



UiT Norges arktiske universitet

Handelshøgskolen ved UiT

Automatisert Aksjeanalyse med GPT

Hvor effektivt er GPT som et verktøy for å predikere avkastningen basert på årsrapporter for selskaper notert på Oslo Børs?

Thomas Larsen og Tim Andre Strøm Øyen

Masteroppgave i økonomi og administrasjon, BED-3901, juni 2024

Forord

Denne masteroppgaven er utarbeidet som avsluttende arbeid for et 2-årig masterstudium i Økonomi og Administrasjon ved Handelshøgskolen, UiT Norges arktiske universitet. Oppgaven bygger på den kunnskapen vi har tilegnet oss gjennom vår spesialisering i økonomisk styring og finans. Arbeidet med oppgaven tilsvarer ett semester med fulltidsstudier.

På bakgrunn av den omfattende omtalen rundt kunstig intelligens i media, ble vi nysgjerrige på om det var mulig å benytte kunstig intelligens for å oppnå meravkastning i aksjemarkedet. Arbeidet med denne oppgaven har gitt oss muligheten til å fordype oss i et fremtidsrettet, spennende og svært dagsaktuelt tema. Vi har tilegnet oss kunnskap som er relevant og betydningsfullt for mange aspekter av samfunnet.

Vi ønsker å rette en stor takk til vår veileder Espen Sirnes. Han har bidratt til mange nyttige råd til oppgaven, og gitt oss mye god data for analysedelen i oppgaven vår. Videre vil vi takke våre medstudenter for fine samtaler og mange gode latter. I tillegg ønsker vi å rette en takk til familie og venner for støtten de har gitt oss.

Tromsø, juni 2024

Thomas Larsen og Tim Øyen

Sammendrag

Etter flere år med stor vekst og utallige nye bruksområder, kan vi være sikre på at kunstig intelligens er kommet for å bli. Ettersom det har utviklet seg nye teknologier med kunstig intelligens, synes vi det er svært spennende å undersøke mulighetene for nyere metoder for predikering av finansmarkedet. Historisk sett har prediksjon av aksjemarkedet vært en utfordrende og kompleks oppgave, og det er i teorien ofte beskrevet som et forsøk på å forutse en uforutsigbar fremtid. Med fremveksten av maskinlæring, presenterer kunstig intelligens en unik mulighet til å automatisere og forbedre analyseprosessene for aksjemarkedet.

I denne oppgaven ser vi nærmere på om kunstig intelligens kan brukes til å predikere aksjemarkedet. Dette gjøres gjennom å bruke Generative Pre-trained Transformer (GPT) analysere innhold i årsrapporter. Målet er å undersøke om det er mulig å oppnå en meravkastning sammenlignet med indeksen på Oslo Børs. Datamaterialet som studeres er hentet fra databasen i TITLON i tidsperioden 2010 og 2011. På bakgrunn av dette har vi formulert følgende problemstilling for undersøkelsen vår:

Hvor effektivt er GPT som et verktøy for å predikere avkastningen basert på årsrapporter for selskaper notert på Oslo Børs?

Oppgaven undersøker også hvorvidt GPT er i stand til å identifisere selskaper med positive eller negative forutsetninger for fremtidig vekst. Informasjon fra årsrapportene er brukt som grunnlag for at GPT skal gi karakterer til de ulike selskapene. Karakterskalaen går fra 0-10, der 10 er høyeste karakter og 0 er laveste karakter. Det er anvendt en kvantitativ metode for å besvare problemstillingen og forskningsspørsmålene. Karakterene blir flere steder kategorisert i karaktergruppe 1, 2 og 3. Karakter 0-3 er karaktergruppe 1, karakter 4-6 er karaktergruppe 2 og karakter 7-10 er karaktergruppe 3.

I 2010 opplevde børsen en positiv utvikling, men 2011 var et år preget av negative resultater. Dette reflekterer de generelle markedsforholdene i de respektive årene. Figurene i analysedelen illustrerer at GPT har en evne til å identifisere både underpresterende selskaper og de som klarer seg bedre, selv om det ikke resulterer i meravkastning sammenlignet opp mot Oslo Børs.

Nøkkelord: GPT, Kunstig intelligens, Oslo Børs, Årsrapporter, Aksjeanalyse

Innholdsfortegnelse

Forord	i
Sammendrag	ii
Figurer	v
Tabeller	vi
1 Introduksjon	1
1.1 Aktualisering.....	1
1.2 Problemstilling	2
1.3 Begrepsavklaring.....	2
1.4 Strukturen i arbeidet.....	3
2 Teoretiske perspektiver	5
2.1 Kunstig intelligens	5
2.1.1 Kunstig intelligens i et historisk perspektiv	5
2.1.2 GPT	7
2.2 Ulike felt innen AI.....	8
2.2.1 Maskinl�ring.....	9
2.2.2 Dyp l�ring og forsterkende l�ring.....	10
2.2.3 Dypl�ring	11
2.2.4 Transformer-modellen.....	12
2.2.5 Kunstige nevrale nettverk.....	12
2.3 AI i finanssektoren	13
2.3.1 Fordeler knyttet til AI i bank og finanssektoren.....	13
2.4 Udfordringer knyttet til AI.....	15
2.4.1 Juridiske og etiske problemstillinger.....	15
2.4.2 Datapersonvern og sikkerhet i finanssektoren.....	15
2.4.3 Black box.....	16
2.5 Finansanalyse og prediksj�n	17
2.5.1 Teknisk analyse	17
2.5.2 Fundamental analyse	17
2.5.3 Maskinl�ring i finans	18
2.6 Markedet.....	18
2.6.1 Markedseffisiens	18
2.6.2 Random Walks	20
2.6.3 Moderne portef�ljeteori.....	20
2.7 Faktormodeller	21
2.7.1 Markedsmodellen	21
2.7.2 Fama og French sin trefaktormodell	23
2.7.3 Carhart fire-faktormodell	24
2.8 Faktorforklaringer	25
2.8.1 SMB (Small Minus Big)	25
2.8.2 HML (High Minus Low).....	25
2.8.3 LIQ (Liquidity).....	25
2.8.4 MOM (Momentum).....	26

3. Forskningsdesign og metode	27
3.1 Forarbeid av metode.....	27
3.1.2 Fremgangsmåte	27
3.2 Forskningsdesign.....	28
3.2.1 Valg av forskningsmetode.....	29
3.2.2 Innsamling og analyse av data.....	29
3.3 Evaluering av metode	30
3.3.1 Validitet.....	30
3.4.2 Reliabilitet	31
3.4.3 Vurdering av GPT	31
4 Analyse og resultater.....	32
4.1 Avkastning for karakter grupper sammenlignet med Oslo Børs	32
4.2 Statistikk.....	34
4.2.1 Standardfeil	35
4.2.2 T-verdi	35
4.2.3 Analyse av standardfeil og T-verdi	36
4.2.4 Konfidensintervall	36
4.3 Regresjonsmodell.....	37
4.4 Test av heteroskedasitet.....	38
4.5 OLS regresjon.....	39
4.6 Oppsummering.....	41
5 Diskusjon.....	42
5.1 Sammendrag av resultater	42
5.2 Konklusjon	43
5.3 Metodekritikk.....	44
5.3.1 Styrker ved metodikken.....	44
5.3.2 Svakheter ved metodikken	45
5.4 Forslag til videre forskning	45
Referanseliste.....	47
Vedlegg.....	51

Figurer

Figur 1: Markedsavkastningslinjen.....	22
Figur 2: Gjennomsnittlig avkastning for hver karaktergruppe og Oslo børs fra 1 juni 2010 til 1 juni 2011.	32
Figur 3: Gjennomsnittlig avkastning for hver karaktergruppe og Oslo børs fra 1 juni 2011 til 1 juni 2012.	33
Figur 4: OLS regresjon: T-Alfa mot karakternivå.....	40

Tabeller

Tabell 1: Sammenligning mellom antall GPT-modellparametere og mengden forhåndstreningsdata (OpenAI, 2024).....	8
Tabell 2: Resultater år 2010: Viser avkastning og statistikk for hver karaktergruppe for 2010.....	34
Tabell 3: Resultater år 2011: Viser avkastning og statistikk for hver karaktergruppe for 2011.....	35
Tabell 4: Regresjonsmodell med karakter 1 som referansenivå og alfa som avhengig variabel.....	37
Tabell 5: Viser tallene for Breusch-Pagan test	39

1 Introduksjon

Vi befinner oss i en ny industriell revolusjon, *Industri 4.0*, som kjennetegnes av at digitalisering av industrielle systemer øker. Data og informasjon flyter i sann tid, og avanserte algoritmer, maskinlæring og kunstig intelligens brukes for å forbedre industrisystemets produktivitet, kvalitet og adaptivitet (Martinsen, 2024). Den typen av teknologi som får størst oppmerksomhet for tiden er kunstig intelligens, også kalt AI, som kommer fra det engelske begrepet *Artificial intelligence*. AI refererer til simulerte menneskelige intelligensprosesser utført av maskiner, spesielt datamaskiner (AISuksess, u.å.). Teknologi kombinerer avanserte algoritmer og maskinlæring med mål om å utføre oppgaver som tradisjonelt krever menneskelig intelligens. AI har mange bruksområder. Alt fra automatisering av oppgaver og forbedring av beslutningsprosesser til innovasjon innen for eksempel helsevesenet, finans og transport. Gjennom å etterligne kognitive funksjoner som læring og problemløsning, har AI et stort potensial til å revolusjonere mange aspekter i dagens samfunn (AISuksess, u.å.).

1.1 Aktualisering

I dagens stadig mer komplekse og dynamiske finansmarkeder søker investorer etter nye metoder og verktøy for å analysere selskapets ytelse og predikere fremtidig avkastning. Tradisjonelle metoder som teknisk og fundamental analyse er fremdeles veldig relevante, men kunstig intelligens (AI) har fått stadig større oppmerksomhet ettersom det er et kraftig verktøy som kan behandle store mengder data raskt og effektivt, og identifisere mønstre som ellers kunne blitt oversett. En av de mest populære AI-modellene for naturlig språkbehandling er GPT, utviklet av OpenAI. Modellen har vist seg lovende på mange områder grunnet dens evne til å forstå og generere menneskelig språk. Dette gjør den til et kraftig verktøy i mange ulike applikasjoner, og har potensial til å forbedre effektivitet og innovasjon på tvers av mange sektorer.

Denne oppgaven tar sikte på å utforske om GPT kan brukes til å analysere regnskaps- og nøkkeltall fra årsrapporter til selskaper listet på Oslo Børs. Gjennom koding skal vi hente ut og analysere innhold fra årsrapportene som til slutt skal interagere med GPT og som skal kategorisere selskapenes fremtidige utsikter ved å gi en karakter fra 0-10, der 10 er høyeste karakter og 0 er laveste karakter. Deretter skal vi liste opp selskapene med tilhørende

karakter, og sammenligne karakternivåene de ulike selskapene fikk. Videre skal vi undersøke om GPT har evnen til å forutse hvilke selskaper som har potensial til å gjøre det bra. Resultatene vil vi presentere i ulike figurer og tabeller, samt gjennomføre analyser og regresjoner for å vurdere og sammenligne funnene våre.

Denne studien er svært spennende fordi den utforsker et dagsaktuelt og fremtidsrettet tema. Det er interessant å undersøke om kunstig intelligens har potensiale til å analysere årsrapporter. Dersom det viser seg at kunstig intelligens er troverdig i denne sammenhengen, kan denne forskningen bidra til å forbedre beslutningsprosesser, øke effektiviteten og muliggjøre mer presise investeringer i aksjemarkedet.

1.2 Problemstilling

Kunstig intelligens er den nye store teknologien som veldig mange ønsker å utforske for å undersøke om den kan effektivisere arbeid. Formålet med denne studien er derfor å evaluere hvor god GPT er til å identifisere om aksjekursen til selskaper har positive eller negative fremtidsutsikter basert på analyser av årsrapporter. Vi undersøker om det dermed er et effektivt og troverdig verktøy for å oppnå en høyere meravkastning enn indeksen på Oslo Børs. Den overordnede problemstillingen for undersøkelsen vår er:

Hvor effektivt er GPT som et verktøy for å predikere avkastningen basert på årsrapporter for selskaper notert på Oslo Børs?

For å besvare denne problemstillingen har vi formulert to forskningsspørsmål:

1. Hvordan presterer AI-genererte karakternivåer sammenlignet med hovedindeksen på Oslo Børs i samme periode?
2. Hva er sammenhengen mellom karakterer gitt av GPT og selskapenes faktiske avkastning?

1.3 Begrepsavklaring

I dette kapittelet vil vi definere og forklare sentrale begreper som benyttes i denne masteroppgaven. Formålet er å sikre en felles forståelse av viktige konsepter som er relatert til oppgaven. En presis begrepsavklaring er viktig for å forstå sammenhengen mellom de ulike delene av oppgaven og den overordnede problemstillingen.

- Aksjekursutvikling: Refererer til endringen i verdien av en aksje over tid.
- GPT: Generative Pre-trained Transformer (GPT) er en avansert språkmodell utviklet av OpenAI, som genererer naturlig tekst basert på en gitt inngang ved bruk av transformer-arkitektur.
- Karaktergrupper: En sammensatt portefølje for selskapene som har fått karakter 1-3, 4-6 og 7-10 gitt av GPT. Disse karakterene deles inn i 3 karaktergrupper hvor 1-3 tilsvarer 1, 4-6 tilsvarer 2, og 7-10 tilsvarer 3.
- Maskinlæring: Bruk av maskinlæringsalgoritmer for å analysere store mengder data og finne mønstre som kan brukes til å forbedre beslutningsprosesser.
- Dyp læring (Deep Learning): En undergren av maskinlæring som benytter dype nevralt nettverk til å lære fra store mengder data.
- Forsterkende læring (Reinforcement Learning): En type maskinlæring der algoritmen lærer gjennom å samhandle med miljøet og få tilbakemelding i form av belønninger eller straff.
- Markedseffisiens: En hypotese som antyder at markeder raskt reflekterer den økonomiske påvirkningen av ny informasjon, noe som gjør det vanskelig for aktører å overgå markedet.

1.4 Strukturen i arbeidet

Oppgaven er strukturert i flere kapitler som til sammen skal gi en analyse om hvordan kunstig intelligens kan brukes til å predikere aksjekursutviklingen basert på årsrapporter for selskaper notert på Oslo Børs i perioden 2010 til 2011.

I innledinga, *kapittel 1*, introduserer vi bakgrunnen for studien, problemstillingen og forskningsspørsmålene, og forklarer hvorfor kunstig intelligens er relevant og spennende for finansmarkedet i dag. I *kapittel 2*, teoretiske perspektiver, presenterer vi det teoretiske rammeverket for oppgaven som dekker grunnleggende konsepter innen kunstig intelligens, maskinlæring, og dyp læring, samt deres anvendelser i finanssektoren. Videre diskuteres tidligere forskning og teoretiske modeller som er relevante for studiens problemstilling. I *kapittel 3* om forskningsdesign og metode beskriver vi hvilke forskningsdesign og metoder som er brukt for å samle inn og analysere data. Valget av kvantitativ metode begrunnes, og prosessen for datainnsamling fra TITLON-databasen forklares. Det gis så en detaljert beskrivelse av hvordan dataene er behandlet og analysert ved hjelp av GPT-modellen. I *kapittel 4* presenteres og analyseres resultatene fra studien. Det gjøres en gjennomgang av

avkastning for ulike karaktergrupper sammenlignet med Oslo Børs, samt resultater fra regresjonsmodeller og tester for heteroskedastisitet. Hovedfunnene diskuteres i lys av den teoretiske rammen. I *kapittel 5* diskuterer vi implikasjonene av funnene våre, begrensninger ved studien, og mulige fremtidige forskningsområder, samt hvordan resultatene kan anvendes praktisk i finanssektoren. Avslutningsvis oppsummerer vi i *kapittel 6* de viktigste funnene fra studien, trekker linjene tilbake til problemstillingen og forskningsspørsmålene, og gir anbefalinger for videre forskning og anvendelse av kunstig intelligens i aksjemarkedet.

