitet *och* ett bestämt pris för vilket denne vill köpa eller sälja en aktie [81]. När denna ordern läggs så hamnar den i orderboken, vars funktion beskrivs senare i denna bilaga.

Att **blanka** en aktie (från engelskans *short selling*) är en speciell typ av order som innebär att en aktie "lånas" under en bestämd tidsperiod [82]. Den som lånar aktien spekulerar att aktiens pris kommer att falla. Efter att aktien lånats kan den säljas till sitt rådande marknadspris, och när aktien sedan ska lämnas tillbaka köps en ny aktie från marknaden. Om priset på aktien då har fallit under den period som aktien lånades kan investeraren köpa aktien billigare än vad den såldes för. Investeraren har således gjort en vinstaffär. Om istället aktiepriset har stigit har investeraren gjort en förlustaffär. Förenklat innebär detta att det är möjligt att göra vinstaffärer på en aktiekurs som faller.

A.2 Orderboken

Orderboken består av en samling limitordrar vid olika prisnivåer både från köpsidan och säljsidan, de billigaste priserna som investerare vill sälja aktien för, samt det lägsta priserna som investerare vill köpa aktien för [83]. Orderboken kan exempelvis se ut som på bilden nedan.

Antal	Köp	Sälj	Antal
189	69,90	70,10	569
624	69,85	70,20	1 358
404	69,80	70,30	3 990
341	69,75	70,35	850
14	69,70	70,40	3 150

Figur A.1: Exempel på orderboken. I figuren för Tobii AB 2021-04-29. Hämtad från Avanza. Köp- och säljpriser visas i SEK.

Som kan ses i figur A.1 visas fem prisnivåer från både sälj- och köpsidan, samt den ackumulerade volymen av aktier vid varje prisnivå. Orderboken kan därför sägas vara en sammanställning av utbud och efterfråga i aktien. Om det utgås från detta orderdjup skulle en investerare, som lägger en köpmarknadsorder, köpa aktier till priset 70,10 kr per aktie, vilket är det billigaste priset som är tillgängligt [60]. Om köpordern är större än den ackumulerade volymen vid priset av 70,10 SEK, i.e. 569 st, handlas resterande aktier till det näst billigaste priset. På samma sätt skulle en investerare som lägger en säljmarknadsorder sälja sina aktier till det högsta köppriset, vilket är 69,90 kr.

Om en investerare istället lägger en *limitorder* fungerar de flesta börser så att ordern hamnar längst bak i kön vid den prisnivå som ordern läggs på [60]. Detta innebär att om en säljlimitorder läggs vid priset 70,30 kr måste de första 3 990 aktierna köpas innan denna investerares order kan fyllas. Följden av detta är att det behövs tas hänsyn till flera faktorer när det handlas med limitordrar, såsom position i kön och även ifall en annan aktör placerar en limitorder till ett bättre pris. I fallet med orderdjupet i figur A.1 så skulle någon kunna placera en säljlimitorder till priset 70,05 kr och då ta positionen med det billigaste säljpriset.

Då det istället handlas med marknadsordrar köper eller säljer investeraren alltid till det bästa tillgängliga priset, vilket försäkrar att dennes order alltid kommer att fyllas. Priset att betala för detta är den så kallade *budfördelningen*, vilket är differensen mellan det lägsta köppriset och det högsta säljpriset [60]. Budfördelningen beskrivs noggrannare under följande delkapitel.

Enligt [7, s. 145] kan obalanser mellan köp- och säljsidan i orderboken identifieras och användas i investeringsalgoritmer. En obalans i utbud och efterfråga kan innehålla information om aktieprisets kortsiktiga rörelser.

A.3 Transaktionskostnader

Transaktionskostnaderna är den kostnad en investerare måste betala för att köpa eller sälja aktier [7, s. 67]. Denna kostnad är uppbyggd av handelsavgifter, budfördelning, marknadspåverkan, och fördröjning. Konsekvensen av att bortse från transaktionskostnader är att en förlustdrivande algoritm misstas för en vinstdrivande algoritm.

