

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Kevin Kiil 211592IABM

13F finantsraportite automatiseeritud uudiste genereerimine GPT mudeli abil

Magistritöö

Juhendaja: Innar Liiv
PhD

Tallinn 2023

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Kevin Kiil

10.05.2023

Annotatsioon

Käesoleva töö eesmärgiks on uurida, kuidas kasutada tehisintellekti ja teostada andmeanalüüsi, et luua automatiseeritud uudisartikleid 13F raportitest ning hinnata nende usaldusväärsust ja kvaliteeti. Töö põhineb *desk research* meetodikal, kus kogutakse ja eeltöödeldakse andmeid, uuritakse uudisartiklite sisu ja struktuuri, teostatakse andmeanalüüs. Samuti kasutatakse *prompt engineering* meetodeid, et genereerida GPT mudeli abil analüüsitulemusi kirjeldavad kokkuvõtlikud uudisartiklid. Magistritöö pakub praktilist näidet GPT mudeli kasutamisest ja liidestamisest andmeanalüüsi tulemustega, näidates sellega, et see tehnoloogia on potentsiaalselt kasulik ka teistes valdkondades, kus on vaja luua tekstilisi ülesandeid. Töö annab aluse edasisteks uurimisteks ja arendusteks, näiteks GPT mudeli peenhäälestamise ja teiste andmeallikate integreerimise kaudu.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 47 leheküljel, 5 peatükki, 20 joonist, 8 tabelit.

Abstract

Automated news generation from 13F financial reports using the GPT model

The aim of this study is to explore how to use artificial intelligence and perform data analysis to automatically create news articles from 13F reports and evaluate their reliability and quality. The work is based on the desk research methodology, where data is collected and preprocessed, the content and structure of news articles are examined, data analysis is conducted. Also prompt engineering methods are used to generate news articles describing the analysis results with the help of the GPT model. The master's thesis provides a practical example of using and integrating the GPT model with data analysis, demonstrating that this technology has potential benefits in other areas where text-based tasks need to be created. The work lays the foundation for further research and development, such as fine-tuning the GPT model and integrating other data sources.

The thesis is in Estonian and contains 47 pages of text, 5 chapters, 20 figures, 8 tables.

Lühendite ja mõistete sõnastik

Keelemudel	Masinõppe mudel, mis ennustab sõnade või lausete tõenäosusi.
GPT-3	<i>Generative Pre-trained Transformer 3</i> ehk keelemudel, mis on loodud OpenAI poolt
SEC	<i>Securities and Exchange Commission, USA</i> järelevalveasutus, mis reguleerib USA väärtpaberiturgu
13F raport	SEC-i poolt nõutav kvartaalne esitatav dokument, milles institutsionaalsed investorid peavad avaldama oma aktsiapositsioonid, võlakirjad ja muud väärtpaberid, mille koguväärtus ületab 100 miljonit USA dollarit.
CUSIP	<i>Committee on Uniform Securities Identification Procedures</i> , väärtpaberite identifitseerimise standard
CIK	<i>Central Index Key</i> , SEC-i poolt välja antav identifitseerimisnumber
Ticker	Lühendatud sümbol aktsiate või muude väärtpaberite identifitseerimiseks börsil
Token	Ühik, mille abil keelemudel teksti töötleb ja genereerib.
Prompt	GPT mudelile edastatav teksti kujul sisend
Proof of Concept	Teostus, kus demonstreeritakse idee tehnilist teostatavust ja potentsiaalset kasulikkust
API	API (<i>Application Programming Interface</i>) on rakendusliides, mida tarkvara kasutab teise tarkvaraga suhtlemiseks või andmete vahetamiseks

Sisukord

1 Sissejuhatus	11
1.1 Taust ja probleem	11
1.2 Töö eesmärk	11
1.3 Töö struktuur	12
2 Metoodika.....	13
2.1 Objekti detailne kirjeldus.....	13
2.2 GPT-3.5 mudel	13
2.3 Desk research.....	15
2.4 Prompt engineering.....	15
2.5 Andmed	16
2.5.1 SEC 13F raportid	16
2.5.2 Fails-to-Deliver andmekogum.....	17
2.5.3 USA börsidel kaubeldavad aktsiad.....	18
2.6 Tööriistade kirjeldus	18
2.6.1 SQLiteStudio	18
2.6.2 Python.....	18
2.6.3 Visual Studio Code.....	19
2.6.4 Yahoo_fin	19
2.6.5 GPT mudeli API.....	19
2.6.6 Pandas.....	19
2.6.7 Matplotlib	20
2.7 Tööprotsessi kirjeldus.....	20
2.8 Tulemuste valideerimine	21
3 Töö tulemused	23
3.1 Andmete kogumine ja eeltöötlus	23
3.2 Andmebaasi struktuur.....	24
3.2.1 13F raportite andmekogum.....	24
3.2.2 Täiendavad andmed.....	25
3.3 Andmeanalüüs	26