Denne strukturen sikrer en logisk og sammenhengende fremstilling av studien, fra introduksjon til konklusjon, og gir leseren en klar forståelse av hvordan vi har brukt kunstig intelligens til å predikere aksjekursutviklingen til selskapene.

2 Teoretiske perspektiver

Formålet med teorikapitlet er å redegjøre for det teoretiske rammeverket som ligger til grunn for å besvare oppgavens problemstilling. Vi vil utforske de teoretiske grunnlagene som er essensielle for å forstå og vurdere den anvendte metodikken i denne studien, som sentrerer seg rundt bruk av kunstig intelligens AI for finansiell analyse. Vi begynner med en generell innføring i kunstig intelligens (AI), inkludert den historiske utvikling og de grunnleggende prinsippene som styrer denne teknologien. Deretter vil vi dykke dypere inn i de spesifikke underområdene av AI som er viktig for å forstå hvordan denne teknologien er bygd opp.

Videre vil vi se hvordan AI benyttes i finanssektoren i dag, samt vise til dens styrker og svakheter. Vi vil også gi en oversikt over aksjemarkedet i dag, hvor den eneste verdipapir børsen er Oslo Børs, og forklare teorien om markedseffisiens. Til slutt vil vi presentere faktormodeller som benyttes i finansielle analyser, med hovedfokus på Capital Asset Pricing Model (CAPM), som forklarer at en investor krever høyere avkastning når risikoen øker.

2.1 Kunstig intelligens

Kunstig intelligens er et vidt begrep som i dag brukes om mange former for algoritmer, maskinlæring, modeller og statistiske metoder (Direktoratet for forvaltning og IKT, u.å.). Ordet kunstig intelligens blir ofte forkortet til KI, men forkortelsen AI - for det engelske begrepet artificial intelligence, er den mest brukte. Vi vil i fortsettelsen av denne oppgaven benytte oss av begrepet AI. I den nasjonale strategien for AI er det definert som “Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål” (Direktoratet for forvaltning og IKT, u.å.). Likevel er det et ganske omfattende begrep, og det er ingen generell enighet om hva en definisjon bør inkludere. AI anses også å være et dynamisk konsept, hvor definisjonen endres i takt med utviklingen av teknologiske fremskritt innenfor feltet (Iriondo, 2018).

2.1.1 Kunstig intelligens i et historisk perspektiv

Opprinnelsen til AI kan spores tilbake til midten av 1950-tallet. I et historisk perspektiv, når man ser på etableringen av AI, går man tilbake til Dartmouth-konferansen i 1956 hvor ulike vitenskapsmenn og professorer som John McCarthy, Marvin Minsky og Allen Newell møttes for å diskutere hvordan maskiner kunne etterligne en form for menneskelig intelligens (Moor,

2006, s.87). På 1950-tallet var forskningen av AI preget av en såkalt symbolsk AI, som fokuserte på å representere kunnskap i form av ulike regler og symboler. Symbolsk AI handler om at datamaskiner manipulerer ulike symbolers tilnærming til hverandre, og dette kan brukes for å løse mindre komplekse problemstillinger som matematiske ligninger (Autoblock, 2024).

På 1960- og 1970-tallet ble det gjort betydelige fremskritt innenfor ekspertsystemer. Disse systemene er spesialiserte systemer som baseres på en slutningsmotor som programmeres slik at de inneholder informasjon for å løse problemstillinger. I 1972, på Stanford University, produserte de et ekspertsystem som spesialiserte seg innenfor blodsykdommer og reseptbelagt medisin. Ved å legge inn informasjon om pasienter kunne systemet diagnostisere og legge frem kunnskap på et høyt medisinsk nivå (Council of Europe, 2024).

Ekspertsystemene hadde et stort potensial for videre utvikling, men på grunn av manglende progresjon og ressurser falt interessen for videre forskning på 1980-tallet. Å programmere kunnskap på en slik måte krevde en betydelig innsats og inputs som krevde mellom 200 og 300 “regler” for å kunne bruke informasjonen. Dette førte så til en “black box”-effekt der maskinenes resonnering ikke var transparent. Utviklingen og vedlikehold av AI-systemene ble dermed svært problematisk og ulønnsomme, ettersom det fantes raskere, mindre komplekse og billigere alternativer tilgjengelig. Videre på 1990-tallet ble begrepet kunstig intelligens nærmest tabubelagt og begreper som “avansert databehandling” erstattet begrepet i det akademiske språket (Council of Europe, 2024).

Mot begynnelsen av det 21. århundre begynte maskinlæring å dominere AI-landskapet. Maskinlæring, spesielt dyp læring, førte til revolusjonerende fremskritt innen bilde-, tale- og naturlig språkbehandling. Det er to store faktorer som har bidratt til dette. For det første, en enorm økning i tilgang til data førte til at AI ble integrert i en rekke applikasjoner og bransjer, fra helsevesen og transport til finans og produksjon. Etter det ble den svært effektive ytelsen til grafikkort prosessorer oppdaget for å akselerere beregningen av læringsalgoritmer. Prosessen er svært repetitiv, og før 2010 kunne det ta uker å behandle hele datamengden. Grafikkortenes beregningskraft, som er i stand til å utføre mer enn tusen milliarder transaksjoner per sekund, har muliggjort betydelige fremskritt (Council of Europe, 2024).

2.1.2 GPT

Ifølge ChatGPT som er en spesialisert implementering av GPT, forklares GPT slik:

Generative Pre-trained Transformer (GPT) er en banebrytende språkmodell utviklet av OpenAI, som har fått bred anerkjennelse for sin evne til å generere sammenhengende og kontekstuelle relevante tekster. GPT er en type kunstig intelligens (AI) modell som faller under kategorien av nevralt nettverk, spesielt designet for naturlig språkprosessering (NLP) (OpenAI, 2024).

GPT er basert på en nevralt nettverksarkitektur kalt Transformer, som forskere lanserte i 2017 (Smith, 2024). Transformatorer er spesielt flinke til å behandle datastrenger, for eksempel tekst, ved hjelp av en mekanisme kalt «selvoppmerksomhet». Den mekanismen gjør at modellen kan fokusere på ulike deler av teksten avhengig av hvor viktige de er for oppgaven som modellen prøver å løse.

Språkmodellen er designet for å generere menneskelige samtaler og svare på en rekke emner. Modellen er trent på enorme mengder tekst og data fra internett for å forstå og produsere et menneskelignende naturlig språk, og brukes av millioner hver eneste dag (Tzur, 2023). GPT har større beregningskapasitet og et stort antall treningsparametere, med trening basert på forsterkende læring fra menneskelige tilbakemeldinger og en betydelig mengde manuelt merket data. Modellen kan håndtere kompliserte oppgaver som artikkelskriving, tekstoppsummering, språklig oversettelse, utvikling av dataprogrammer og andre avanserte oppgaver. Ifølge Tzur er de mest populære bruksområdene til GPT å oppsummere tekster, skrive e-poster og tekster. Videre er også programmering og koding blitt populære verktøy man kan bruke GPT til. Særlig programmer som RStudio og Python kan GPT brukes for å kode ulike oppgaver, og for å feilsøke eventuelle feilkoder fra eget arbeid.

Utviklingen av GPT var sterkt motivert av behovet for AI-språkmodeller som både var komplekse og allsidige. GPT ble først utviklet i 2018, men har siden utviklet seg fra den første prototypen GPT-1 til GPT-4 i 2023 (Wu et al., 2023, s. 1123). Det er tre betydelige vendepunkter i dens vekst. For det første bruker GPT Transformer-arkitektur, noe som muliggjør opprettelsen av effektive og utvidbare språkmodeller (Casella et al., 2023). For det andre viser utviklingen av GPT-serien potensialet til AI-språkmodeller i en rekke anvendelser, inkludert tekstgenerering, oversettelse og dataanalyse. For det tredje er GPT utgitt på grunnlag av tidligere generasjoner av GPT-produkter, med forbedringer innenfor forståelse, nøyaktighet og allsidighet (Sallam et al., 2023).

Forskjellen fra den første lanserte GPT-en, GPT-1, til GPT-4 har vært enorm. GPT-1, som ble lansert i 2018, hadde sine største styrker ved å kunne generere språk når den ble gitt en tekst eller kontekst. Modellen var basert på to datasett, nemlig Common Crawl og BookCorpus. Svakheterne til GPT-1 var at den ofte genererte samme tekst om og om igjen ved gitte spørsmål, og den slet med å gi fornuftige svar når man stilte nye spørsmål innenfor en tidligere gitt kontekst. I tillegg var den svært begrenset når det kom til svarene som ble gitt, og disse besto ofte av korte setninger (Ali et al., 2023).

GPT-4 representerer en markant forbedring over sine forgjengere, både i form av teknisk kapasitet og praktisk anvendbarhet. Med bedre forståelse av kontekst, økt nøyaktighet, redusert skjevhet, og større tilpasningsevne, har GPT-4 tatt et stort skritt fremover i utviklingen av kunstig intelligens for naturlig språkprosessering. Dette gjør GPT-4 i stand til å utføre en bredere rekke oppgaver mer effektivt og pålitelig enn noen tidligere GPT-modell (OpenAI, 2024).

I juni 2018 lanserte OpenAI GPT-1, som hadde 117 millioner parametere og omtrent 5GB med forhåndstreningsdata. Siden da har antallet parametre i hver generasjon av GPT-modeller økt eksplosivt, slik som vist i tabellen nedenfor (Guo & Zhang, 2023).

GPT-modell	Utgivelsestid	Antall parametere	Mengde forhåndstreningsdata
GPT-1	Juni 2018	117 millioner	Omtrent 5GB
GPT-2	Februar 2019	1,5 milliarder	40GB
GPT-3	Mai 2020	175 milliarder	45TB
GPT-4	November 2022	Hundre milliarder-nivå	Hundre TB-nivå

Tabell 1: Sammenligning mellom antall GPT-modellparametere og mengden forhåndstreningsdata (OpenAI, 2024).

2.2 Ulike felt innen AI

I dette delkapitlet vil vi gi en overordnet teoretisk innføring i forskjellige områder innen AI, inkludert maskinlæring, dyplæring, kunstige nevralt nettverk, automatisk talegjenkjenning og naturlig språkbehandling. GPT er bygd opp ved hjelp av maskinlæring, dyp læring og transformer-modellen, og vil derfor være viktig for oppgaven å få et innblikk i. Det er nødvendig å påpeke at dette er svært komplekse emner, hvor flere av dem bygger på hverandre og er sammenvevd.

2.2.1 Maskinlæring

Maskinlæring er en form for kunstig intelligens som fokuserer på å utvikle algoritmer og teknikker som lar datamaskiner lære og forbedre seg selv basert på erfaring og data, uten å være direkte programmert for bestemte oppgaver. Målet med maskinlæring er å utvikle modeller som kan generalisere fra data og gjøre forutsigelser eller ta beslutninger på nye, uventede data (Bi et al., 2018, s.2222). Dette kalles også for å trene en modell. En slik modell trenes på et treningssett for så å undersøke om modellen har lært noe, ved å teste det på et nytt datasett.

Maskinlæring kan deles opp i tre ulike kategorier: veiledet/overvåket læring, ikke-veiledet/overvåket læring og dyplæring/forsterket læring. Overvåket læring handler om at maskinlæringsalgoritmer som utleder mønstre mellom et sett med inputs (x) og et ønsket output (y). Disse mønstrene er så brukt for å kartlegge gitte inputs til det predikerte outputtet. For å lette overvåket læring, kreves det et merket datasett som inneholder matchede par av observerte inndata og deres tilknyttede utdata. Prosessen med å bruke dette datasettet til å se en sammenheng mellom inndata og utdata kalles å trene datasettet. Når datasettet er trent, kan algoritmen gjøre prediksjoner for utdataverdier basert på inndata som ikke er til stede i trenings datasettet (DeRose et al, 2024).

Et eksempel på overvåket læring er multiplikasjonsregresjon, der en regresjonsmodell bruker matchede data (X , Y) for å estimere parametere som definerer forholdet mellom de uavhengige variablene (X), og den avhengige variabelen (Y). Disse estimerte parameterne gjør det mulig å forutsi Y for en ny mengde X -verdier. Forskjellen mellom de forutsagte og faktiske Y -verdiene fungerer som en metrikk for å vurdere den prediktive ytelsen til regresjonsmodellen på ukjente data (DeRose et al, 2024).

Uovervåket læring er maskinlæring som ikke bruker merkede data. I en uovervåket læring har vi inndata (X -er) som brukes til analyse uten at noe utdata (Y) blir gitt. I uovervåket læring får ikke algoritmen merkede treningsdata og dermed forsøker algoritmen å forstå strukturer innenfor dataene selv. Dermed er uovervåket læring nyttig for å utforske nye datasett fordi det kan gi innsikt i større datasett som vil ta lang tid å analysere, eller er for komplekse (DeRose et al, 2024).

To viktige typer problemer innenfor uovervåket maskinlæring er å redusere dimensjonen av data og å sortere data inn i klynger, kjent som dimensjonsreduksjon. Dimensjonsreduksjon fokuserer på å redusere antall funksjoner samtidig som variasjonen over observasjonene bevares for å beholde informasjonen som er inneholdt i den variasjonen.

Dimensjonsreduksjon kan ha flere formål. Det kan brukes på data med et stort antall funksjoner for å produsere en lavere dimensjonal representasjon (dvs. med færre funksjoner), som for eksempel kan vises på en dataskjerm. Dimensjonsreduksjon brukes også i mange kvantitative investerings- og risikostyringsapplikasjoner der det er avgjørende å identifisere de mest forutsigende faktorene som ligger til grunn for bevegelsene i eiendomspriser (DeRose et al, 2024).

Klynging fokuserer på å sortere observasjoner inn i grupper (klynger) slik at observasjoner i samme klynge er mer lik hverandre enn de er til observasjoner i andre klynger. Grupper dannes basert på en rekke kriterier som kan være forhåndsdefinerte eller ikke (for eksempel antall grupper). Klynging har blitt brukt av kapitalforvaltere for å sortere selskaper inn i grupperinger drevet av data (f.eks. basert på deres finansielle data eller selskap karakteristika) i stedet for konvensjonelle grupperinger (f.eks. basert på sektorer eller land) (DeRose et al, 2024).

2.2.2 Dyp læring og forsterkende læring

Mer generelt innen kunstig intelligens, er ytterligere kategorier av maskinlæringsalgoritmer skilt ut. I dyp læring løser sofistikerte algoritmer komplekse oppgaver, som bildeklassifisering, ansiktsgjenkjenning, talegjenkjenning og naturlig språkbehandling. Dyp læring er basert på nevrale nettverk (NN-er), også kalt kunstige nevrale nettverk (ANN-er) - svært fleksible maskinlæringsalgoritmer som har blitt vellykket anvendt på en rekke overvåkede og usuperviserte oppgaver preget av store datasett, ikke-lineariteter og interaksjoner mellom egenskaper. I forsterkende læring lærer en datamaskin fra samhandling med seg selv eller data generert av den samme algoritmen. Prinsippene for dyp læring og forsterkende læring har blitt kombinert for å skape effektive algoritmer for å løse et spekter av svært komplekse problemer innen for eksempel robotikk, helsevesen og finans.