Handelsavgifter består dels den avgift som aktiemäklaren begär för att genomföra en transaktion på en investerares förfrågan (även känt som *courtage*) [84]. Hur denna kostnad beräknas är olika beroende på mäklaren. Den mäklartjänst som använts under realtidshandeln i detta projekt (Alpaca) erbjöd gratis mäklaravgift [85]. Handelsavgifter består även av reglerande kostnader, vilket på den amerikanska marknaden är i form av avgifter till Securities and Exchange Commission (SEC) samt Financial Industry Regulatory Authority (FINRA). Avgiften som betalas till SEC är \$22 per \$1 000 000 och avgiften till FINRA är \$0,000119 per aktie [85].

Budfördelning (engelskans *bid/ask-spread*) är som tidigare nämnt differensen mellan det lägsta köppriset och det högsta säljpriset, och är en kostnad som uppstår då det handlas med marknadsordrar. Teorin bakom budfördelningen är att köparen måste möta det billigaste säljpriset, vilket innebär att denne betalar mer för aktien än om den hade sålts med en limitorder [60].

Om investeraren handlar med limitordrar talas det istället om möjlighetskostnader, vilket är en implicit kostnad som motsvarar kostnaden av att exempelvis missa en uppgång på grund av att köparen lade en köplimitorder som ej fylldes innan aktiekursen steg.

Enligt [9, s. 23] kan budfördelningskostnaden estimeras till halva budfördelningen. Med utgångspunkt i orderboken i figur A.1 hade en säljmarknadsorder genomförs till kursen 69,90 kr, och budfördelningskostnaden blir i detta exempel 0,10 kr.

Om en säljlimitorder hade lagts till priset 70,05 kr hade möjlighetskostnaden i detta exempel uppstått om aktiekursen skulle falla, vilket resulterar i att köparen i ett senare skede kan behöva sälja aktien ännu billigare.

Fördröjning (engelskans *slippage*) är den eventuella extrakostnaden som betalas på grund av tidsfördröjning från det att en order placeras tills det att aktiemäklaren registrerar och bekräftar ordern.

A.4 Hur mikrostrukturen hanterades i projektet

Investeringsalgoritmen handlar endast med marknadsordrar för att försäkra att varje order som placeras fylls. Detta garanteras inte då limitordrar används. För att simulera att investeringsalgoritmen handlar med marknadsordrar då den testas mot historiska aktiekurser, genomförs köpet alltid till öppningspriset för den efterföljande datapunkten. Detta öppningspris skiljer sig något från det pris som triggat köpsig-

nalen, på så vis simuleras även fördröjningen, vars konsekvenser beskrevs ovan.

I avsnittet ovan beskrivs vilka handelsavgifter som uppstår vid aktiehandel på den amerikanska marknaden. För att efterlikna verkligheten gjordes en estimering av vad dessa avgifter motsvarar procentuellt. Denna estimering beräknades på följande sätt:

Snittpriset för en aktie som ingår i DJIA är \$176,2. Detta snittpris är beräknat från stängningskurserna för samtliga ingående aktier den 7:e maj 2021. Hur många aktier som handlas vid varje transaktion i förhållande till genomsnittspriset beräknas enligt:

$$n = \frac{x}{\$176,2} \quad (\text{A.1})$$

Där x är storleken på kapitalet som investeras och n är antalet aktier som köps.

Avgiften som betalas till SEC är \$22 per \$1 000 000, och avgiften som betalas till FINRA är \$0,000119 per aktie. Tillsammans beräknas den procentuella estimeringen genom att lägga ihop handelsavgifterna för en order och dividera med kostnaden för ordern enligt:

$$\begin{aligned} \text{Handelsavgift} &= \frac{\frac{\$22}{\$1000000} \cdot x + \$0,000119 \cdot n}{x} \\ &= \frac{\$22}{\$1000000} + \frac{\$0,000119 \cdot \frac{x}{\$176,2}}{x} \\ &= 0,00002267 \\ &\approx 0,0023\%. \end{aligned} \quad (\text{A.2})$$