3.3.1 LHV artikli tabel „Top omatud aktsiad“	26
3.3.2 „Top omatud aktsiad“ päringu tehniline teostus	27
3.3.3 LHV artikli tabel „Portfellig suurima osakaaluga“	28
3.3.4 „Portfellig suurima osakaaluga“ päringu tehniline teostus.....	28
3.3.5 LHV artikli tabel „Top ostud 2022. aasta esimeses kvartalis“	29
3.3.6 „Top ostud 2022. aasta esimeses kvartalis“ päringu tehniline teostus	29
3.3.7 Sektorite osakaalud investorite portfelliges aegreana	29
3.3.8 Investorite portfellig suurused	32
3.3.9 Aktsiate osakaalud sektorite lõikes	34
3.4 Liidestamine GPT mudeliga.....	34
3.4.1 GPT mudeli API.....	35
3.4.2 GPT mudeli API-ga kasutatud parameetrid	36
3.5 Rakenduse poolt loodud tulemuste võrdlemine LHV artikli lõikudega	38
3.5.1 „Top omatud aktsiad“ tabelite võrdlemine	38
3.5.2 „Top omatud aktsiad“ lõikude võrdlemine.....	40
3.5.3 „Portfellig suurima osakaaluga“ tabelite võrdlemine	41
3.5.4 „Portfellig suurima osakaaluga“ lõikude võrdlemine	42
3.5.5 „Top ostud“ tabelite võrdlemine	43
3.5.6 „Top ostud“ lõikude võrdlemine	44
3.6 Rakenduse poolt loodud lõikude ja sisendandmete võrdlemine.....	45
3.7 Rakenduse poolt loodud lõikude hindamine GPT mudeli poolt.....	48
3.8 Automatiseeritud uudisartiklite loomine GPT mudeliga.....	49
4 Analüüs ja järeldused.....	51
4.1 Tehnilise teostuse põhjendus	51
4.2 Järeldused GPT mudeli kasutamisest	52
4.2.1 Sisendandmete põhjaliku kirjeldamise olulisus GPT mudeli kasutamisel	52
4.2.2 GPT API muutuja <i>max_tokens</i> mõju sisendandmete analüüsimisele.....	53
4.2.3 GPT mudeli usaldusväärsus sisendandmete peegeldamisel	54
4.2.4 GPT mudeli kasutamine loodud lõikude hindamiseks	54
4.2.5 Välised tegurid, mis mõjutavad andmete võrreldavust erinevates allikates..	55
4.3 Alternatiivid teostatud tööle	56
4.3.1 Olemasolevad lahendused	56
4.4 Edasine töö	57
5 Kokkuvõte	58

Kasutatud kirjandus	59
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	62

Jooniste loetelu

Joonis 1. SEC 13F raportite andmekogumis olevad tabelid ja atribuudid [15].....	17
Joonis 2. SQL päring vahetabeli cik_symbol_sector_value loomiseks 2022 aasta neljanda kvartali kohta.....	30
Joonis 3. SQL päring vahetabeli grouped_quarters loomiseks kõikide kvartalite kohta	31
Joonis 4. SQL päring sektorite osakaalude andmete pärimisest.....	31
Joonis 5. Investorite portfelli osakaalud sektorite lõikes aegreana kasutades Pandast ja Matplotlibi	32
Joonis 6. SQL päring valitud investorite portfelli keskmise väärtuse leidmiseks	33
Joonis 7. Valitud investorite portfelli suuruste jaotus sektordiagrammil	33
Joonis 8. Päringu tulemus aktsiate osakaalude muutustest sektorites	34
Joonis 9. Illustreeriv koodilõik GPT mudeliga liidestamisest [26]	35
Joonis 10. Kuvatõmmis LHV artikli "Top omatud aktsiad" tabelist [33]	38
Joonis 11. Kuvatõmmis LHV artikli "Top omatud aktsiad" lõigust [33]	40
Joonis 12. Kuvatõmmis GPT mudeli loodud "Top omatud aktsiad" lõigust.....	40
Joonis 13. Kuvatõmmis LHV artikli "Top omatud aktsiad" tabelist [33]	41
Joonis 14. Kuvatõmmis LHV artikli "Portfelli suurima osakaaluga" lõigust [33]	42
Joonis 15. Kuvatõmmis GPT mudeli loodud "Portfelli suurima osakaaluga" lõigust .	43
Joonis 16. Kuvatõmmis LHV artikli "Top ostud" tabelist [33]	43
Joonis 17. Kuvatõmmis LHV artikli "Top ostud" lõigust [33].....	44
Joonis 18. Kuvatõmmis GPT mudeli loodud "Top ostud" lõigust	45
Joonis 19. Kuvatõmmis sisendandme mittekorrektsest kasutamisest lõigus.....	47
Joonis 20. Kuvatõmmis Pythoni scripti loodud HTML tüüpi uudisartiklist	50