2.2.3 Dyplæring

Dyplæring er en læreprosess som brukes innenfor maskinlæring, og går ut på å “trene opp” såkalte “dype kunstige nevralt nettverk” (Tidemann, A. 2023). Prinsippet rundt dyplæring handler om at datamaskiner skal absorbere til seg kunnskap om noe den ikke kan fra før. Dyplæring drives av nevralt nettverk og bruker klassifiserte bilder, tekst eller lyder for å trekke konklusjoner. Ved bruk av dyplæring kan en datamaskin lære å utføre klassifiseringsoppgaver, og er en nøkkel for datasyn, behandling av naturlig språk og talegjenkjenning. I tillegg til utvikling av anbefalingssystemer innenfor film og musikk (Regjeringen, 2019).

Dyplæring har vært en kjent metode i flere tiår og har fundamentalt endret vår forståelse og bruk av kunstig intelligens. Allerede på 1940-tallet ble de første kunstige nevralt nettverkene utforsket, og interessen for disse teknologiene fikk en markant økning på 1980-tallet da forskere startet å utforske nettverk med flere lag. Til tross for lovende tidlige resultater, var utviklingen begrenset av den periodens datateknologi og tilgjengelige dataressurser (Tidemann, 2023).

Fra begynnelsen av 2000-tallet, med betydelige forbedringer i datamengder og databehandling, begynte dyplæring å få fotfeste innen forskning og industri. Et signifikant gjennombrudd skjedde i 2012, da et dypt nevralt nettverk oppnådde en ny rekord i ImageNet-konkurransen, en fremtredende bildegjenkjenning konkurranse som utfordrer deltakerne til å klassifisere bilder i en av tusen kategorier. Dette nettverket oppnådde en feilrate på kun 16 prosent, en signifikant forbedring sammenlignet med tidligere standarder på rundt 25 prosent, og markerte starten på en ny æra innen kunstig intelligens. Videreutvikling av teknologien reduserte raskt feilraten til bare noen få prosent (Tidemann, 2023).

I dag spiller dyplæring en ledende rolle i maskinell forståelse av komplekse dataformer som bilder, tekst og lyd. Den er en vesentlig komponent i mange digitale tjenester, inkludert de digitale assistentene vi bruker på våre smarttelefoner. Denne teknologiske fremgangen har ikke bare endret måten vi interagerer med teknologi på daglig basis, men også vår oppfatning av kunstig intelligens sitt potensial til å omforme samfunnet (Tidemann, 2023).

2.2.4 Transformer-modellen

En transformatormodell er en type dyplæringsmodell som ble introdusert i 2017. Disse modellene ble raskt veldig grunnleggende i naturlig språkbehandling (NLP), og har blitt brukt på et bredt spekter av oppgaver innen maskinlæring og kunstig intelligens (IBM, u.å.). Modellen ble først beskrevet i en artikkel fra 2017 kalt "Attention is All You Need" av Ashish Vaswani og et team fra Google Brain, samt forskere fra University of Toronto. Publiseringen av denne artikkelen markerte et gjennombrudd i feltet, ettersom transformatorer nå er blitt svært utbredt i applikasjoner som opplæring av store språkmodeller (LLM-er) (IBM, u.å.).

Disse modellene kan oversette tekst og tale i tilnærmet sanntid. For eksempel finnes det apper som gjør det mulig for turister å kommunisere med lokalbefolkningen på deres eget språk. Det hjelper forskere med å forstå DNA bedre og akselererer legemiddeldesign. I finans og sikkerhet brukes de til å oppdage avvik og forhindre svindel. Visjonstransformatorer anvendes også i datasynoppgaver (IBM, u.å.).

OpenAI's populære tekstgenereringsverktøy GPT benytter transformator arkitekturer for oppgaver som prediksjon, oppsummering og spørsmålssvar. Transformatorene lar modellen fokusere på de mest relevante delene av inndatateksten. "GPT" i verktøyets forskjellige versjoner (f.eks. GPT-3, GPT-4) står for "Generative Pre-trained Transformer." Tekstbaserte generative AI-verktøy som GPT drar nytte av transformatorer fordi de effektivt kan forutsi det neste ordet i en tekstsekvens basert på et omfattende datasett (IBM, u.å.).

2.2.5 Kunstige nevralt nettverk

Kunstige nevralt nettverk (ANN) er en teknologi basert på studiet av hjernen og nervesystemene til mennesker. ANN er et sentralt konsept innenfor kunstig intelligens og maskinlæring, og brukes primært for å analysere komplekse sammenhenger (Tidemann, 2019). Akkurat som den menneskelige hjernen, fungerer ANN ved at nevroner har en sammenheng med hverandre og prosesserer informasjon ved å identifisere fenomen, veie muligheter opp mot hverandre for så å komme til en konklusjon. Hvert nevralt nettverk består av lag med noder, eller kunstige nevroner som sender signaler til hverandre.

2.3 AI i finanssektoren

AI har revolusjonert hvordan forbrukere og selskaper får tilgang til og leder økonomien. Med estimerte kostnadsbesparelser på opptil 447 milliarder dollar innen 2023 fra bruk av AI-applikasjoner, utforsker blant annet banker stadig nye måter å integrere denne teknologien i sine tjenester (eMarketer, 2023). AI og maskinlæring spiller en stadig viktigere rolle i finanssektoren, med anvendelser som strekker seg fra chatbot-assistenter og svindeloppdagelse til automatisering av rutineoppgaver. Ifølge en rapport fra Insider Intelligence er så mange som 80% av bankene godt informert om de betydelige fordelene AI kan tilby (eMarketer, 2023).

Beslutningen om å implementere AI i finansinstitusjoner vil bli drevet fremover av teknologiske fremskritt, økende aksept blant brukere og endringer i regelverket. Banker som tar i bruk AI, har muligheten til å effektivisere monotone prosesser betydelig og forbedre kundeopplevelsen ved å tilby tilgang til kontoer og finansielle rådgivningstjenester døgnet rundt (eMarketer, 2023).

AI automatiserer manuelle oppgaver, som for eksempel dataregistrering, risikovurdering og kundeservice. Dette gir ansatte i ulike selskaper tid til å fokusere på andre nytteområder og mer komplekse oppgaver. Blant annet er det mange selskaper som har innført Robotic Process Automation (RPA) som bruker kunstig intelligens til å utføre diverse repeterende oppgaver, som for eksempel behandling av faktura og kontoavstemninger (Florkin, u.å.).

Ved hjelp av diverse AI-algoritmer kan man analysere store mengder med historiske data for å forsøke å predikere fremtidige hendelser. I sammenheng med finansmarkedet er dette svært tids- og kostnadsbesparende for vurdering av risiko og investeringsbeslutninger. Ved å bruke AI-baserte handelsalgoritmer vil man kunne analysere markedstrender og utføre kjøp og salg raskere enn man kan gjøre manuelt, noe som kan bidra til en bedre likviditet og prisutvikling (Florkin, u.å.).

2.3.1 Fordeler knyttet til AI i bank og finanssektoren

Fordelene ved å implementere AI i finanssektoren er betydelige, spesielt når det gjelder automatisering. Ved å automatisere arbeidet gjennom bruk av AI vil det bidra til enorme kostnadsbesparelser. AI i administrative oppgaver alene har potensial til å spare nordamerikanske banker opp mot 70 milliarder dollar innen 2025 (Insider Intelligence, 2021). Totalt sett er de potensielle kostnadsbesparelsene for banker som tar i bruk AI-applikasjoner

anslått til 447 milliarder dollar innen 2023, hvor front- og midtkontorene står for 416 milliarder dollar av dette beløpet (Insider Intelligence, 2021). Dette understreker AI-teknologiens monumentale påvirkning på finanssektoren, og viser hvor viktig det kan være å ta i bruk AI for å holde seg konkurransedyktig.

Implementeringen av AI i front- og administrative oppgaver kan revolusjonere finansnæringen ved å muliggjøre sømløse kundeinteraksjoner tilgjengelige 24/7, redusere behovet for repetitivt arbeid, senke antallet falske alarmer og menneskelige feil, samt redusere kostnader (Insider Intelligence, 2021).

Sømløse kundeinteraksjoner:

Med AI-drevne chatboter og virtuelle assistenter kan kundene få tilgang til kontoinformasjon, rådgivning og andre tjenester når som helst på dagen eller natten. Dette gir en bedre kundeopplevelse ved å gi raskere svar på spørsmål, veiledning om finansielle produkter, og øyeblikkelig bistand med transaksjoner. Virtuelle assistenter kan også bidra med personlig rådgivning ved å analysere kundenes økonomiske situasjon og tilby tilpassede investeringsstrategier (Insider Intelligence, 2021).

Redusert repetitivt arbeid:

Ved å automatisere rutineoppgaver som datainntasting, transaksjonsbehandling og rapportgenerering, kan ansatte i finanssektoren frigjøre tid til mer komplekse og verdiskapende oppgaver. For eksempel kan automatiserte prosesser effektivt håndtere bankoverføringer, kredittsjekker og kundeoppslag, noe som reduserer tiden brukt på disse oppgavene betydelig (Insider Intelligence, 2021).

Redusert antall falske alarmer og menneskelige feil:

AI-modeller kan brukes til å oppdage uregelmessigheter i transaksjoner og flagge potensielt svindelaktige aktiviteter mer nøyaktig enn tradisjonelle systemer. Dette reduserer antallet falske alarmer og sikrer at teamene kan fokusere på de tilfellene som virkelig krever oppmerksomhet. Videre kan AI-teknologier som maskinlæring bidra til å forbedre kvaliteten på dataanalyse, noe som fører til mer nøyaktige beslutninger og færre feil (Insider Intelligence, 2021).

2.4 utfordringer knyttet til AI

I dette delkapitlet undersøker vi utfordringene knyttet til bruk av kunstig intelligens (AI) i finanssektoren. Vi vil fokusere på juridiske og etiske problemstillinger, samt datapersonvern og sikkerhet.

2.4.1 Juridiske og etiske problemstillinger

Bruken av AI i finanssektoren bringer med seg ulike etiske og juridiske vurderinger som er avgjørende å adressere for å sikre en ansvarlig bruk av den nye teknologien. De primære etiske og juridiske utfordringene knyttet til bruk av AI i finanssektoren ligner på de som oppstår i andre industrier, men de er forsterket av behovet for å beskytte mot finansielle risikoer som kan påvirke både forbrukeres og bankers økonomiske stabilitet.

Datasekk brukt til trening av GPT kan inneholde sensitive finansielle informasjon eller være beskyttet av immaterielle rettigheter, noe som kan føre til brudd på opphavsrett, patentrett og lover og forskrifter om datavern. Hvis finansinstitusjoner stoler på ulovlig informasjon når de tar beslutninger, kan de også bli straffet (Wang, Li, Yang, et al., 2023).

Det finnes også en viss moralsk risiko med GPT på grunn av dens evne til å generere menneskelignende språk og skrivning. Denne evnen kan brukes til identitetsforfalskning, svindel og villedning, phishing-angrep og spredning av uønsket informasjon (Ray, 2023). I forsikringsindustrien, på grunn av GPT's "black-box teori", kan løsningene som tilbys av GPT bli påvirket av faktorer som rase, kjønn og bosted. Derfor kan introduksjonen av GPT føre til fordommer mot de forsikrede (Deshpande & Joshi, 2022). Teorien om "black box" vil vi presentert i delkapittel 2.4.3.

2.4.2 Datapersonvern og sikkerhet i finanssektoren

Data har stor verdi, og det er en viktig årsak til at ditt online personvern er truet (Kaspersky, u.å.). All informasjon som lastes opp på internett, kan brukes med dårlige hensikter. Data som lastes opp til AI blir lagret i databaser, og man må derfor være forsiktig med å laste opp kritisk informasjon.

AI-teknologiens innflytelse i banksektoren har reist flere etiske og juridiske spørsmål, spesielt rundt personvern, sikkerhet, åpenhet og algoritmisk bias. I forhold til personvern skaper AI-systemene utfordringer knyttet til behandling og lagring av personlige data uten passende godkjenninger. Finanssektoren involverer store mengder sensitiv informasjon, som for

eksempel makroøkonomiske indikatorer, markedsdata, selskapsfinansielle data og personlige transaksjonsdata. (Wang, Li, Yang, et al., 2023). Både personvernet og stabiliteten i finansmarkedene kan bli alvorlig truet av datalekkasjer.

I bankindustrien kan det være nødvendig å behandle store mengder personlig sensitiv informasjon når bankorganisasjoner bruker GPT til å analysere kundeinformasjon. Hvis de relaterte dataene samles inn og brukes feil, vil muligheten for potensielle trusler mot kundens personvern øke, noe som påvirker bankens omdømme og kundetilfredshet. Med tanke på sikkerhet er det bekymringer for AI-systemenes sårbarhet mot ondsinnede angrep som kan forstyrre operasjoner og forårsake økonomiske tap (Deshpande & Joshi, 2022).

En annen bekymring er mangel på åpenhet, hvor det kan være vanskelig å identifisere kildene til data og prosessene bak AI-beslutninger. Den kanskje mest alvorlige utfordringen er imidlertid potensialet for algoritmisk bias, hvor AI-systemer kan lære og reprodusere eksisterende skjevheter i treningsdata, noe som kan føre til urettferdige eller diskriminerende beslutninger (Deshpande & Joshi, 2022).

Disse utfordringene indikerer at AI-baserte beslutninger kan være skjevfordelte, unøyaktige, eller i noen tilfeller, diskriminerende. I tillegg kan svake sikkerhetstiltak åpne for misbruk av teknologien, som for eksempel hvitvasking og innsidehandel, som kan foregå raskt og være vanskelig å oppdage på grunn av AI-teknologiens behandlingshastighet. Derfor er det avgjørende at banker implementerer nøye utformede retningslinjer, prosedyrer og protokoller for å håndtere bruk av AI, samtidig som de minimerer tilhørende etiske og juridiske risikoer (Deshpande & Joshi, 2022).

2.4.3 Black box

Når man skal diskutere utfordringer og ulemper med bruk av kunstig intelligens, kommer man ikke foruten begrepet *black box*, som på norsk betyr "sort boks". Dette begrepet refererer til AI-systemer med interne funksjoner som er usynlig for brukeren. Black box fungerer som en slags metafor for de utilgjengelige beregningene som kan oppstå når man generer data fra et AI-nettverk. Ved bruk av AI kan man generere data som ikke alltid er korrekt, ved et black box tilfelle har vi manglende forståelse for beslutninger som AI-modellen foretar seg, og det kan være vanskelig å feilsøke problemet (Sonne, 2017).

2.5 Finansanalyse og prediksjon

I dette kapittelet vil vi utforske ulike metoder for finansanalyse. I det første delkapitlet starter vi med å presentere teknisk analyse, som er en tradisjonell metode for å studere markedsbevegelser gjennom bruk av grafer for å identifisere trender i markedet. Deretter ser vi på fundamental analyse, som vurderer de økonomiske kreftene bak tilbud og etterspørsel av aksjer for å bestemme den iboende verdi. Til slutt ser vi på hvordan maskinlæring brukes i finans for å håndtere store datamengder, avdekke mønstre og forbedre prediksjonsnøyaktigheten.

2.5.1 Teknisk analyse

Historisk sett er teknisk analyse det mest brukte verktøyet for å predikere bevegelser på aksjemarkedet. Det er en svært tradisjonell metode, og ble definert av John J. Murphy som studien av markedsbevegelser, primært gjort ved bruk av charter for å predikere fremtidige pristrender (Murphy, 1999, s. 32). Det er ifølge Murphy tre premisser for hvor en teknisk analyse kan iverksettes, dette er hvor alt som kan påvirke kursen på en aksje, faktisk reflekteres i prisen. Det andre er at prisen beveger seg i ulike trender, altså poenget med å chartre prisen på en aksje er å identifisere trender i tidlige stadier av kursendringer. Det tredje prinsippet er at historien gjentar seg selv. Det man tar med seg fra det tredje prinsippet er at nøkkelen til å predikere fremtiden handler om hva som har skjedd i fortiden, og at fremtiden ofte bare er en repetisjon av hva som har tidligere (Murphy, 1999, s. 34-36).