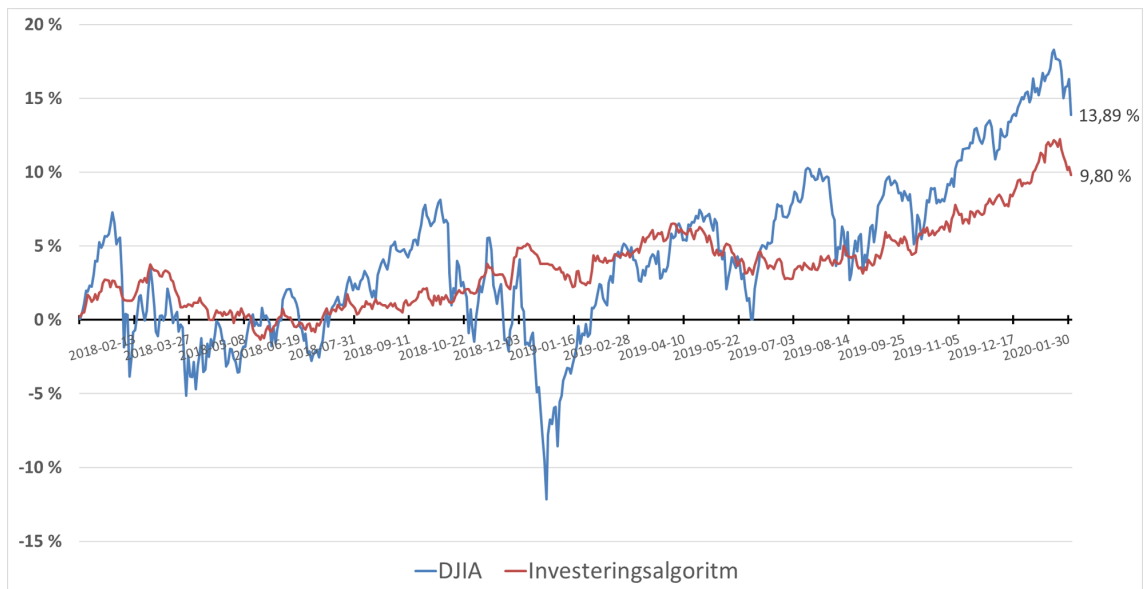
Alltså användes en handelsavgift på 0,0023 % vid varje transaktion. Det vill säga såväl köporder som säljorder.

B

Investeringsalgoritmernas simuleringsresultat

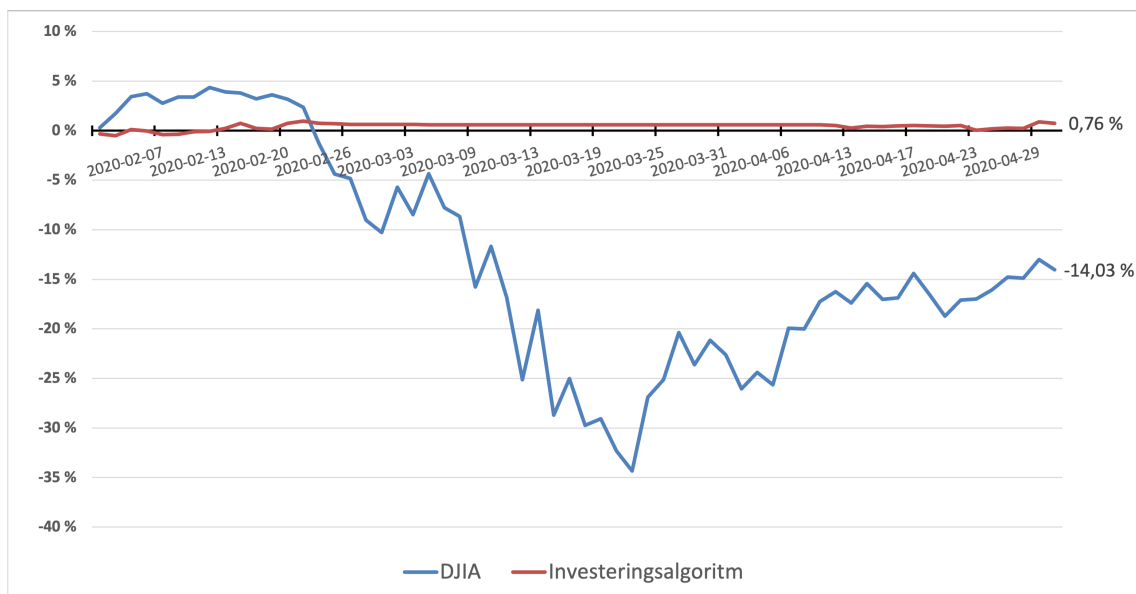
Här presenteras samtliga grafer för den historiskt simulerade avkastningen från investeringsalgoritmerna i jämförelse med Dow Jones Industrial Average. Först presenteras investeringsalgoritmen för intradagshandel där inga positioner hålls över natten och sedan investeringsstrategin för interdagshandel där positioner över flera dagar tillåts. Graferna presenteras i kronologisk ordning där första grafen är träningssetet, andra grafen är coronasetet och tredje grafen är det slutliga testsetet.