Tabelite loetelu

Tabel 1. Fails-To-Deliver tabeli atribuudid [19]	18
Tabel 2. 13F raportite andmekogumi tabelid ja nende kirjeldused	25
Tabel 3. Andmebaasi tabelid täiendavate andmete kohta.....	26
Tabel 4. GPT mudeli API parameetrid [26]	37
Tabel 5. "Top omatud aktsiad" päringu tulemused 2022. aasta esimese kvartali kohta.	39
Tabel 6. "Portfellig suurima osakaaluga" päringu tulemused 2022. aasta esimese kvartali kohta	42
Tabel 7. "Top ostud" päringu tulemused 2022. aasta esimese kvartali kohta	44
Tabel 8. GPT mudeli hinnangute tulemused	48

1 Sissejuhatus

1.1 Taust ja probleem

OpenAI GPT keelemudel on üks tänapäeva kõige edukamaid ja populaarsemaid keelemudeleid, mida kasutatakse laialdaselt erinevates rakendustes, sealhulgas erinevate tekstide koostamises ja kokkuvõtete loomises [1]. Häid näiteid GPT mudeli rakendamise kohta on võimalik tuua erinevatest valdkondadest nagu kunstimaailmast, meditsiinist või mikrobioloogiast [2] [3] [4]. Mudeli populaarsus on viimase aastaga märkimisväärselt kasvanud tänu selle erakordsetele tulemustele tekstide loomises erinevates keeltes. Sellest tulenevalt saab see olla suureks abiks paljudes valdkondades, kus on vaja automatiseerida või kiirendada tekstipõhiseid ülesandeid [5]. Teadlased ja ettevõtted on huvitatud, kuidas uued masinõppe lahendused võivad aidata luua uusi praktilisi rakendusi, mis tõstaksid praeguste lahenduste tootlikkust ja tõhusust [6] [7]. Käesoleva töö sarnased näited annavad paremat aimu GPT mudeli potentsiaalsetest ja võimetest, mis on heaks suuniseks uute lahenduste loomisel [8].

1.2 Töö eesmärk

Magistritöö eesmärk on uurida, kuidas kasutada SEC 13F raportitest koosnevat andmebaasi ja andmeanalüüsi meetodeid ning luua automaatne süsteem, mis loob kokkuvõtvaid uudisartikleid GPT-3 keelemudeli abil. Töö eesmärk on näidata, kuidas ühildada andmeanalüüsi meetodid ja keelemudelid praktiliste lahenduste loomiseks finantsvaldkonnas ning anda ülevaade sellest, kuidas selline süsteem võiks olla kasulik finantsuudiste loomisel. Käesoleva töö raames on püstitatud järgnevad uurimisküsimused:

1. Kuidas automatiseerida 13F raportite analüüs ja uudisartikli loomine?
2. Kuidas võrrelda tehisintellekti abil loodud uudisartikleid finantsanalüütiku poolt loodud uudisartiklitega ja hinnata nende usaldusväärsust ja kvaliteeti?

1.3 Töö struktuur

Käesoleva magistritöö struktuur koosneb neljast põhiosast: metoodika, töö tulemused, analüüs ja järeldused. Metoodika peatükis antakse ülevaade uurimisprotsessist, sealhulgas valitud andmekogumist, tööriistadest ja meetoditest. Töö tulemuste peatükis kirjeldatakse andmetöötlust ja -analüüsi, töö protsessi ja saavutatud töö tulemusi. Analüüsi peatükis hinnatakse kvalitatiivsete ja kvantitatiivse analüüsi tulemusi, võrreldakse neid varasemate finantsväljaannetega ning antakse ülevaade nende tulemuste piirangutest. Lõpuks, järelduste peatükis, tuuakse välja töö peamised järeldused, arutatakse nende tulemuste tähendust ja esitatakse soovitusel tulevastele lahendustele.

2 Metoodika

Metoodika peatükk annab ülevaate objekti detailsest kirjeldusest, kasutatud GPT mudelist, metoodikatest, andmetest, tööriistadest, tööprotsessi kirjeldusest ja tulemuste valideerimisest.

2.1 Objekti detailne kirjeldus

Käesolev töö koosneb SEC-i raportitest koosnevast andmebaasist, SQL päringutest ja Python skriptidest, mis ühildatuna GPT-3 mudeliga loovad uudisartiklite loomise süsteemi, mis võimaldab automatiseeritult genereerida kokkuvõtlikke uudiseid investorite portfellide, aktsiate hindade ja muude SEC raportitest kogutud andmete kohta.