2.5.2 Fundamental analyse

Der hvor teknisk analyse av aksjer omhandler studier av markedshandlinger, ser man innenfor fundamental analyse på de økonomiske kreftene til tilbud og etterspørsel av aksjer. Den fundamentale tilnærmingen ser på alle relevante faktorer som påvirker priser i markedet, for å bestemme den iboende prisen (Murphy, 1999, s. 36). Hvis den iboende prisen er under markedsprisen, er markedet overpriset og man bør selge aksjene sine. Hvis den iboende prisen derimot er over markedspris, er markedet underpriset og man bør kjøpe aksjer ifølge den fundamentale analysen (Murphy, 1999, s. 36).

Likheten mellom en fundamental og en teknisk analyse handler om at analytikerne ønsker å løse samme problem, altså å fastslå hvilken retning aksjekursene vil bevege seg. Forskjellen mellom tilnærmingene er hvordan man velger å angripe problemstillingen. Murphy siterer

med “The fundamentalist studies the cause of market movement, while the technician studies the effect” (Murphy, 1999, s. 36). For en som driver teknisk analyse er kun effekten det som er relevant for studien, mens for fundamentalisten er det viktig å vite årsaken til effekten eller prisendringen.

2.5.3 Maskinlæring i finans

Innenfor finansfeltet er maskinlæringssystemer til stor hjelp for å forstå enorme datamengder og avdekke viktige mønstre i dem. Dette er nyttig informasjon som deretter brukes til å forbedre forretningsprosesser, ta informerte beslutninger og hjelpe til med prediksjonsoppgaver (Coursera Staff, 2024). Maskinlæring-teknologien brukes ofte i finansbransjen for å forsøke å forutsi aksjekurser og påvirke beslutninger. Maskinlæring er særlig nyttig ved å bruke store datasett for å komme med spådommer om fremtiden. maskinlæringsteknologi muliggjør to typer handel. Den første er algoritmisk handel. Det betyr å identifisere mønstre og utvikle handelsstrategier med hastighet og nøyaktighet. Den andre typen er High-Frequency trading (HFT), som handler om å identifisere handelsmuligheter og utføre kjøp eller salg i høye hastigheter.

2.6 Markedet

Den eneste verdipapirbørsen i Norge er Oslo Børs. Dette er en regulert markedsplass for omsetning av verdipapirer, og tilbyr handel av aksjer, egenkapitalinstrumenter, derivater og renteprodukter. Oslo Børs eies i dag av Euronext (Gram, 2024).

I vår oppgave sammenligner vi våre funn opp mot hovedindeksen i Norge, Oslo Børs. En indeks er en sammensatt statistikk, et mål på endringer i en representativ gruppe av individuelle datapunkter eller med andre ord et sammensatt mål som samler flere indikatorer (Nordnet, u.å.). Oslo Børs hovedindeks (OSEBX) består av 69 selskaper, der Equinor helt klart er det største selskapet. I denne oppgaven skal vi sammenligne eventuell meravkastning mot OSEBX.

2.6.1 Markedseffisiens

Markedseffisiens er en hypotese som antyder at markeder, spesielt finansmarkeder, har evnen til raskt å reflektere den økonomiske påvirkningen av ny informasjon. I et effisient marked vil prisene raskt justeres til å gjenspeile den underliggende verdien av eiendelen, basert på

tilgjengelig kunnskap. Dette innebærer at det er vanskelig for aktører å overgå markedet, det vil si å oppnå høyere avkastning enn gjennomsnittet av markedet. Markedseffisiens er et sentralt konsept innen finansiell teori som forklarer hvordan finansmarkeder opererer (Gårseth-Nesbakk, 2023). Ifølge teorien om en effisient markedssetting vil ikke AI kunne utnytte tilgjengelig informasjon for å oppnå en betydelig økonomisk fordel, da prisene allerede reflekterer all kjent informasjon.

Det ligger en del forutsetninger til grunn for at et marked skal være effisient. Markedet må ha tilstrekkelig stor konkurranse blant aktører i finansmarkedet, som er villig til å gjøre vurderinger og handle basert på ny informasjon. Det vil si at det trengs tilstrekkelig med kjøpere og selgere til enhver tid. Markedseffisiens legger også til grunn at markedsaktørene gjør rasjonelle valg og at de jevnt over er kompetente. Samtidig må de være motivert til å oppnå en høyere avkastning ved å forsøke å utnytte antatt feilprising eller ved å handle tidlig på nyheter som kan ha betydning for finansielle produkter (Gårseth-Nesbakk, 2023).

Samlet sett fører dette til at den nye informasjonens effekt raskt blir innarbeidet i vurderingen av verdien til produkter som handles på finansmarkeder, for eksempel aksjer, obligasjoner og valuta. Selv om hver enkelt aktør ikke har full forståelse eller treffer perfekt med sine vurderinger, vil andre markedsaktører enten korrigere dette eller utnytte forbigående feilprising raskt. Dermed vil den samlede vurderingen og handlingene til markedsaktørene bidra til en korrekt prising i markedet (Gårseth-Nesbakk, 2023).

Virkingen av markedseffisiens er at det over tid ikke skal være mulig for enkeltaktører å slå effisiente markeder, gitt at man tar samme grad av risiko. Da et effisient marked skal prise verdipapirene "riktig", vil det ikke ha noen hensikt å lete etter investeringsobjekter som er feilpriset, og det vil dermed ikke kunne gi noe ekstraordinær avkastning. Resultatene man oppnår i markedet skal derfor reflektere markedet generelt (Gårseth-Nesbakk, 2023).

Et kjent bilde fra boken "A Random Walk Down Wall Street", hvor forfatteren Burton Malkiel hevder at en ape med bind for øynene som kaster dartpiler på en finansavis vil ha like gode muligheter til å plukke ut en konkurransedyktig aksjeportefølje som det ekspertene klarer. Poenget til Malkiel er at det ikke er umulig å slå markedet noen ganger, men at det kun vil være mulig på kort sikt, og dermed sterkt preget av flaks og tilfeldigheter, og derfor ikke vil la seg gjøre over tid (Gårseth-Nesbakk, 2023).

Hypotesen rundt et effisient marked oppsto på 1960-tallet etter at Paul Samuelson og Eugene Fama gjennomførte forskning på temaet. Samuelson og Fama inkluderte i sin studie at det

finnes tre typer effisiens i finansmarkedet: sterk form, semi-sterk form og svak form (Fama 1970).

Sterk effisiens av effisiens innebærer at all informasjon i markedet er offentlig tilgjengelig, slik at selv innsideinformasjon ikke gir investorer noen fordel. Dette betyr at det er umulig å oppnå unormal avkastning uansett hvor mye forskningsinnsats eller tilgang til informasjon man har. Denne typen effisiens betraktes som ekstrem og er sjelden i praksis (Dalen, 2014).

Semi-sterk effisiens innebærer at dagens aksjepris reflekterer all tilgjengelig offentlig informasjon. Det betyr at det ikke er mulig å oppnå en avkastning som er høyere enn markedet, gjennom verken fundamental eller teknisk analyse. Eneste mulighet for investorer til å oppnå en fordel for å oppnå unormal avkastning er gjennom informasjon som ikke er offentlig tilgjengelig (Dalen, 2014).

Svak effisiens er historiske aksjekurser reflektert i dagens priser. Derfor er teknisk analyse ikke effektiv for å forutsi aksjekursen eller overgå markedet. All historisk informasjon for å forutsi fremtidige priser er allerede inkludert av investorer. Samtidig indikerer teorien om svak effisiens at fundamental analyse kan brukes til å vurdere om selskaper er under- eller overvurdert. Fundamental analyse innebærer å evaluere selskapets økonomiske prestasjoner, inkludert avkastning, omsetning, kontantstrøm osv (Dalen, 2014).

2.6.2 Random Walks

Random Walks teorien antyder at endringer i aksjepriser skjer tilfeldig, og at det er umulig å forutsi retningen på disse prisendringene. Denne teorien støtter markedseffisiens, hvor all tilgjengelig informasjon som kan brukes til å forutsi aksjepriser allerede er reflektert i dagens priser. Dermed vil Random Walks naturlig oppstå som resultatet av priser som kontinuerlig gjenspeiler dagens informasjon. (Bodie et al., 2009)

2.6.3 Moderne porteføljeteori

Markowitz (1952) sin moderne porteføljeteori kom med flere forutsetninger hvor en av dem omhandler investorens villighet til å ta risiko. Teorien er en av de mest banebrytende vitenskapelige publikasjonene innen finansiell økonomi, og markerer typisk begynnelsen på det som kalles moderne porteføljeteori. Den omhandler investorens villighet til å ta risiko og hvordan investorer setter sammen sine porteføljer, som avhenger av risikoaversjon.

Risikoaverse investorer ønsker helst å unngå risiko.

Markowitz (1952) er en av de mest banebrytende vitenskapelige publikasjonene innen finansiell økonomi og markerer typisk begynnelsen på det som kalles moderne porteføljeteori. Artikkelen tar for seg at hvordan investorer setter sammen sine porteføljer, avhenger av risikoaversjon. Den presenterer et viktig bidrag til forståelsen av forholdet mellom valg av spesifikke finansielle eiendeler, deres avkastning, og den generelle porteføljerisikoen. Gitt antagelsen om at en investor er rasjonell og at risiko er noe som bør unngås, blir et teoretisk rammeverk for å maksimere avkastningen for et bestemt risikonivå, eller minimere risikoen for en gitt avkastning.

2.7 Faktormodeller

Faktormodeller er finansielle modeller som bruker fundamentale, tekniske, makroøkonomiske og lignende faktorer for å definere risikopremier og forventet avkastning (Srivastava, 2020). Modellene er lineære og eksempler på faktormodeller kan er CAPM, Fama-French eller Carharts firefaktormodell. Vi forklarer i neste kapittel hvilke faktorer som er brukt i vår analysedel.

2.7.1 Markedsmodellen

En av de mest grunnleggende av faktormodellene er Capital Asset Pricing Model (CAPM), og ble utviklet av Treynor (1962), Sharpe (1964), Lintner (1965) og Mossin (1966). Modellen er utviklet i et forsøk på å gi en teoretisk forklaring på hvorfor det finnes risikopremier, og beskriver sammenhengen mellom forventet avkastning og risiko (Håkon R., u.å.): En investor krever høyere avkastning når risikoen øker, og denne lineære sammenhengen kan skrives slik:

$$r_i - r_f = \alpha_i + \beta[r_M - r_f] + e_i$$

Variablene forklares på følgende måte:

$r_i - r_f$ = Avkastningen på eiendel i

α = Krysningen av regresjonslinjen

$[r_M - r_f]$ = Avkastningen på markedsporteføljen minus risikofri rente

SMB = Avkastning på verdivektet faktortportefølje for størrelse

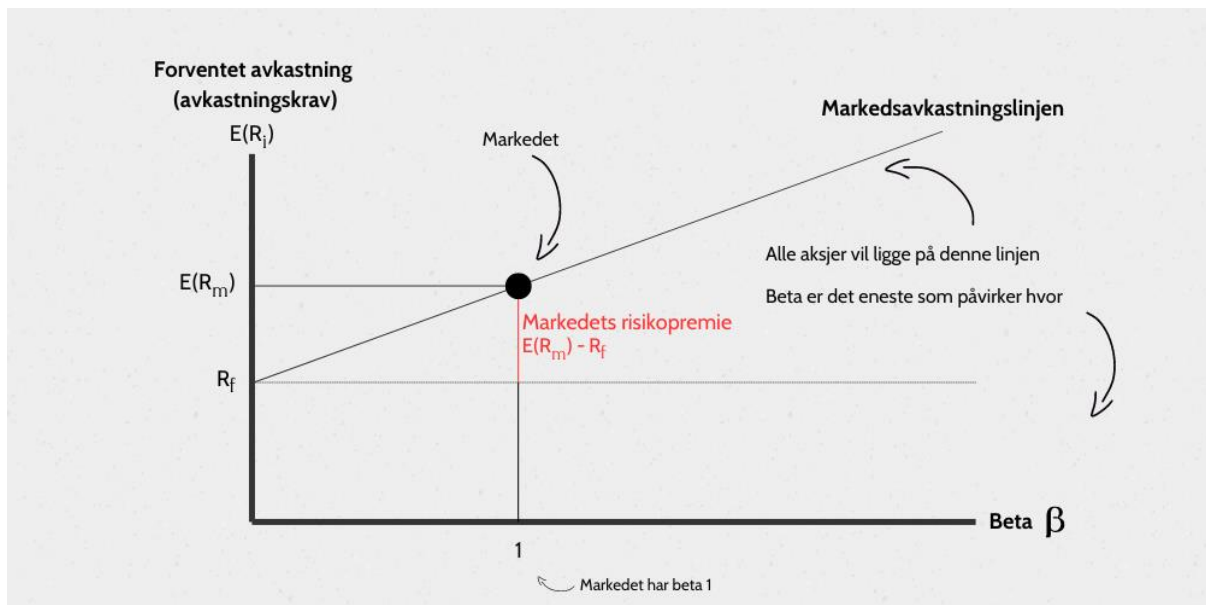
HML = Avkastning på verdivektet faktortportefølje for verdi

ei = Feilledet

β, s, h , = Koeffisienten verdiene på de tre uavhengige variablene

Kapitalverdimodellen forteller at den forventede avkastningen er lik risikofri rente pluss markedets risikopremie som er multiplisert med beta, hvor beta er et mål på aksjens systematiske risiko (Håkon R., u.å.).

Siden det er en likevektsmodell antar man at alle investorer handler rasjonelt, er risikoaverse og har tilgang til samme informasjon. Dermed antar man at prisene i samfunnet reflekterer all tilgjengelig informasjon (Håkon R., u.å.).



Figur 1: Markedsavkastningslinjen

Denne figuren illustrerer Kapitalverdimodellen (CAPM) som viser forholdet mellom forventet avkastning og beta (β) for en investering. X-aksen representerer beta, som måler markedsrisiko, mens Y-aksen viser forventet avkastning. Risikofri rente (R_f) er den avkastningen som kan oppnås uten risiko. Markedets forventede avkastning ($E(R_m)$) ligger over R_f og forskjellen er markedets risikopremie ($E(R_m) - R_f$). Markedsavkastningslinjen (SML) viser at alle aksjer skal ligge på denne linjen hvis markedet er i likevekt. Jo høyere beta, desto høyere forventet avkastning, gitt den økte risikoen.

CAPM er basert på sterke antagelser, og blir ofte beskrevet som et "empirisk nederlag" av forskere som Fama og French i 2015. Modellen kan ikke forklare observerte mønstre og

anomalier i aksjeavkastninger gjennom sine enkle lineære sammenhenger. Det har blitt utviklet to alternative modeller for å adressere disse begrensningene. Den vi skal gå litt inn på er Arbitrage Pricing Theory (APT) av Ross i 1976. APT antar at avkastningen er en lineær kombinasjon av flere systematiske risikofaktorer pluss en aksje-spesifikk avkastning. Ross påpekte også at idiosynkratisk risiko kan diversifiseres bort ved å investere i en portefølje i stedet for enkeltaksjer, slik at avkastningen hovedsakelig reflekterer systematisk risiko hvis porteføljen er tilstrekkelig diversifisert.

2.7.2 Fama og French sin trefaktormodell

Fama og French sin trefaktormodell er en prisingsmodell fra 1992 som bygger videre på kapitalverdimodellen (CAPM). Den utvider CAPM ved å inkludere faktorer for størrelsesrisiko og verdirisiko, i tillegg til markedsrisikofaktoren. Modellen tar hensyn til at verdiaksjer og små aksjer ofte gir bedre avkastning enn markedet generelt. Ved å legge til disse to faktorene, gir modellen en bedre justering for denne overprestasjonen, noe som gjør den til et mer nøyaktig verktøy for å evaluere fondsforvalteres prestasjoner (Hayes, 2024).