B.1 Intradagshandel

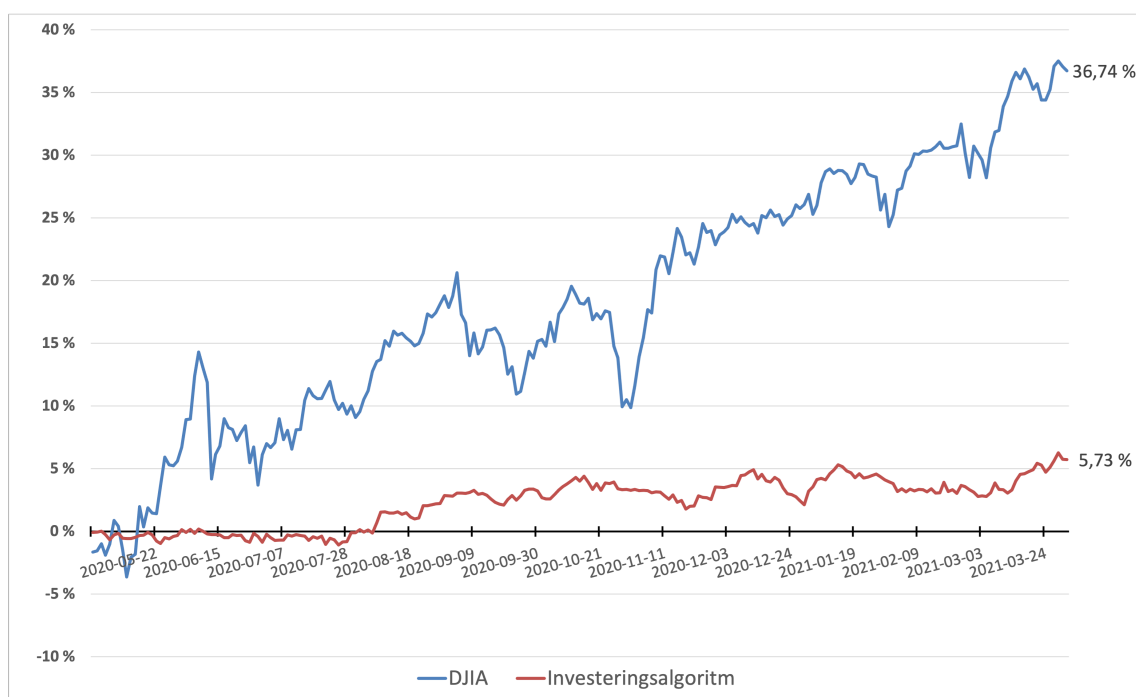


Figur B.1: Intradagsalgoritmen vs. DJIA - Träningsset.