2.2 GPT-3.5 mudel

GPT mudel on lühend sõnadest "Generative Pre-trained Transformer" ning on sügavõppe meetodil treenitud keelemudel, mille erakordne ennustustäpsus on saavutatud ajaloo ühe suurima tekstidest koosneva andmebaasi ehk treeningandmete abil [9]. GPT mudeli on arendanud ettevõtte OpenAI ning GPT-3.5 on GPT mudeli seeria kolmanda põlvkonna kõige uuem mudel. GPT mudelit on võimalik kasutada erinevates tekstiloomise ja tekstide mõistmise ülesannetes. Mudel suudab kirjutada inimlikult sarnast teksti ning mõista keerukaid kontekste ja suhteid tekstis, mida kasutatakse erinevate probleemide lahendamiseks, alates tekstide loomisest kuni suuremate andmekogumite analüüsini [10]. Ehkki keelemudel saab suurepäraselt hakkama lihtsamate tekstide mõistmise ülesannetes, võivad keerulisemad või spetsiifilisemad päringud põhjustada raskusi. Keerulisemate päringute puhul võivad suured keelemudelid genereerida fiktiivseid ja ebatäpseid vastuseid [11].

Vaatamata sellele, et mahukas eksperimentaalne analüüs GPT3.5 kohta näitab, et keelemudel seisab endiselt silmitsi raskustega loomuliku keele järeldamise ülesannetes, osutus GPT-3.5 käesolevas töös valituks, kuna OpenAI API toetatud mudelitest on GPT-

3.5 kõige uuem ehk olenemata GPT-4 olemasolust ei saa viimast mudelit läbi API liidestuse kasutada, mis on oluline nõue automatiseeritud lahenduse puhul [12].

GPT-3.5 seeria koosneb erinevatest mudelitest, mida kasutatakse vastavalt konkreetse ülesande keerukusele ja nõudmistele. GPT-3.5 mudelid on erinevad üksteisest suuruse, võimekuse ja ressursikasutuse poolest [10].

GPT-3.5 seeria mudelid on järgmised:

1. **gpt-3.5-turbo**: Tegemist on kõige võimekama ja kulutõhusama mudeliga GPT-3.5 seeriast. Peamiselt optimeeritud vestluse jaoks, kuid töötab hästi ka traditsiooniliste tekstiülesannete jaoks. Selle mudeli ressursikasutust võrreldes `text-davinci-003`'ga hinnatakse 1/10 [10].
2. **gpt-3.5-turbo-0301**: Mudeli *snapshot* ehk hetketõmmis pärineb 1. märtsist 2023. Erinevalt `gpt-3.5-turbo` mudelist ei saa `gpt-3.5-turbo-0301` uuendusi ja tehnilist tuge. Sellest tulenevalt openAI ei soovita kasutajatel `gpt-3.5-turbo-0301` mudelit kasutada [10].
3. **text-davinci-003**: Mudel suudab teha erinevaid keelelisi ülesandeid parema kvaliteedi ja pikema väljundiga võrreldes GPT-3 seeria mudelite `curie`, `babbage` või `ada-ga`. Samuti toetab see puudulike tekstide täiendamist [10].
4. **text-davinci-002**: Sarnane `text-davinci-003` mudeli võimekusega, kuid treenitud juhendatud õppe abil, mitte stiimulõppega [10].
5. **code-davinci-002**: Mudel on optimeeritud koodi täitmise ülesannete jaoks [10].

Käesolevas töös kasutati `text-davinci-003` mudelit, kuna see pakub parimat võimekust keerukamate tekstiloomise ja analüüsiülesannete lahendamiseks. Samuti valitud mudeli kasutamisega tagatakse tulemuste kõrgem kvaliteet ja täpsus. Alternatiivina oleks võinud valida `gpt-3.5-turbo`, kuid arvestades seda, et `gpt-3.5-turbo` on *chat*-tüüpi mudel, siis käesoleva lahenduse jaoks on sobilikum `text-davinci-003`, kus GPT poolt saadud väljundites ei pea formaati ja üleliigseid detaile eemaldama.

2.3 Desk research

Desk research on uurimismeetod, mille käigus kogutakse, analüüsitakse olemasolevat teavet erinevatest allikatest, et saada parem ülevaade uurimistest. *Desk research* hõlmab peamiselt andmete kasutamist, mis on juba olemas erinevates avalikes allikates, nagu uuringud, uudised, artiklid, aruanded, statistika ja teised avalikud andmebaasid. *Desk research* on tihti aluseks esmasele andmekogumisele.