Fama og French-modellen inkluderer tre faktorer: størrelsen på selskapene, bokførte verdier, og meravkastning i forhold til markedet. Dette betyr at modellen bruker følgende: "liten minus stor" (SMB), "høyt minus lavt" (HML), og porteføljens avkastning minus den risikofrie avkastningen. SMB viser til at mindre selskaper med lav markedsverdi ofte gir høyere avkastning. HML viser til verdiaksjer, som er aksjer med høye bokførte verdier, som generelt gir bedre avkastning enn markedet (Hayes, 2024).

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2SMB_t + \beta_3HML_t + \varepsilon_{it}$$

Forklaring til variablene ser slik ut:

- R_{it} = total avkastning på en aksje eller portefølje ved tid t
- R_{ft} = risikofri avkastning ved tid t
- R_{Mt} = total markedsavkastning ved tid t
- $R_{it} - R_{ft}$ = forventet overskytende avkastning
- $R_{Mt} - R_{ft}$ = overskytende avkastning på markedsporteføljen (indeks)
- SMB_t = størrelse-premie (små minus store selskaper)
- HML_t = verdi-premie (høy minus lav verdi)

- $\beta_1, \beta_2, \beta_3$ = faktorkoeffisienter

2.7.3 Carhart fire-faktormodell

Carhart sin fire-faktormodell er en utvidelse av den mer kjente Fama-French tre-faktor modellen, og ble foreslått av Mark Carhart mot slutten av 1990-tallet (Thuen, 2021). Da Fama-French tre faktor modellen ble utviklet, hevdet den at de fleste aksjemarkedsavkastninger kunne forklares av tre ulike faktorer: risiko, pris og størrelsen på tilhørende selskap. I Carharts fire-faktormodell la han til en momentumfaktor for prisendringer på aksjer og fond. Denne faktoren forklarer Carhart med at fond og aksjer som har prestert godt de siste 12 månedene vil fortsette å prestere godt, og motsatt for fond og aksjer som har prestert dårlig de siste 12 månedene vil fortsette den negative trenden (Carhart, 1997). Teorien om at firefaktormodellen økte nøyaktigheten i måling av porteføljer, baserte Carhart på studien sin hvor han analyserte nesten 2000 fond på S&P.

I regresjonsform ser Carhart sin firefaktormodell slik ut:

$$r_i - r_f = \alpha + \beta_1 (r_m - r_f) + \beta_{2i} (\text{SMB}) + \beta_{3i} (\text{HML}) + \beta_{4i} (\text{WML}) + \epsilon_i$$

Forklaring til variablene ser slik ut:

r_i = Avkastningen til eiendel i

r_f = risikofri rente på statsobligasjoner

α = Krysningen på regresjonslinjen

r_m = Avkastning på markedsporteføljen

(SMB) = Avkastning på størrelsesfaktor

(HML) = Avkastning på BE/ME faktoren

(WML) = Avkastning på momentumfaktor

ϵ_i = Residualene til regresjonen

β_1 - β_2 - β_3 - β_4 = Betaverdien på de uavhengige variablene $r_m - r_f$, SMB, HML og WML

2.8 Faktorforklaringer

Faktorene som blir presentert i analysedelen av oppgaven er rm , SMB, HML, LIQ og MOM, hvor SMB, HML og rm er utbredt fra Fama-Frenchs trefaktormodell, stammer LIQ fra Pastor-Stambaugh modellen, og MOM fra Mark Carharts firefaktormodell som er en utvidelse fra Fama-French. LIQ er en likviditetsfaktor som måler likviditetsrisikoen. Denne beregnes som avkastningen på aksjer med lav likviditet minus avkastningen på aksjer med høy likviditet. Disse 5 faktorene er de som benyttes i TITLONs database og er derfor de vi har størst fokus på i denne undersøkelsen.

2.8.1 SMB (Small Minus Big)

SMB (Small Minus Big) er faktoren som representerer forskjellen i avkastning mellom små og store selskaper i markedet. Teorien om Small Minus Big handler om at positiv SMB gir mindre selskaper en bedre avkastning på markedet enn større selskaper over lengre tid, mens en negativ SMB indikerer det motsatte. SMB blir sammen med andre faktorer brukt til å forklare avkastningen på porteføljer, og blir også kalt for størrelseseffekten, hvor størrelsen baserer seg på de ulike selskaperenes markedsverdi (Kenton, 2020).

2.8.2 HML (High Minus Low)

HML (High Minus Low) refererer til selskaper med en høy bokført verdi kontra markedsverdien, disse blir også omtalt som "value stocks", og motsatt for selskaper med en lav bokført verdi kontra markedsverdien. Selskaper med en lav bokført verdi i forhold til markedsverdien er ofte omtalt som "growth stocks". Tidligere forskning har vist at porteføljer med en større mengde med value stocks presterer bedre enn de porteføljene med flere growth stocks på lengre sikt (Kenton, 2020).

2.8.3 LIQ (Liquidity)

LIQ (Liquidity) representerer forskjellen i avkastning mellom porteføljer av likvide aksjer og porteføljer av illikvide aksjer. Likviditeten beskriver til hvor stor grad en aksje eller et verdipapir kan handles uten å påvirke prisen (Pontiff et al, 2019). Likvide aksjer er enklere å kjøpe og selge uten å påvirke prisen, mens de illikvide aksjene kan være vanskeligere å omsette. Målene på LIQ forklares slik, en positiv LIQ indikerer at likvide aksjer har gitt en høyere omsetning enn illikvide, om motsatt for en negativ LIQ (Pontiff & Singla, 2019).

2.8.4 MOM (Momentum)

MOM (Momentum)-faktoren måler forskjeller i avkastning mellom porteføljer bestående av aksjer med en stigende trend, og porteføljer av aksjer med en fallende trend. En positiv MOM indikerer at aksjer med høy tidligere avkastning vil fortsette å gi økt avkastning over en tidsperiode, mens en negativ MOM indikerer det motsatte (Fajasy, 2023).

3. Forskningsdesign og metode

I dette kapitlet presenteres den metodiske tilnærmingen som er valgt for å besvare vår problemstilling og forskningsspørsmål. Metoden demonstrerer hvordan data ble samlet inn, og hvordan den skal analyseres og tolkes. Innledningsvis presenteres forarbeidet som ble gjort for å velge riktig metode. Deretter presenterer vi en steg-for-steg-gjennomgang av forskningsdesignet, etterfulgt av en begrunnelse for valg av forskningsmetode og en detaljert beskrivelse av studietypen. Videre beskriver vi prosessen for datainnsamling og databehandlingen, samt en oversikt over datamaterialet. Metoden demonstrerer hvordan data ble samlet inn og hvordan de skal analyseres og tolkes. Til slutt vil vi gjennomgå en evaluering av studiens kvalitet.

3.1 Forarbeid av metode

Først vil vi presentere forarbeidet som ble gjort for å sikre at metodedelen var gjennomførbar. Det som var viktig i starten var å undersøke om GPT kunne gi en karakter basert på innledningen, nøkkeltallene og regnskapet fra årsrapporter. For å teste dette, analyserte vi manuelt årsrapportene fra de 10 største selskapene på Oslo Børs. Vi fant ut at det var mulig å få en karakter fra GPT ved å la den analysere årsrapportene, men prosessen var svært tidkrevende. For å få mest mulig valide resultater, fant vi ut at det måtte analyseres årsrapporter fra flere hundre selskaper. Dette er nødvendig for å se om karakterene i gjennomsnitt har en reell sammenheng med selskapenes prestasjoner, eller om de var tilfeldige. Med hjelp fra vår veileder er det utviklet en kode som kan automatisere denne prosessen. Denne automatiseringen danner grunnlaget for resten av metodedelen i vår studie.

3.1.2 Fremgangsmåte

Først ble et omfattende datasett, bestående av 328 årsrapporter fra selskaper opererende i ulike bransjer fra Oslo Børs, samlet inn fra databasen i TITLON. Årsrapporter er store PDF filer med mye informasjon uten noen form for fast struktur. Å hente ut tekst med python var håndterbart, men for å få tabeller med høy kvalitet har veileder utviklet en kode med følgende prompts:

1. Først ble prompten «is_mainly_text» brukt for å avgjøre om innholdet hovedsakelig er tekst eller tall. Hvis det er tekst, ignoreres det.

2. Deretter prøver man å hente linjene før tabellen for å finne en eventuell overskrift. Avhengig av om disse linjene finnes eller ikke, brukes det enten prompten «fix_table_header» eller «fix_table_simple».
3. Hvis linjene ble funnet, forsøkes det deretter å bestemme hva overskriften kan ha vært ved hjelp av prompten «decide_table_caption».
4. PDF'en deles så inn i avsnitt, hvor hvert avsnitt blir analysert med promptet “summarize”. Oppsummeringene av de ulike avsnittene blir så satt sammen til en lang tekst som deretter analyseres med promptet “conclude”

Neste steg er at alle de tidligere responsene samles i et dokument som gis til GPT. Deretter brukes prompten «analyze_all» for å be GPT gi en karakter fra 0 til 10 samt en begrunnelse for denne vurderingen. Når det så kjøres API-spørring brukes GPT til å behandle og analysere dataen.

Variabler vi får ut:

1. Avkastningstall for ett og to år fra 1. juni etter at rapporten er publisert. Disse inkluderer alfa og beta for aksjene i henhold til en firefaktormodell (SMB, HML, LIQ, MOM), med p-verdier for å vurdere signifikans.
2. En anbefaling fra GPT på en skala fra 0 til 10, samt en begrunnelse for denne vurderingen.

Dette er variablene som utgjør grunnlaget for datasettet vårt, og som brukes for å gjøre beregningene i vår analysedel. Basert på disse variablene og informasjonen lastet ned fra TITLON, har vi regnet oss frem til avkastningen for alle selskapene.

3.2 Forskningsdesign

Valg av forskningsdesign for vår masteroppgave er en eksperimentell tilnærming hvor vi anvender AI, spesifikt GPT, for å gi karakter til selskapene basert på hvor lovende fremtidsutsiktene til aksjekursen er for norske børsnoterte selskaper, basert på å analysere årsrapportene. For å utforske potensiale til kunstig intelligens gjennom å analysere årsrapportene, har vi samlet et omfattende datasett bestående av årsrapporter fra ulike børsnoterte selskaper i Norge. Selskapene vi har valgt til datasettet vårt er fra et bredt spekter og representerer alle sektorene. Det er 212 årsrapporter fra 2010, og 116 fra 2011. Dataen vi har samlet inn er hentet og lastet ned fra databasen TITLON. Ettersom årsrapportene er fra

tidligere år, vil vi kunne sammenligne avkastningen opp mot hovedindeksen på Oslo Børs OSEBX.

3.2.1 Valg av forskningsmetode

Vi har i vår studie valgt kvantitativ tilnærming. Kvantitative metode er forskningsmetoder som brukes ved innsamling og analyse av kvantitative data. Dette er data som foreligger i form av tall eller andre mengdetermer (Grønmo, 2024). Kvantitativ metode er best egnet for vår oppgave da vi analyserer store mengder numeriske data og trenger statistiske verktøy for å identifisere mønstre og sammenhenger. Dette er viktig for å kunne validere om GPT kan gi pålitelige karakterer og prediksjoner av karakterer basert på innhold i årsrapportene. Ved å automatisere prosessen, vil vi kunne håndtere store datasett effektivt og minimere subjektiv tolkning, noe som øker påliteligheten og gyldigheten av våre funn. Denne kvantitative tilnærmingen gjør det også mulig å generalisere resultatene og dra konklusjoner basert på statistisk signifikans.

3.2.2 Innsamling og analyse av data

Dataene er samlet inn fra TITLON-databasen, som er utviklet gjennom et samarbeid mellom universiteter og høyskoler i Norge. Denne databasen inneholder detaljerte opplysninger om selskaper, fond, obligasjoner og andre derivater (TITLON, 2024). Det er lastet ned over 300 årsrapporter, og det er denne informasjonen som utgjør grunnlaget for det som er nødvendig for at vi skal kunne besvare forskningsspørsmålene våre. Dataene består av aksjedata for selskaper notert på Oslo Børs. Selskapet med færrest observasjoner har kun 9 observasjoner i løpet av året, mens de selskapene med flest observasjoner hadde opptil 512 observasjoner gjennom året. De fleste selskapene har 256 observasjoner, men vi ser at det kan være stor variasjon i antall observasjoner. Denne variasjonen kan ha betydning for dataens konsistens da selskaper med færre observasjoner kan gi mindre pålitelige resultater og påvirke signifikansen av analysen.

Observasjonene i datasettet vårt inneholder tall for aksjene fra 2010 og 2011. For å sikre en presis analyse har vi delt opp datasettet etter hvilket år observasjonene er fra. Etter å ha delt datasettet for år 2010 og år 2011, har vi laget separate tabeller for hvert år, hvor disse tabellene gir ulike statistiske mål. For å forenkle tabellene våre, deler vi karakterene som GPT har gitt oss i tre kategorier. Den første kategorien tar for seg karakterene 7-10, og får en

verdi på 3, karakterene 4-6 får en verdi på 2, og karakterene 3-0 omgjøres til verdi 1. Den høyeste karakteren gitt av GPT var 8, og den laveste karakteren var 2.

3.3 Evaluering av metode

I dette delkapitlet vurderes kvaliteten på forskningen, med fokus på studiens pålitelighet og relevans. Vi benytter sekundærdata hentet fra årsrapporter i databasen TITLON, og har derfor ikke hatt kontroll over utarbeidelsen av datagrunnlaget. Viktige kriterier i vurderingen av forskningens kvalitet er validitet og reliabilitet (Sander, 2022). Det er imidlertid diskutert i hvilken grad disse kriteriene er egnet for å vurdere kvantitative og kvalitative forskningsmetoder. Det argumenteres for at kriteriene er mer egnet til å vurdere kvantitative studier enn kvalitative (Sander, 2022). For å kunne si noe om resultatene av denne studien, må vi derfor vurdere om forskningsmetoden oppfyller disse kriteriene. Evalueringen i denne studien gjøres ved å vurdere flere kritiske aspekter som validitet og reliabilitet.

3.3.1 Validitet

En av de viktigste delene i en forskningsoppgave er å undersøke om resultatene man får har en tilstrekkelig validitet. Hvis en undersøkelse mangler validitet kan den anses som verdiløs (Cohen et al, 2007). Når man ser på betydningen av validitet innenfor et forskningsemne, handler det om graden av nøyaktighet og hvorvidt et verktøy måler det det faktisk er ment å måle. I konteksten av denne studien undersøker vi nøyaktigheten til GPT, og om det kan anvendes for å predikere aksjekurser basert på årsrapporter.

For å vurdere validiteten til studiet gjør vi en regresjon mellom den uavhengige og avhengige variabelen. I vårt tilfelle vil den uavhengige variabelen være "karakteren" vi får fra årsrapportene (0 til 10). Denne scoren bygges opp av kvantitative og kvalitative data som blir hentet fra årsrapportene og analysert gjennom GPT. Den avhengige variabelen er selve aksjekursene som vi har da det er årsrapporter fra tidligere år. Ved å be GPT analysere dataene fra årsrapportene vil den gi oss en karakter, og denne karakteren vil bli analysert og sammenlignet den gir en fornuftig mening sammenlignet med faktisk avkastning.