B. Investeringsalgoritmernas simuleringsresultat

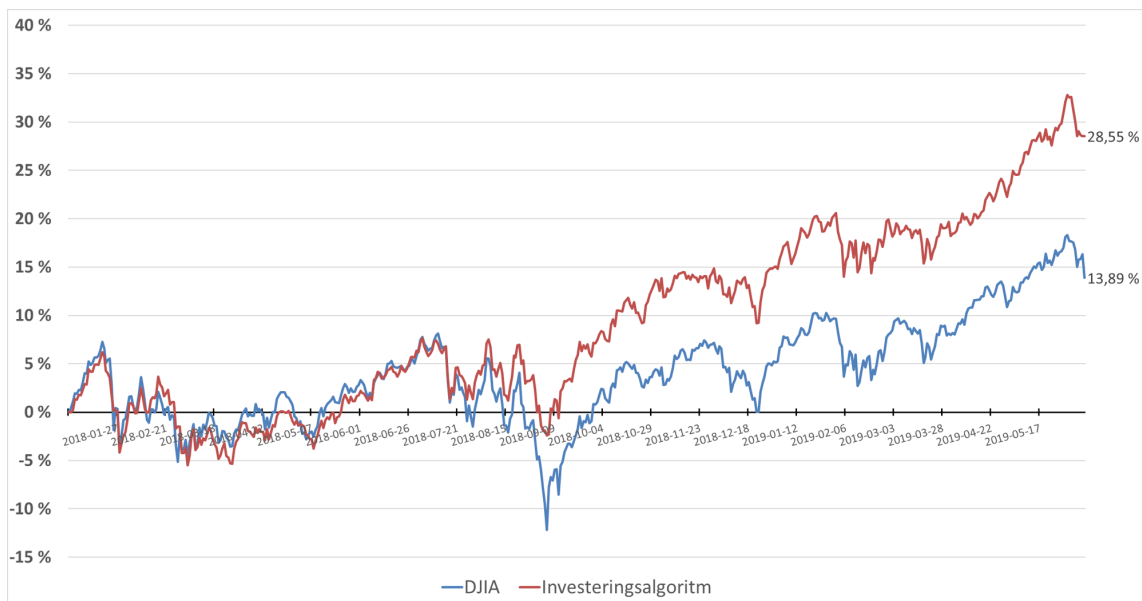


Figur B.2: Intradagsalgoritmerna vs. DJIA - Coronaset.

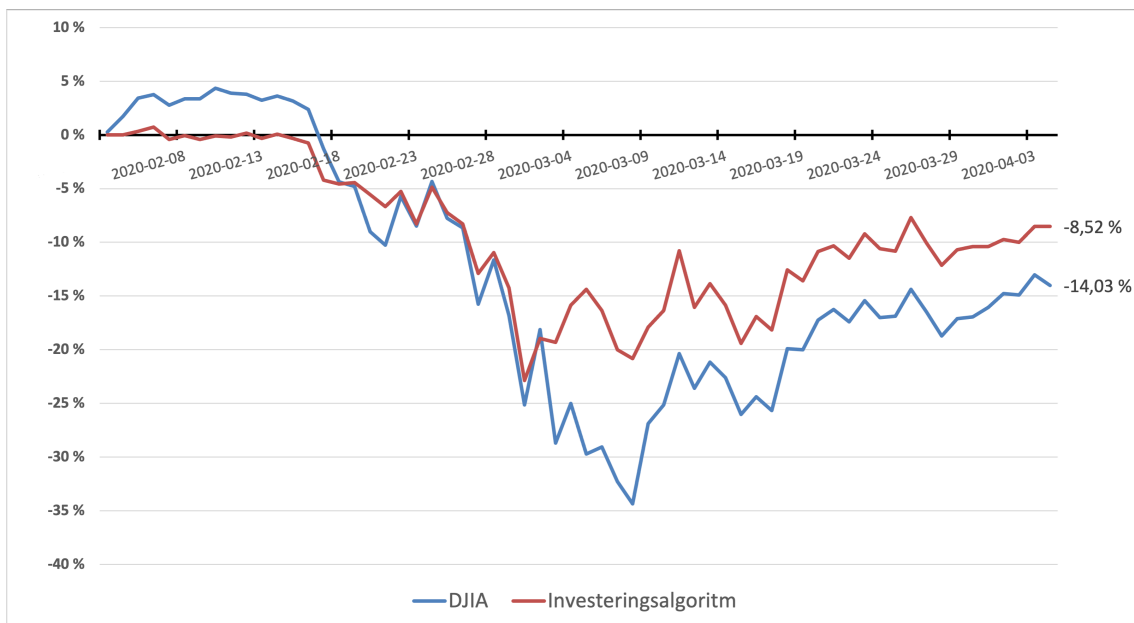


Figur B.3: Intradagsalgoritmerna vs. DJIA - Testset.