Järgnevalt on liigitatud magistr töö etapid, mis hõlmasid *desk research* meetodit:

- Andmete kogumine: 13F raportite ja teiste uudisartiklite koostamiseks vajaminevate andmete kogumine erinevatest allikatest.
- Andmete eeltötlus: Olulise teabe eraldamine, nt. ettevõtete, aktsiate nimed, investeeringud, positsioonide muutused.
- Teadmiste omandamine: Finantsartiklite struktuuri ja nendes sisalduva teabe uurimine, et mõista, kuidas uudisartikleid luua.
- Analüüs: 13F raportite andmete analüüs, et tuvastada olulised suundumused ja investeeringute muutused raporteerinud investorite seas.

2.4 Prompt engineering

Prompt engineering viitab protsessile, kus kasutaja saab peenhäälestada sisendeid, et saavutada soovitud väljund, sageli lisades ülesannete kirjeldusi, et mudelit konkreetsemalt suunata. Seega, oma olemuselt on *prompt engineering* seotud sellega, kuidas rääkida tehisintellektiga, et see teeks seda, mida sa soovid [13].

Käesolevas magistr töö kasutatakse *prompt engineering* meetodeid, et parandada OpenAI GPT mudeli tulemusi 13F raportitest põhinevate uudisartiklite loomisel. Peamine eesmärk on aidata mudelil paremini mõista sisendandmeid ning anda asjakohaseid ja kvaliteetseid väljundeid.

Magistr töö kasutatakse *prompt engineering* meetodeid mitmes etapis:

- Promptide loomine: GPT mudeli sisendandmete eeltöötlus, kus andmeanalüüsi tulemusi muudetakse GPT mudelile loetavaks tekstiks.
- Uudisartiklite genereerimine: GPT mudelile promptide esitamine ja uudisartiklite loomine vastuseks.
- Iteratiivne täiustamine ja testimine: *Promptide* korrigeerimine ja täiustamine, et parandada mudeli tulemusi.
- Kvaliteedi hindamine ja kontroll: Genereeritud uudisartiklite kvaliteedi hindamine ja võrdlemine finantsanalüütikute poolt loodud artiklitega.

2.5 Andmed

Andmete peatükk annab ülevaate töös kasutatud andmetest, milleks on SEC-i poolt loodud kokkuvõtlikud andmekogumid ja andmed USA börsidel kaubeldavate aktsiate kohta (aktsiate sektorid, ajaloolised hinnad).

2.5.1 SEC 13F raportid

13F raportid on regulatiivsed dokumendid, mida suuremad institutsionaalsed investorid esitavad USA Väärtpaberite ja Börsikomisjonile (SEC). Need raportid on loodud selleks, et suurendada avalikkusele kättesaadavat teavet nende investorite väärtpaberiportfellide kohta. 13F vormide esitamine on kohustuslik kõigile institutsionaalsetele USA investoritele, kelle halduses on vähemalt 100 miljoni USA dollari ulatuses USA börsidel kaubeldavaid varasid. 13F raportid annavad ülevaate investorite portfelli positsioonidest, sealhulgas aktsiatest, optsioonidest ja muudest väärtpaberitest, mis on USA börsidel kaubeldavad. Raportid esitatakse kvartaalselt ning aitavad finantsturgudel osalejatel paremini mõista suuremate institutsionaalsete investorite portfelli muutusi ja nende mõju turule [14].

Antud töös kasutatakse SEC-i poolt jagatud ametlike 13F raportite andmekogumeid, mis koosnevad perioodil mai 2013 kuni märts 2023 esitatud raportitest [15]. Andmekogumid on väljavõte XML-põhistest 13F aruannetest SEC-i EDGAR süsteemist. SEC-i poolt esitatud kokkuvõtlikud andmekogumid on loodud eesmärgiga pakkuda avalikkusele

hõlpsamat viisi 13F andmete kogumiseks. Kvartaalsed andmekogumid sisaldavad sama teavet erinevate institutsionaalsete investorite aktsiaportfellide kohta, mida alternatiivina saaks pärida ka EDGAR süsteemist [16]. EDGAR (*Electronic Data Gathering, Analysis, and Retrieval*) on SEC-i elektrooniline süsteem, kuhu ettevõtted peavad esitama kõik olulised regulatiivsed dokumendid ja finantsaruanded [17]. Andmekogumite kasutus töös osutus valituks, sest EDGAR süsteemi API kasutus on reguleeritud ja puudub efektiivne viis kokkuvõtlike kvartaalsete andmete kogumiseks otse EDGAR süsteemist [18].