I forskningsteorien er det vanlig å skille mellom ulike typer av validitet. I eksperimentelle studier, som denne oppgaven er et eksempel på, skiller man mellom indre og ytre validitet (Grønmo, et al. 2024). Indre validitet handler om at eksperimentet er utført på en forsvarlig måte. Det betyr at slutningen om årsakssammenheng er gyldig under de ulike betingelsene

man ser på i en undersøkelse. Dersom man ser at årsakssammenhengen også er gyldig i andre situasjoner enn selve eksperimentet, kan den generaliseres og man sier dermed at eksperimentet har en høy ytre validitet. I vår oppgave kan vi ikke konkludere med en høy ytre validitet ettersom vi ikke kan garantere at svarene som GPT gir oss er identiske hver gang.

3.4.2 Reliabilitet

Reliabilitet handler om pålitelighet og er en angivelse av om en undersøkelse eller et eksperiment viser den virkelige situasjonen, og til hvor stor grad resultatene kan ettertestes (Sander, 2022). En høy reliabilitet betyr at alle uavhengige målinger andre gjør alltid skal resultere i et tilnærmet identisk resultat eller konklusjon. Altså hvis andre forskere gjennomfører en identisk undersøkelse, vil de komme frem til et likt resultat. Hvis undersøkelsene fører til ulike resultater, kan det tyde på en lav reliabilitet. I denne oppgaven hvor vi bruker en generativ språkmodell som GPT til å gi oss datamateriale, kan det være utfordringer knyttet til reliabiliteten i resultatene vi får. Dette diskuterer vi nærmere i diskusjonskapittelet.

3.4.3 Vurdering av GPT

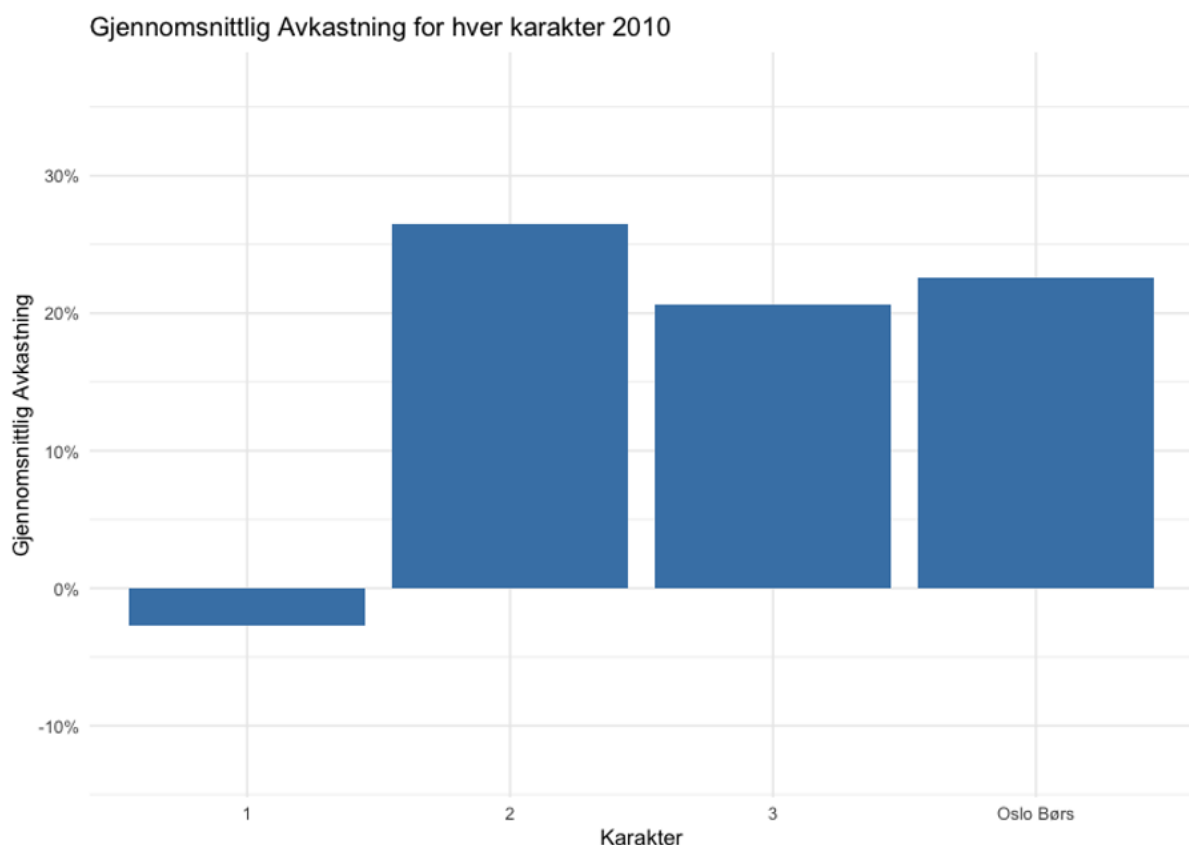
Tradisjonelle maskinlæringsmodeller innenfor finans er typisk designet for prediksjon, uten noe spesiell form for forklaringsevne av dataen som gis ut. I motsetning til dette tilbyr de nye språkmodellene som GPT i tillegg til prediksjon av data, en forklaring tilhørende dataene som genereres. Denne evnen til å generere data skiller GPT og andre former for kunstig intelligens fra de mer tradisjonelle modellene. I vår analyse har vi oppdaget at GPT konsekvent tilbyr både forutsigelser og forklaringer for de ulike dataene vi har bedt språkmodellen om å predikere.

4 Analyse og resultater

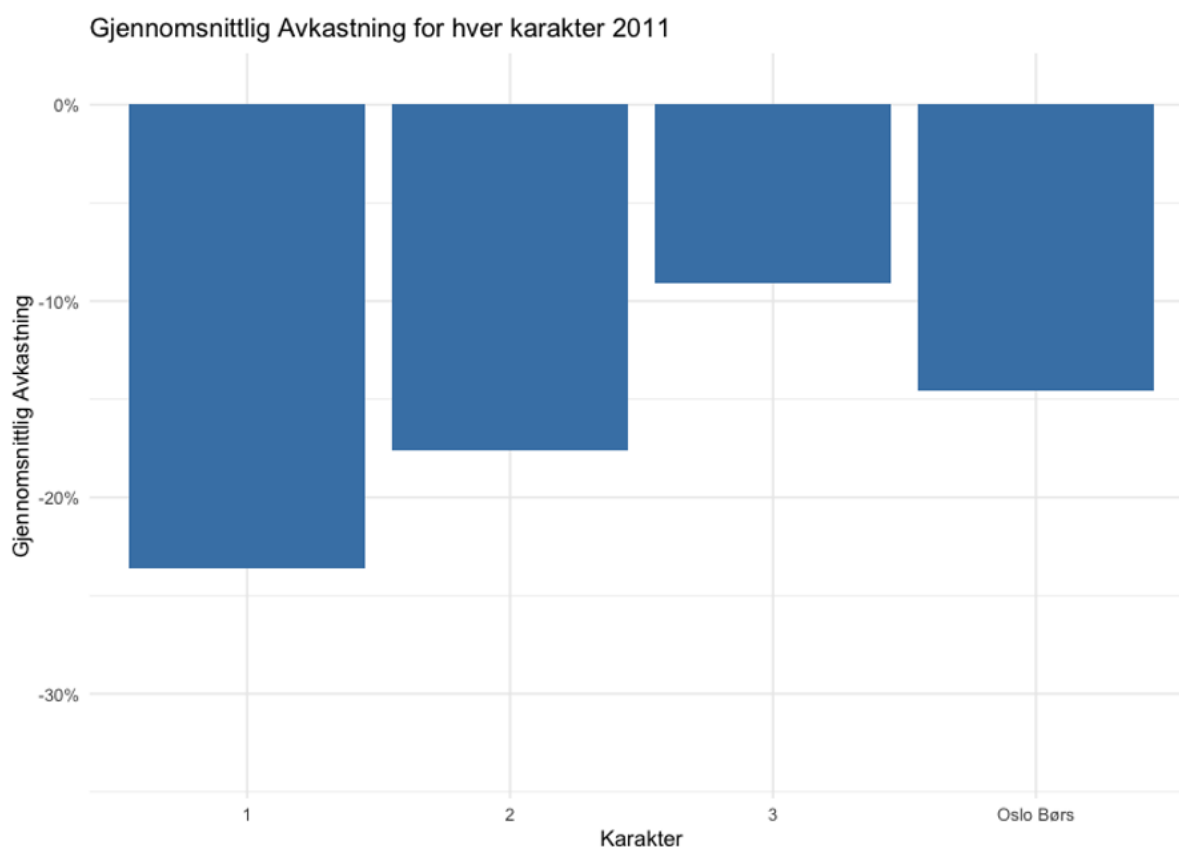
I dette kapitlet blir studiens resultater lagt frem. Datasettet som er lagt som vedlegg er brukt som grunnlag for å komme frem til de ulike resultatene som blir presentert. Karakterene i datasettet er gitt fra 0-10, men vi har valgt å slå disse sammen til 3 karaktergrupper, slik at hver gruppe får flere observasjoner. Årsaken til denne justeringen er for å få likere antall observasjoner mellom de ulike karaktergruppene. Først vil vi presentere de ulike karaktergruppene avkastning i årene 2010 og 2011. Deretter presenterer vi en regresjonsmodell, og til slutt en OLS regresjonsanalyse.

4.1 Avkastning for karakter grupper sammenlignet med Oslo Børs

Vi starter med å se på avkastningen for årene 2010 og 2011 i et søylediagram. Vi har lagt til et fjerde nivå som inkluderer indeksen. I figur 2 og 3 presenteres avkastningen for hver av karakternivåene, samt indeksen Oslo Børs, for hvert av årene.



Figur 2: Gjennomsnittlig avkastning for hver karaktergruppe og Oslo børs fra 1 juni 2010 til 1 juni 2011.



Figur 3: Gjennomsnittlig avkastning for hver karaktergruppe og Oslo børs fra 1 juni 2011 til 1 juni 2012.

Figur 2 og 3 viser at det er tydelig sammenheng mellom GPTs karakterrangering av selskaper og den påfølgende aksjeavkastningen. Vi ser for år 2010 at karakter 1 gir en negativ avkastning på -2,7%, karakter 2 gir 26,6% avkastning og karakter 3 gir 21%. For 2011 har karakter 1 en negativ avkastning på -23,6%, karakter 2 en negativ avkastning på 17,6% og karakter 3 en negativ avkastning på -9,1%.

For året 2010 observeres det en markant avkastning for karakterene 4-6 (karaktergruppe 2), som er vesentlig høyere enn de andre gruppene. Dette funnet er imidlertid noe misvisende, da det hovedsakelig skyldes den eksepsjonelle avkastningen til SAS AB, som var på 2425%. Dersom vi ekskluderer SAS AB, ville den gjennomsnittlige avkastningen for denne karaktergruppen vært mellom gruppene 0-3 (karaktergruppe 1) og 7-10 (karakter gruppe 3). Til tross for denne urimeligheten, viser dataene fortsatt en trend der høyere GPT-karakterer generelt korrelerer med høyere avkastning.

Dataene for 2011 viser at alle karaktergruppene, samt Oslo Børs, hadde negativ avkastning. Likevel viser funnene våre at karakter 1 har den svakeste avkastningen sammenlagt, noe som

tyder på at GPT klarer å identifisere selskapene som har en dårlig forutsetning for vekst i aksjen. Dette antyder at GPT har en evne til å plukke ut selskaper som presterer dårlig, noe som kan være svært nyttig informasjon for investorer. Det er også verdt å merke seg at karaktergruppe 7-10 hadde bedre avkastning enn indeksen på Oslo børs i 2011.

Sammenlagt av de to figurene ser vi en tydelig sammenheng mellom GPTs karakterrangeringer og påfølgende aksjeavkastning. Generelt korrelerer høyere karakterer med høyere avkastning, noe som støtter hypotesen om at GPTs rangeringer er en verdifull indikator for fremtidig aksjeutvikling. I 2010 hadde karaktergruppe 2 (4-6) den høyeste avkastningen, men dette skyldes hovedsakelig den eksepsjonelle prestasjonen til SAS AB. Uten denne påvirkningen, følger trenden fortsatt at høyere karakterer gir høyere avkastning. I 2011 opplevde alle grupper negativ avkastning, men GPT identifiserte de dårligste selskapene (karakter 0-3) som hadde størst negativ avkastning.

4.2 Statistikk

I dette kapittelet presenterer vi resultatene fra vår analyse av sammenhengen mellom GPTs karakterrangeringer og aksjeavkastning. Vi viser avkastning, standardfeil og T-verdier for ulike karaktergrupper for årene 2010 og 2011. Disse statistikkene gir innsikt i påliteligheten og signifikansen av de observerte avkastningene, og hjelper oss å forstå hvordan GPTs vurderinger kan brukes som indikator for fremtidig aksjeutvikling. Vi vil også diskutere konfidensintervallene for å bedre forstå spredningen og variasjonen i avkastningen.

Resultater 2010

Karaktergruppe	Avkastning	Standardfeil	Antall	T-verdi
0-3	-0.027	0.157	88	2.086
4-6	0.266	0.240	103	1.983
7-10	0.210	0.056	21	1.988

Tabell 2: Resultater år 2010: Viser avkastning og statistikk for hver karaktergruppe for 2010.

Resultater 2011

Karaktergruppe	Avkastning	Standardfeil	Antall	T-verdi
0-3	-0.236	0.248	12	2.204
4-6	-0.176	0.055	60	2.007
7-10	-0.091	0.056	44	2.021

Tabell 3: Resultater år 2011: Viser avkastning og statistikk for hver karaktergruppe for 2011

Tabell 2 og 3 presenterer avkastning og statistikk for de ulike karaktergruppene for årene 2010 og 2011. De viser også avkastningen, standardfeil, antall observasjoner per karaktergruppe og til slutt T-verdien til karaktergruppene.

4.2.1 Standardfeil

Standardfeilen gir et mål på hvor mye observasjonene avviker fra gjennomsnittet, og den gir en indikasjon på hvor presist gjennomsnittet er estimert. En lav standardfeil indikerer at dataene ligger tett rundt gjennomsnittet, mens en høy standardfeil viser større spredning.

I 2010 hadde karaktergruppene 4-6 den høyeste standardfeilen (0,240), noe som tyder på større spredning i avkastningen innenfor denne gruppen sammenlignet med de andre gruppene, dette gir mening da det er størst svingninger i avkastningen for denne karaktergruppen. Karaktergruppen 7-10 hadde den laveste standardfeilen (0,056), som indikerer at avkastningene var mer konsistente i denne gruppen. I 2011 var standardfeilen størst i karaktergruppen 0-3 (0,248), noe som viser stor spredning i avkastningene.

Karaktergruppene 4-6 og 7-10 hadde liknende standardfeil (0,055 og 0,056), noe som tyder på at avkastningene i disse gruppene var mer konsistente.

4.2.2 T-verdi

T-verdien brukes for å teste hypoteser om middelverdien av en populasjon. Den indikerer hvor mange standardavvik en observasjon er fra nullhypotesen (ofte er middelverdien lik null). En høy T-verdi (i absolutt verdi) antyder at vi kan avvise nullhypotesen og konkludere med at det er en signifikant forskjell.

I 2010 hadde karaktergruppen 0-3 en T-verdi på 2,086, karaktergruppen 4-6 en T-verdi på 1,983, og karaktergruppen 7-10 en T-verdi på 1,988. Disse verdiene indikerer at avkastningene i alle gruppene er signifikant forskjellige fra null.

I 2011 hadde karaktergruppen 0-3 en T-verdi på 2,204, karaktergruppen 4-6 en T-verdi på 2,007, og karaktergruppen 7-10 en T-verdi på 2,021. Igjen viser disse verdiene at avkastningene i alle gruppene er signifikant forskjellige fra null.