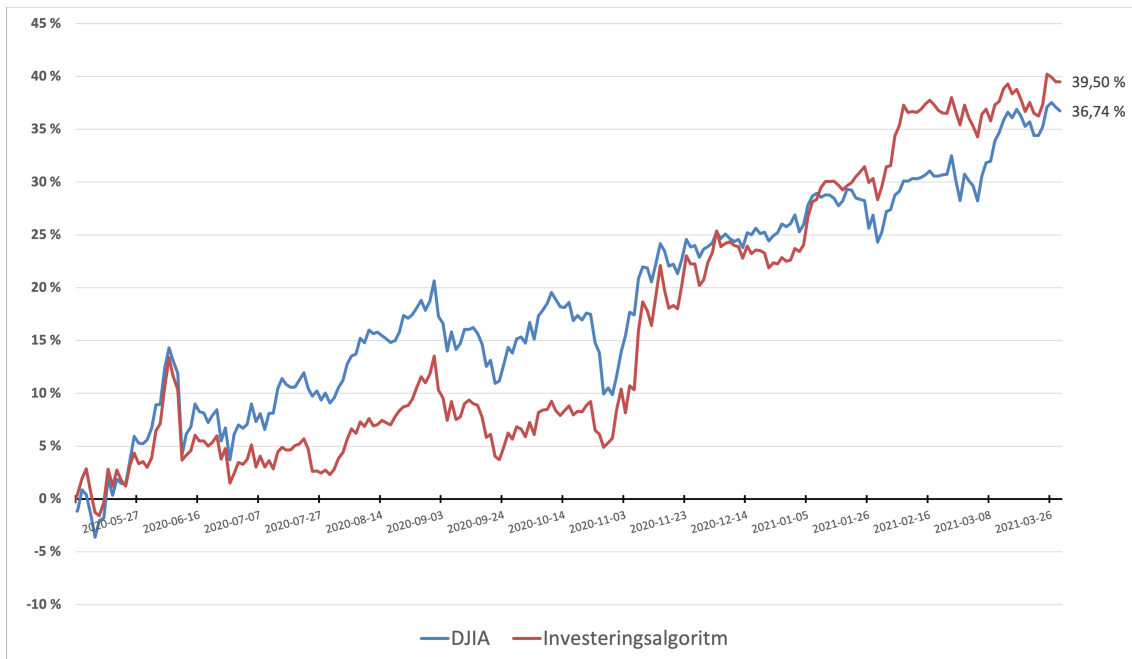
B.2 Interdagshandel



Figur B.4: Interdagsalgoritmerna vs. DJIA - Träningsset.



Figur B.5: Interdagsalgoritmerna vs. DJIA - Coronaset.



Figur B.6: Interdagsalgoritmen vs. DJIA - Testset.

C

Dow Jones Industrial Average (DJIA)

Tabell C.1: Lista på alla 30 företag som ingick i DJIA i augusti 2020 [86].

Företag	Symbol
3M	MMM
American Express	AXP
Amgen	AMGN
Apple Inc.	AAPL
Boeing	BA
Caterpillar	CAT
Chevron	CVX
Cisco Systems	CSCO
The Coca-Cola Company	KO
Dow Inc.	DOW
Goldman Sachs	GS
The Home Depot	HD
Honeywell	HON
IBM	IBM
Intel	INTC
Johnson & Johnson	JNJ
JPMorgan Chase	JPM
McDonald's	MCD
Merck & Co.	MRK
Microsoft	MSFT
NIKE	NKE
Proctor & Gamble	PG
Salesforce	CRM
The Travelers Companies	TRV
UnitedHealth Group	UNH
Verizon	VZ
Visa	V
Walmart	WMT
Walgreens Boots Alliance	WBA
The Walt Disney Company	DIS

INSTITUTIONEN FÖR DATA- OCH INFORMATIONSTEKNIK
CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden

www.chalmers.se



GÖTEBORGS
UNIVERSITET



CHALMERS