COVERPAGE		SUBMISSION		OTHERMANAGER		INFOTABLE	
PK	ACCESSION_NUMBER VARCHAR2(25)	PK	ACCESSION_NUMBER VARCHAR2(25)	PK	ACCESSION_NUMBER VARCHAR2(25)	PK	ACCESSION_NUMBER VARCHAR2(25)
	REPORTCALENDARORQUARTER DATE		FILING_DATE DATE		OTHERMANAGER_SK NUMBER(38)		INFOTABLE_SK NUMBER(38)
	ISAMENDMENT CHAR(1)		SUBMISSIONTYPE VARCHAR2(10)		CIK VARCHAR2(10)		NAMEOFISSUER VARCHAR2(200)
	AMENDMENTNO NUMBER(2)		CIK VARCHAR2(10)		FORM13FFILENUMBER VARCHAR2(17)		TITLEOFCLASS VARCHAR2(150)
	AMENDMENTTYPE VARCHAR2(20)		PERIODOFREPORT DATE		NAME VARCHAR2(150)		CUSIP CHAR(9)
	CONFIDENIEXPIRED CHAR(1)		UNIQUE (ACCESSION_NUMBER)		UNIQUE (ACCESSION_NUMBER, OTHERMANAGE)		VALUE NUMBER(12)
	DATEDENIEXPIRED DATE						SSHPRNAMT NUMBER(16)
	DATEREPORTED DATE						SSHPRNAMTTYPE VARCHAR2(10)
	REASONFORNONCONFIDENTIALITY VARCHAR2(25)						PUTCALL VARCHAR2(10)
	FILINGMANAGER_NAME VARCHAR2(150)						INVESTMENTDISCRETION VARCHAR2(10)
	FILINGMANAGER_STREET1 VARCHAR2(40)						OTHERMANAGER VARCHAR2(100)
	FILINGMANAGER_STREET2 VARCHAR2(40)						VOTING_AUTH_SOLE NUMBER(16)
	FILINGMANAGER_CITY VARCHAR2(30)						VOTING_AUTH_SHARED NUMBER(16)
	FILINGMANAGER_STATEORCOUNTRY CHAR(2)						VOTING_AUTH_NONE NUMBER(16)
	FILINGMANAGER_ZIPCODE VARCHAR2(10)						UNIQUE (ACCESSION_NUMBER, INFOTABLE_SK)
	REPORTTYPE VARCHAR2(30)						
	FORM13FFILENUMBER VARCHAR2(17)						
	PROVIDEINFORINSTRUCTIONS CHAR(1)						
	ADDITIONALINFORMATION VARCHAR2(4000)						
	UNIQUE (ACCESSION_NUMBER)						

Joonis 1. SEC 13F raportite andmekogumis olevad tabelid ja atribuudid [15]

2.5.2 Fails-to-Deliver andmekogum

Fails-to-Deliver (edaspidi FTD) andmekogum on kogum teabest, mis sisaldab infot väärtpaberitehingute tõrgetest Ameerika Ühendriikide börsidel. Andmed on kogutud perioodil veebruar 2004 kuni märts 2023. FTD andmekogumist leiab seose aktsiate CUSIP- numbrite ja nendega seotud *ticker*-ite kohta, mida kasutatakse käesolevas magistris töös 13F raportites esitatud aktsiate CUSIP- numbrite sidumiseks nendele määratud *ticker*-itega. Käesoleva töö raames kasutati FTD andmeid osaliselt ehk vajaminevateks atribuutideks olid SETTLEMENT DATE, CUSIP ja SYMBOL (vt Tabel 1) [19].

Field Name	Field Description	Maximum Size
SETTLEMENT DATE	SETTLEMENT DATE	Number - 8 digits
CUSIP	CUSIP	9 characters
SYMBOL	TICKER SYMBOL	10 characters

QUANTITY (FAILS)	TOTAL FAILURE-TO-DELIVER SHARES	Number - unlimited
DESCRIPTION	COMPANY NAME	30 characters
PRICE	CLOSING PRICE ON PREVIOUS DAY	Number - unlimited

Tabel 1. Fails-To-Deliver tabeli atribuudid [19]

2.5.3 USA börsidel kaubeldavad aktsiad

Käesolevas töös oli oluline lisada USA börsidel kaubeldavate aktsiate ajaloolised hinnad, et saada täpne ülevaade aktsiate hinnamuutustest erinevate kvartalite lõikes. Lisaks oli töös vajalik lisada andmed aktsiate sektorite kohta, mis võimaldasid aktsiaid sektoripõhiselt liigitada ja analüüsida investeerimistrende erinevates majandussektorites. Selline sektoripõhine analüüs aitab mõista, millised sektorid on raporteerinud institutsionaalsete investorite jaoks atraktiivsed, ning annab parema ülevaate turudünaamikast ja investeerimisstrateegiatest. Andmed USA börsidel kaubeldavate aktsiate hindade ja sektorite kohta pärinevad Yahoo Finance finantsportaalist ning kasutatud on andmeid ajaperioodi detsember 2013 kuni september 2022 kohta [20].