4.2.3 Analyse av standardfeil og T-verdi

Standardfeilene og T-verdiene gir oss viktig innsikt i spredningen og signifikansen av avkastningene i de ulike karaktergruppene. En lav standardfeil kombinert med en høy T-verdi indikerer pålitelige og signifikante resultater. I denne analysen viser resultatene for karaktergruppen 7-10 i 2010 en lav standardfeil og relativt høy T-verdi, noe som gjør resultatene mer pålitelige. Samtidig viser dataene for 2011 at karaktergruppen 0-3 har en høy standardfeil, men også en høy T-verdi, som indikerer at når det er stor spredning i dataene, er forskjellene fra null signifikante.

Samlet sett tyder disse funnene på at avkastningene i alle karaktergruppene for begge årene er signifikant forskjellige fra null, noe som gir en viss grad av tillit til at de observerte avkastningene ikke er tilfeldige. Disse resultatene vil være nyttige for å forstå prestasjonene til de ulike karaktergruppene og kan danne grunnlaget for videre analyser og beslutninger.

4.2.4 Konfidensintervall

Konfidensintervallene for årene 2010 og 2011 gir en indikasjon på hvor den sanne gjennomsnittlige avkastningen for hver karaktergruppe sannsynligvis ligger. For 2010 viser konfidensintervallet for karaktergruppe 0-3 at den gjennomsnittlige avkastningen med 95% sikkerhet ligger mellom -35.56% og 30.08%. For karaktergruppe 4-6, som er sterkt påvirket av SAS ABs avkastning på 2425%, ligger intervallet mellom -21.01% og 74.06%. Uten denne eksepsjonelle avkastningen ville intervallet vært mindre ekstremt, og avkastningen ville vært nærmere midten av de andre gruppene. For karaktergruppe 7-10 er intervallet mellom 9.56% og 31.78%.

For 2011 er konfidensintervallet for karaktergruppe 1 (tilsvarer 0-3) mellom -78.2% og 31.0%, for karaktergruppe 2 (tilsvarer 4-6) mellom -28.6% og -6.63%, og for karaktergruppe 3 (tilsvarer 7-10) mellom -20.3% og 2.14%. Dette viser at der alle grupper opplevde negative

gjennomsnittlige avkastninger, er usikkerheten og variasjonen høy, spesielt for lavere karaktergrupper.

Oppsummert indikerer konfidensintervallene for begge årene at høyere karakterer generelt er assosiert med bedre avkastning, selv om det er stor variasjon. I 2010 var avkastningen for karaktergruppe 4-6 kunstig høy på grunn av en enkelt aksje, mens i 2011 viste alle grupper negative avkastninger.

Basert på dataene og de beregnede konfidensintervallene, viser aksjene med høyere karakterer fra GPT en tendens til å prestere bedre enn de med lavere karakterer. Spesielt gruppen med karakterene 7-10 har en statistisk signifikant positiv gjennomsnittlig avkastning, som tyder på at høyere karakterer kan være en god indikasjon på fremtidig avkastning. Dette må imidlertid analyseres videre for å forstå de underliggende faktorene som påvirker disse resultatene.

4.3 Regresjonsmodell

I dette kapittelet presenterer vi en regresjonsanalyse som undersøker effekten av ulike karakterer på en avhengig variabel, med karakter 1 som referansenivå. Modellen estimerer hvordan karakterene 2 og 3 påvirker den avhengige variabelen sammenlignet med karakter 1. Tabellen nedenfor viser estimater, standardfeil, t-verdier og p-verdier for hver karakter. Denne analysen gir innsikt i hvorvidt karakterene har en signifikant effekt på den avhengige variabelen, og hvor mye av variasjonen som kan forklares av modellen.

Karakter	Estimat	Standardfeil	T-verdi	P-verdi
Ref. nivå (1)	-0.0025	0.00075	-3.456	0.0006
2	0.0016	0.0008	1.905	0.0577
3	0.0030	0.0008	3.475	0.0006

Tabell 4: Regresjonsmodell med karakter 1 som referansenivå og alfa som avhengig variabel.

Estimatet for referansenivået er $-0,0025821$ med en standardfeil på $0,0007471$. T-verdien er $-3,456$, og p-verdien er $0,000622$. Dette betyr at estimatet for referansenivået er signifikant forskjellig fra null på et konfidensnivå på 95%. Den negative verdien indikerer at karakter 1 har en negativ effekt på den avhengige variabelen sammenlignet med karakter 3 og karakter 2.

Estimatet for karakter 3 er $0,0029072$ med en standardfeil på $0,0008366$. T-verdien er $3,475$, og p-verdien er $0,000581$. Dette indikerer at karakter 3 har en signifikant positiv effekt på den avhengige variabelen sammenlignet med referansenivået (karakter 1). Estimatet er signifikant på et konfidensnivå på 95%.

Estimatet for karakter 2 er $0,0015606$ med en standardfeil på $0,0008193$. T-verdien er $1,905$, og p-verdien er $0,057688$. Dette indikerer at effekten av karakter 2 på den avhengige variabelen ikke er signifikant forskjellig fra null på et konfidensnivå på 95%. Likevel ser vi at den positive verdien antyder en mulig positiv effekt, men den er ikke statistisk signifikant.

R-squared-verdien for modellen er $0,03737$. Dette indikerer at modellen forklarer omtrent 3,737% av variasjonen i den avhengige variabelen. Selv om R-squared-verdien er relativt lav, kan de signifikante koeffisientene fortsatt gi innsikt i forholdet mellom karakterer og den avhengige variabelen.

Oppsummert ser vi i regresjonsmodellen at karakter 3 har en signifikant positiv effekt på den avhengige variabelen sammenlignet med referansenivået (karakter 1), mens effekten av karakter 2 ikke er signifikant. Den lave R-squared-verdien antyder at det er andre faktorer som også påvirker den avhengige variabelen, som ikke er inkludert i denne modellen.

4.4 Test av heteroskedastisitet

I denne delen av analysen undersøker vi om det er tilstedeværelse av heteroskedastisitet i den lineære regresjonsmodellen som er brukt. Heteroskedastisitet refererer til tilfeller der variansen av residualene i en regresjonsmodell ikke er konstant. Dette kan føre til ineffisiente estimater.

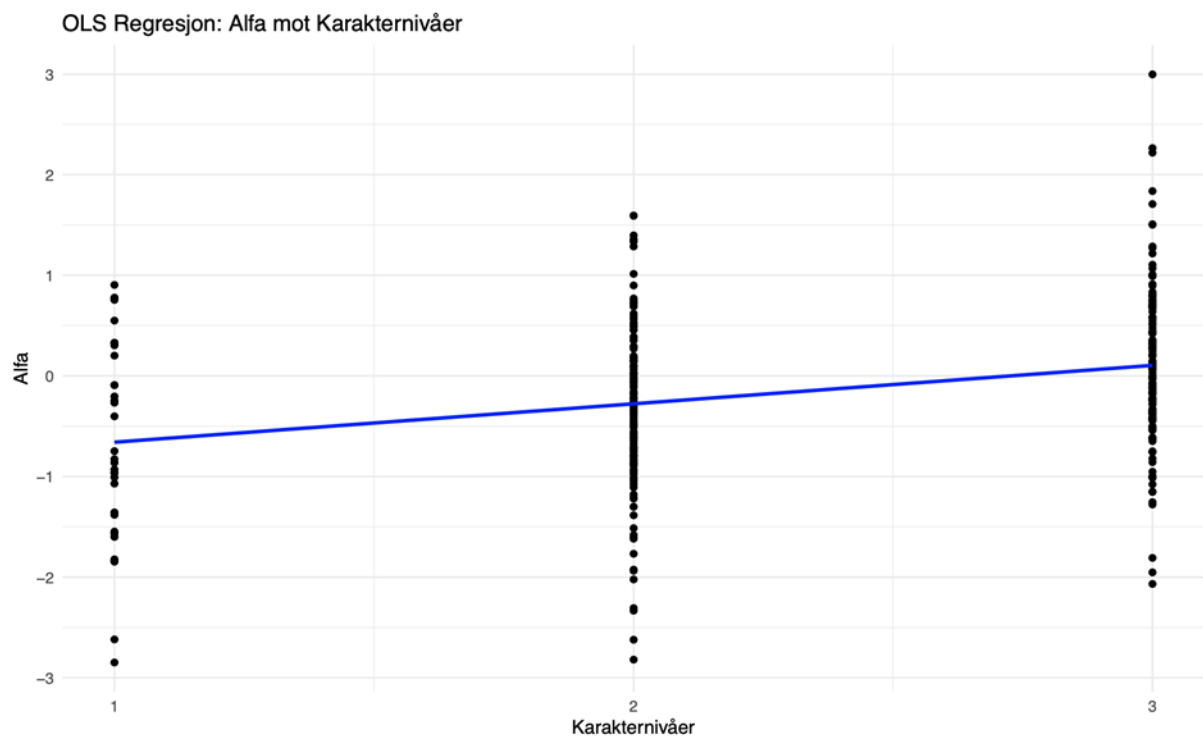
Breusch-Pagan Test	0,17011
Df	2
P-verdi	0,9185

Tabell 5: Viser tallene for Breusch-Pagan test

For å sikre validiteten av den lineære regresjonsmodellen som er benyttet, er det viktig å undersøke om variansen av residualene er konstant, en egenskap kjent som homoskedastisitet. Heteroskedastisitet, derimot, oppstår når variansen av residualene varierer systematisk med verdiene til en eller flere uavhengige variabler, noe som kan føre til ineffisiente og skjevhet i estimatene. Vi benyttet oss av Breusch-Pagan-testen for å teste for heteroskedastisitet. Breusch-Pagan-testen er en av de vanligste metodene for å diagnostisere heteroskedastisitet i regresjonsmodeller. Nullhypotesen for denne testen er at det ikke er heteroskedastisitet, altså at variansen av residualene er konstant. Testen ble gjennomført på vår regresjonsmodell, og resultatene ser vi i tabell 5.

4.5 OLS regresjon

I denne delen av analysen har vi gjennomført en OLS-regresjonsanalyse i Rstudio for å undersøke sammenhengen mellom t-verdien for alfa og vurderingene gitt av GPT for aksjens fremtidige utsikter. Formålet med analysen er å forstå hvorvidt positive utsikter for en aksje, målt ved en karakter fra 1 til 3, har en signifikant innvirkning på den risikjusterte meravkastningen representert ved t-verdien for alfa.



Figur 4: OLS regresjon: T-Alfa mot karakternivå

X-aksen i figur 4 representerer de tre karakternivåene, og viser forskjellige nivåer av variabelen som er kodet som karakter i datasettet. Y-aksen representerer T-verdiene for alfa som strekker seg fra -3 til 3.

Definisjoner:

- T-Alfa : T-verdien for alfa, som representerer risikojustert meravkastning for et gitt år.
- Karaktergrupper: En numerisk karakter fra 1 til 3 gitt av GPT som vurderer de fremtidige utsiktene for aksjekursen, der 3 indikerer meget positive utsikter og 1 indikerer negative utsikter.

For å illustrere forholdet mellom karaktergrupper og T-Alfa er det generert et scatterplot med en regresjonslinje. Plottet viser at til tross for den signifikante positive sammenhengen, er variabiliteten i t-verdiene høy. Dette indikerer at selv om positive utsikter (høyere karakter) generelt fører til høyere t-verdier for alfa, er det mange andre faktorer som også spiller en rolle. Det ser ikke ut til at det er en klar lineær sammenheng mellom karaktergruppene og T-verdien for alfa, siden alle punktene er spredt uten en tydelig trend.

Resultatene fra OLS-regresjonen viser en positiv sammenheng mellom Alfa og Karakter. Koeffisientene er som følger: Interceptet er -0.6719 , noe som er statistisk signifikant ($p = 3.99e-06$) og representerer den forventede t-verdien for alfa når Karaktergruppen er 1. For Karaktergruppen 2 er koeffisienten 0.4001 (statistisk signifikant, $p = 0.0113$), som indikerer en økning på 0.4001 i t-verdien for alfa ved overgang fra Karaktergruppen 1 til Karaktergruppe 2. For Karaktergruppe 3 er koeffisienten 0.7731 (statistisk signifikant, $p = 2.12e-06$), som indikerer en økning på 0.7731 i t-verdien for alfa ved overgang fra Karaktergruppe 1 til Karaktergruppe 3.

Nøkkeltall fra analysen inkluderer en residual standard error på 0.8226 med 325 frihetsgrader, en multiple R-squared på 0.08216 , en adjusted R-squared på 0.07652 , og en F-statistic på 14.55 (p -verdi = $8.897e-07$). Dette viser at det er en signifikant sammenheng mellom vurderingene (Karakter) og den risikjusterte meravkastningen (Alfa), selv om modellen forklarer en begrenset del av variansen.

4.6 Oppsummering

Funnene i vår analysedel viser at karakterene gitt av GPT har en betydelig innvirkning på den gjennomsnittlige avkastningen av selskapene plassert i de ulike karaktergruppene 1, 2 og 3. Det fremgår tydelig av figur 2 og 3 at høyere karakterer korrelerer med høyere avkastning. Videre viser tabell 2 og 3 at karaktergruppe 1 har høyere standardfeil enn karaktergruppe 2 og 3, noe som indikerer at avkastningen var mer konsistente i de høyere karaktergruppene både for 2010 og 2011. T-verdiene indikerer at avkastningen i alle gruppene er signifikant forskjellig fra null. Konfidensintervallene bekrefter også at høyere karakterer generelt er assosiert med høyere avkastning, selv om det er betydelig variasjon.

I regresjonsmodellen fremkommer det at karakter 3 har en signifikant positiv effekt på den avhengige variabelen sammenlignet med referansenivået. Breusch-Pagan-testen for heteroskedastisitet viser en p -verdi betydelig høyere enn $0,05$, noe som indikerer at det ikke finnes statistisk signifikante bevis for heteroskedastisitet i vår modell. Resultatene fra OLS-regresjonen viser en positiv sammenheng mellom Alfa og Karakter, som understøtter hypotesen om at høyere GPT-karakterer er en god indikator for fremtidig aksjeavkastning.

5 Diskusjon

I dette avsluttende kapitlet vil vi diskutere resultatene fra vår studie, som undersøker GPTs evne til å analysere årsrapporter, for så å gi en karakter basert på fremtidsutsiktene til aksjen. Målet med dette studiet er å vurdere om AI kan brukes til å analysere årsrapporter og forutsi aksjekurser, noe som på sikt kan bidra til å revolusjonere måten investorer tar beslutninger på. Hensikten er å se om AI kan brukes som verktøy for å lage en portefølje basert på karakterer, og se om det er mulig å oppnå meravkastning sammenlignet opp mot Oslo Børs.

5.1 Sammendrag av resultater

Resultatene våre viser sammenhengen mellom karakterene som er gitt av GPT til aksjene og deres tilhørende avkastning og nøkkeltall. Aksjene som fikk høyere karakterer (7-8) oppnådde i gjennomsnitt høyere avkastning kontra aksjene som fikk lavere karakter (2-3) noe som tabell 2 viser. For eksempel ser vi at aksjene som havnet i karaktergruppe 3 fikk 21% avkastning, i motsetning til aksjene i karaktergruppe 1 som endte med en negativ avkastning på 2.7% for aksjene i 2010.

Resultatene viser at høyere karakter korrelerer med høyere avkastning, og lavere karakter korrelerer med lavere avkastning. Selv om vi fant noen avvik som figur 2 viser. Aksjer som havnet i karaktergruppe 2 fikk i gjennomsnitt høyere avkastning enn aksjene i karaktergruppe 3 i 2010, noe som forklares tidligere i oppgaven med en enkelt aksje som trakk gjennomsnittet svært høyt opp.

Sammenlignet med avkastningen på Oslo Børs ser vi i figur 2 at det kun er karaktergruppe 2 som slår Oslo Børs på avkastning i 2010. For perioden vi analyserte (2010-2011) kunne Oslo Børs vise til en avkastning på 22,6%, noe som var mindre enn gjennomsnittlig avkastning for karaktergruppe 2.

Generelt viser figur 2 og 3 at høyere karakter gitt av GPT korrelerer med høyere avkastning, og at lavere karakter gitt av GPT korrelerer med lavere avkastning. Dette indikerer at karakteren GPT gir ut, har en sammenheng med avkastningen. Dette støtter vårt forskningsspørsmål om at det er en sammenheng mellom GPTs tildelte karakterer, noe som kan være en verdifull indikator for fremtidig aksjeutvikling. Dette viser at GPT også kan identifisere svake investeringsobjekter, noe som er nyttig for investorer, enten det er aksjer de

vil unngå i en portefølje eller det er aksjer som kan shortes. Med shorting menes det at man satser på at en aksje skal ha negativ utvikling.

Resultatene fra OLS-regresjonen viser en også en positiv sammenheng mellom T-alfa og karakter. Koeffisientene indikerer at høyere karakter generelt fører til høyere t-verdier i alfa. Det viser en gradvis økning i t-verdien for alfa med høyere karaktergruppe, noe som indikerer at positive utsiktene vurdert av GPT har en signifikant positiv effekt på den risikjusterte meravkastningen. Til tross for den signifikante positive sammenhengen, viser scatterplottet at variabiliteten i t-verdiene er høy, noe som indikerer at det er mange andre faktorer som spiller en rolle i bestemmelsen av t-verdiene for alfa. Dermed er det ikke en klar lineær sammenheng mellom karakternivåene og t-verdiene.

Videre har vi testet for heteroskedastisitet som vises i tabell 5.

En p-verdi over 0.05 indikerer at vi ikke kan forkaste nullhypotesen om homoskedastisitet. I vårt tilfelle er p-verdien 0.9185, som er betydelig høyere enn 0.05. Dette resultatet betyr at det ikke finnes statistisk signifikante bevis for heteroskedastisitet i modellen vår.

Konklusjonen vi trekker fra resultatene i Breusch-Pagan-testen er dermed at variansen av residualene i vår lineære modell antas å være konstant. Dette innebærer at modellen oppfyller kravet om homoskedastisitet, som er en viktig forutsetning for at estimatene i regresjonsanalysen skal være pålitelige og effektive.

5.2 Konklusjon

Basert på vår studie, viser resultatene at GPT har potensiale som et verktøy for å predikere avkastningen basert på årsrapporter for selskaper notert på Oslo Børs, men det er fortsatt betydelige begrensninger. GPT klarte i noen tilfeller å gi karakterer til aksjer som korrelerte med bedre avkastning, spesielt i perioder som 2011 hvor høyere karakterer generelt ga høyere avkastning. Samtidig opplevde vi avvik, hvor aksjer med midt-range karakterer presterte bedre enn de med toppkarakterer. Dette indikerer at selv om GPT kan identifisere aksjer med høy avkastning, har modellen utfordringer med å differensiere nøyaktig mellom aksjer med svært høye karakterer. Sammenlignet med hovedindeksen på Oslo Børs, klarte ingen av AI-genererte porteføljer å overgå markedet konsekvent i den analyserte perioden. Dette antyder at mens GPT kan bidra med verdifull innsikt, er det ikke tilstrekkelig alene for å oppnå meravkastning.

For å svare på problemstillingen vår konkluderer vi med at GPT er et lovende og tidseffektivt, men foreløpig begrenset verktøy for prediksjon av aksjeavkastning. Videre forbedringer og integrering med andre analysemetoder er nødvendig for å øke modellens nøyaktighet og pålitelighet.

Studien viser også en klar sammenheng mellom AI-tildelte karakterer til aksjer og deres faktiske avkastning til aksjene i datasettet. Dette indikerer at GPT til en viss grad har evne til å identifisere selskaper med gode eller dårlige fremtidsutsikter basert på innholdet i årsrapportene. Dette ser vi tydelig i figur 2 og 3 hvor bedre karakter gjenspeiler høyere avkastning, sett bort fra det forklarte avviket i karaktergruppe 2 for 2010.

Fra et teoretisk perspektiv støtter våre funn teorien om markedseffisiens (Gårseth-Nesbakk, 2023), som antyder at all tilgjengelig informasjon allerede er reflektert i aksjekursen. Denne teorien støtter opp om at det er vanskelig å oppnå meravkastning gjennom analyser av årsrapporten i seg selv. Videre ser vi i henhold til teorien om Random Walks (Bodie et al, 2009) at prisendringer i aksjemarkedet er tilfeldige og uforutsigbare, noe som utfordrer GPTs evne til å kunne predikere fremtidige aksjekurser.

5.3 Metodekritikk

Metodekritikk er en vurdering av forskningsmetodene som er benyttet i en studie, med hensikt å identifisere styrker og svakheter. Denne delen av oppgaven tar for seg en kritisk gjennomgang av metodene som er anvendt. Ved å evaluere metodene grundig, kan vi vurdere gyldigheten og påliteligheten av funnene våre.

5.3.1 Styrker ved metodikken

En av de største styrkene ved metodikken vår omhandler effektivitet ved bruk av AI til å analysere årsrapporter. AI gir oss en systematisk og objektiv vurdering av informasjon som kan være vanskelig og svært tidkrevende å oppnå ved en manuell analyse. Automatiseringen gjennom API kodet i Python muliggjør effektiv datainnsamling og håndtering. Ved å integrere dette med GPT-analyse er det mulig å analysere en stor mengde rapporter på kort tid. Denne kombinasjonen forbedrer beslutningstaking ved å levere nøyaktige og relevante resultater raskt. Dette betyr at metoden kan utvides til å inkludere flere selskaper og lengre tidsperioder uten at arbeidsbelastningen økes tilsvarende med antall aksjer og observasjoner.

Ettersom prosessen vår har vært automatisert og følger faste algoritmer, vil det dermed være enkelt å replikere studien, samt verifisere resultatene basert på dataen vi har fått. Andre forskere kan kjøre de samme analysene på ulike datasett for å teste robustheten og generaliserbarhetene av funnene.

5.3.2 Svakheter ved metodikken

Etter arbeidet med å analysere funnene våre, stiller vi spørsmål til størrelsen på datamaterialet. Etter en grundigere analyse ser vi at enkeltaksjer har påvirket den gjennomsnittlige avkastningen til de ulike karaktergruppene i stor grad, noe som har ført til større konfidensintervaller for de ulike karakterene. Særlig for aksjene som tilhører 2011, ser vi et større sprang i konfidensintervallene. Dette forklares enkelt med at det er færre observasjoner, hvor enkeltaksjer som skiller seg ut med enten svært høy avkastning, eller store fall i aksjekursen trekker konfidensintervallene ut. Dette kan skape et misvisende bilde på resultatene som presenteres. Ettersom oppgaven vår baserer seg på svar fra GPT kan vi ikke garantere med 100% sikkerhet at de karakterene GPT presenterer basert på årsrapportene vil være de samme hver gang man henter ut data. Dette fører til utfordringer knyttet til reliabiliteten i oppgaven.

5.4 Forslag til videre forskning

En naturlig forlengelse av vår studie vil være å utvide datasettet med å inkludere enda flere aksjer og teste dem over andre tidsperioder. Man kan gjerne trekke inn aksjer fra andre markeder enn de som er på Oslo Børs for å undersøke om resultatene fortsatt ligner. Et større og mer variert datasett vil kunne gi en mer robust analyse og redusere påvirkningskraften enkelte aksjer gir på resultatene. Dette vil gjøre det mulig å teste algoritmens ytelse i ulike markedsforhold og perioder, noe som kan gi innsikt i hvordan metoden fungerer under ulike økonomiske scenario. Man kan også implementere bruk av kvartalsrapporter i stedet for bare årsrapporter.

En annen interessant retning for videre forskning kan være å undersøke bransjer og sektorer separat. Dette for å se om effekten av analysen varierer mellom bransjer. Noen bransjer kan for eksempel ha mer komplekse årsrapporter eller spesifikke indikatorer som er mer relevante for prediksjon av aksjekurser. Ved å analysere data spesifikt knyttet til sektorer kan man få modeller som gir mer presise prediksjoner for hver enkelt sektor.

Videre kunne en spennende utvidelse av studie være å teste effektiviteten av GPT-modellen i ulike markedssykluser, dette inkluderer bull-markeder og bear-markeder. Dette kan være med å bidra til å forstå hvordan GPT-modellen presterer under ulike økonomiske forhold, og kan avdekke styrker og svakheter med modellen i ulike faser av markedet.

Referanseliste

AISuksess. (u.å.). *Hvorfor ta i bruk AI & Automatisering?* AISuksess.

<https://aisuksess.no/hvorfor-ta-i-bruk-ai-automatisering>

Ali., (2023, 11 April). *GPT-1 to GPT-4: Each of OpenAI's GPT Models Explained and Compared*. Make Use Of. <https://www.makeuseof.com/gpt-models-explained-and-compared/>

Autoblocks. (2024). *symbolic artificial intelligence*. Autoblocks

<https://www.autoblocks.ai/glossary/symbolic-artificial-intelligence> Hentet 14.03.2024

Bi., Goodman., Kaminsky., Lessler., (2018). What is Machine Learning? A Primer for the Epidemiologist, 188(12), 2222-2224. <https://doi.org/10.1093/aje/kwz189>

Bodie, Z., Kane, A. og Marcus, A.J. (2009). *Essentials of Investments*. McGraw Hill/Irwin

Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). *Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios*. Journal of Medical Systems, 47(1), 1-5. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9985086/>

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. London: Routledge.

Council of Europe. (2024). *History of Artificial Intelligence*.

<https://www.coe.int/en/web/artificial-intelligence/history-of-ai> Hentet 14.03.2024

Coursera Staff. (2024, 9. februar). *Machine learning in finance*. Coursera

<https://www.coursera.org/articles/machine-learning-in-finance>

Dalen, P. M. (2014) *Er Oslo børs et effisient aksjemarked?: En test av markedseffisiens på svak form i perioden 1996-2013* [Masteroppgave, Norges miljø- og biovitenskapelige universitet]. NMBU Delarkiv. <https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/217347>

Deshpande, M., & Joshi, S. (2022, August 19). Legal, Ethical and Risk Issues of AI in

Banking. *Lexology*. <https://www.lexology.com/library/detail.aspx?g=ad03c23e-a9bc-4df5-88ec-d9f850e1a82a>

- Direktoratet for forvaltning og IKT. (u.å.). *Hva er kunstig intelligens?*. Digdir. <https://www.digdir.no/kunstig-intelligens/hva-er-kunstig-intelligens/4133>
- Fajasy. (2023, 25. oktober). How to calculate and interpret the Fama and French and Carhart multifactor models. *Stablebread*. <https://stablebread.com/how-to-calculate-and-interpret-the-fama-and-french-and-carhart-multifactor-models/>
- Florkin, J. (u.å.). AI i finans. *Julien Florkin*. <https://julienflorkin.com/no/teknologi/kunstig-intelligens/ai-i-finans/>
- Gårseth-Nesbakk, L. (2023, 2. januar). Markedseffisiens. *Store norske leksikon*. <https://snl.no/markedseffisiens>
- Gram, T. (2024, 10. mai). Oslo Børs. *Store norske leksikon*. https://snl.no/Oslo_B%C3%B8rs
- Grønmo, S. (2024, 15. mai). Kvantitativ metode. *Store norske leksikon*. https://snl.no/kvantitativ_metode
- Grønmo, S., Dahlum, S., Svartdal, F. (2024, 13. mai). Validitet. *Store norske leksikon*. <https://snl.no/validitet>
- Håkon R. (u.å.). Kapitalverdimodellen (CAPM): Formel, Forklaring, Eksempel, Bruk. *Investorkurs*. <https://investorkurs.no/kapitalverdimodellen-capm-formel-forklaring-eksempel-bruk/>
- Hayes, A. (2024, 29. januar). Fama and French Three Factor Model Definition: Formula and Interpretation. *Investopedia*. <https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp>
- IBM. (u.å.). What is a Transformer model. *IBM* <https://www.ibm.com/topics/transformer-model>
- Insider Intelligence. (2023, 2. januar). Artificial Intelligence in Financial Services: Applications and benefits of AI in finance. *EMARKETER*. <https://www.emarketer.com/insights/ai-in-finance/>
- Iriondo, R. (2018). Machine Learning vs. AI, Important Differences Between Them. *TOWARDS AI*. <https://towardsai.net/p/machine-learning/machine-learning-vs-aiimportant-differences-between-them/robiriondo/3432/>

- Kaspersky. (u.å.). Hva er datapersonvern?. *Kaspersky* <https://no.kaspersky.com/resource-center/threats/internet-and-individual-privacy-protection>
- Kenton, W. (2020, 30. november). Small Minus Big (SMB). *Investopedia* https://www.investopedia.com/terms/s/small_minus_big.asp
- Martinsen, K. (2023, 27. Januar). den fjerde industrielle revolusjon. *Store norske leksikon* https://snl.no/den_fjerde_industrielle_revolusjon
- Moor. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years, s.87. DOI <https://doi.org/10.1609/aimag.v27i4>.
- Murphy. (2018). Technical Analysis of the Financial Markets. New York institute of finance. <https://py98.ir/blog/wp-content/uploads/2019/08/Technical-Analysis-of-the-Futures-Markets-John-J.-Murphy.pdf>
- Nordnet. (n.d.). Nordnet Markets: Indeks. <https://www.nordnet.no/no/marked/nordnet-markets/indeks>
- Pontiff, J., & Singla, R. (2019). Liquidity. <https://cfr.pub/published/papers/pontiff2020liquidity.pdf>
- Probesto. (u.å.). Betydningen av kunstig intelligens for finansielle prognoser og analyser. *Probesto*. <https://probesto.com/no/betydningen-av-kunstig-intelligens-for-finansielle-prognoser-og-analyser/>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 1-11.
- Regjeringen. (2019). Nasjonal strategi for kunstig intelligens. <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=3>
- Sallam, M., Salim, N. A., Ala'a, B., & Basha, S. (2023). ChatGPT output regarding compulsory vaccination and COVID-19 vaccine conspiracy: A descriptive study at the outset of a paradigm shift in online search for information. *Cureus*, 15(2), e35029. <https://pubmed.ncbi.nlm.nih.gov/36819954/>
- Sander, K. (2022, 1. desember). Reliabilitet. *eStudie* <https://estudie.no/reliabilitet/>

Smith, J.-A. (2024, 13. mars). Hva er generativ kunstig intelligens (GPT)? *Evolon*.

<https://www.evelon.no/artikler/hva-er-generativ-kunstig-intelligens-gpt>

Sonne, G,H,F. (2017, 25 November). Black box-problemet: Når vi ikke forstår den kunstige intelligens. *Videnskab dk*. [https://videnskab.dk/teknologi/black-box-problemet-naar-vi-](https://videnskab.dk/teknologi/black-box-problemet-naar-vi-ikke-forstaar-den-kunstige-intelligens/)

[ikke-forstaar-den-kunstige-intelligens/](https://videnskab.dk/teknologi/black-box-problemet-naar-vi-ikke-forstaar-den-kunstige-intelligens/)

Srivastava, S. (2020, 30. April). Factor Models. *Medium*. [https://medium.com/wright-](https://medium.com/wright-research/factor-models-744e17e5d0e5)

[research/factor-models-744e17e5d0e5](https://medium.com/wright-research/factor-models-744e17e5d0e5)

Vedlegg

Vedlagt ligger lenke til Github, hvor du finner excel arket med data som er benyttet i oppgaven. Karakteren gitt av GPT finnes på raden W ved navn Answer.

- <https://github.com/thomaslars1/Vedlegg-master-Thomas-Tim/tree/397dc94f03914d5d706b0688e24f10d04141af65>