2.6 Tööriistade kirjeldus

2.6.1 SQLiteStudio

SQLiteStudio on vaba ja avatud lähtekoodiga andmebaasi haldustarkvara, mis võimaldab luua, korrastada, hallata ja vaadata SQLite andmebaase. Kasutades SQLiteStudio-t, loodi magistritöö jaoks vajalik andmebaas ning imporditi 13F raportite tabelid ja teised analüüsiks vajaminevad andmed (vt jaotist 2.3). Andmebaasi haldamine ja andmeanalüüsiks kasutatud SQL päringute loomine toimus suures osas SQLiteStudio liidese abil [21].

2.6.2 Python

Python on programmeerimiskeel, mida kasutatakse laialdaselt andmetöötles ja tehisintellekti rakendustes. Käesoleva töö jaoks kirjutati Python skriptid, mis võimaldasid näiteks andmete importimist csv failidest andmebaasi või andmete eeltöötlemise läbiviimist. Samuti kirjutati GPT mudeli API-ga integreeritud lahendus Python keeles [22].

2.6.3 Visual Studio Code

Visual Studio Code on vaba ja avatud lähtekoodiga koodiredaktor, mida kasutatakse laialdaselt erinevate programmeerimiskeelte arendamisel. Magistritöös kasutati Visual Studio Code'i Python skriptide debugimiseks ja käivitamiseks. Visual Studio Code pakub kasulikke tööriistu ja lisarakendusi, mis aitavad tõrkeotsingut, koodi vormindamist ja skriptide käivitamist, muutes koodi loomise ja testimise protsessi lihtsamaks ja tõhusamaks [23].

2.6.4 Yahoo_fin

Yahoo_fin on Pythoni teek, mis on loodud Yahoo Finance'i veebisaidilt finantsandmete, nagu aktsiahindade, finantsaruannete ja optioonide, hankimiseks. See pakub lihtsat ja tõhusat viisi Yahoo Finance'i platvormil olevate andmete programmiliseks kasutamiseks. Yahoo_fin põhineb MIT litsensil, mis tähendab, et seda võib kasutada nii akadeemilistes kui ka kommertsprojektides, tingimusel et järgitakse litsentsi tingimusi [24].

Käesolevas töös kasutatakse Yahoo_fin teeki aktsiate ajalooliste hindade pärimiseks. Selle teegi abil on võimalik hankida aktsiahindade ajalugu erinevate perioodide ja ajavahemike kohta. Töös kasutatud aktsiahinnad on kuised, mis näitavad valitud aktsia kuu keskmist hinda [25].

2.6.5 GPT mudeli API

GPT mudeli API on OpenAI poolt pakutav liides (API), mis võimaldab arendajatel kasutada GPT mudelit, nagu text-davinci-003, oma rakendustes. API abil on võimalik saada rakendusele ligipääs GPT mudeli funktsionaalsustele, mis käesoleva töö raames tähendab kokkuvõtlike lõikude loomist. API endpoint, mida käesolevas töös kasutatakse, on <https://api.openai.com/v1/completions>, mis koos autentimisvõtmega võimaldab teha päringuid GPT mudelile, et saada vastu genereeritud lõigud tervikliku uudisartikli loomiseks [26].

2.6.6 Pandas

Pandas on avatud lähtekoodiga Pythoni teek, mis pakub erinevaid andmestruktuure, et muuta andmetöötlus ja analüüs lihtsaks ja efektiivseks. Pandas'i peamine andmestruktuur on DataFrame, mis võimaldab töötada tabeliliste andmetega ja pakub erinevaid funktsioone nende andmetega töötamiseks [27]. Pandas on loodud andmete töötlemiseks

ja analüüsiks ning sobib suurepäraselt andmete puhastamiseks, transformeerimiseks ja visualiseerimiseks [28].

2.6.7 Matplotlib

Matplotlib on avatud lähtekoodiga Pythoni teek, mis on loodud andmete visualiseerimiseks. Matplotlib pakub laia valikut graafikatüüpe, nagu joon-, punkt-, tulp- ja sektordiagrammid, mis võimaldavad andmete visuaalset esitamist ja analüüsi. Käesolevas töö ühe näitena kasutati Matplotlib-i, et visualiseerida 100% virtustatud tulpdiaagrammi kujul sektorite osakaale erinevate kvartalite lõikes, mis annab hea ülevaate, kuidas erinevatel perioodidel on investorite portfellides investeerimistrendid muutunud [29].

2.7 Tööprotsessi kirjeldus

Magistritöö protsess algab andmehankemeetodite valimise ja andmete kogumise protsessi planeerimisega. Seejärel kogutakse vajalikud 13F raportite andmed ja lisainfo aktsiate kohta ning viiakse läbi andmete puhastamine ning eeltöötlemine. Selles etapis valitakse täpsem andmeanalüüsi meetoodika, mis sobib 13F raportite analüüsimiseks.

Järgnevalt viiakse läbi andmeanalüüs, tõlgendatakse analüüsitulemusi ja võrreldakse testimises kasutatavate artiklite tabelitega. Pärast andmeanalüüsi ja tulemuste tõlgendamist liidestatakse valitud tehisintellekti lahendus rakendusega, et genereerida uudisartikleid analüüsitulemuste põhjal. Kirjeldatud sammud kuni andmeanalüüsini kuuluvad kasutatud meetoodikate seast *desk researchi* alla. Liidestamine ja mudeli abil tekstide genereerimine järgivad *prompt engineering* meetoodika põhimõtteid, et luua *promptid*, mis tagavad võimalikult kvaliteetsed ja asjakohased lõigud uudisartiklites kasutamiseks.

Andmeanalüüsile järgneb valideerimine, kus kasutatakse erinevaid meetodeid, et tagada tehisintellekti abil loodud uudisartiklite kvaliteet ja usaldusväärsus. Käesoleva töö valideerimine protsessilt kuulub nii *desk researchi* kui ka *prompt engineeringu* meetoodikate alla, kuna see hõlmab nii sisendi kujundamist ja optimeerimist kui ka võrdleva analüüsi läbiviimist.

Viimaks tõlgendatakse tulemusi ja arutletakse tehisintellekti rakendamise üle 13F raportite analüüsimisel ning uudisartiklite loomisel, mis kuulub *desk researchi* alla. Kogu protsessi jooksul keskendutakse magistritöö eesmärkide saavutamisele, mis on automatiseeritud finantsandmete analüüs ja analüüsitulemuste kuvamine ilma finantsanalüütikut kaasamata.

2.8 Tulemuste valideerimine

Tulemuste valideerimise eesmärk on hinnata automaatsete uudisartiklite rakenduse tulemuslikkust ja kvaliteeti. Kuigi GPT mudeli usaldusväärsus on vähe uuritud, siis järgnevalt tuuakse välja peamised meetodid, mida kasutatakse tulemuste valideerimiseks [30]. Tulemuste valideerimine koosneb kolmest osast:

1. Rakenduse poolt loodud uudisartiklite võrdlemine ajalooliste väljaannetega (LHV artikkel). Rakenduse poolt loodud uudisartikleid võrreldakse ajalooliste väljaannetega, et hinnata nende sisu, struktuuri ja keelekasutuse sarnasust.
2. Loodud lõikude analüüsimine sisendandmete vastu. See osa keskendub loodud lõikude analüüsimisele, et kontrollida, kas nende sisu vastab sisendandmetele. See hõlmab kahte alamosa:
 - a. Kas lõigus kirjeldatud andmed on vastavuses sisendandmetega? See analüüs võimaldab hinnata mudeli tulemuslikkust andmete töötlemisel ja nende esitamisel uudisartiklites.
 - b. Kas on GPT poolt lisatud andmeid, mida sisendandmetega ei esitatud? Selles osas kontrollitakse, kas GPT mudel lisab andmeid, mida sisendandmetega ei esitatud, ning hinnatakse, kas need lisatud andmed on kontekstis asjakohased ja kas need parandavad artikli kvaliteeti.
3. Loodud uudisartiklite analüüsimine ja hindamine tehisintellekti poolt. Selles osas kasutatakse tehisintellekti, et analüüsida loodud uudisartiklite sisu, keelekasutust ja struktuuri, et hinnata nende kvaliteeti ja sobivust eesmärgipäraseks kasutamiseks. Tehisintellekt aitab tuvastada võimalikke kitsaskohti ja pakkuda tagasisidet rakenduse täiustamiseks.

On uuritud GPT mudeli kasutamise võimalusi ja efektiivsust hinnangute andmisel, näidates, et GPT hinnangud võivad olla täpsed ja usaldusväärsed pakkudes väärtuslikku tuge näiteks AES tehnoloogias. AES (*Automated essay scoring*) on tehnoloogia, mida kasutatakse kirjutatud esseele automaatse hinde andmiseks [31].

3 Töö tulemused

3.1 Andmete kogumine ja eeltöötlus

Andmete kogumise protsess algas erinevate andmekogumite otsimisega. Peamiseks andmekogumiks olid varasemalt mainitud 13F raportid, mis sisaldavad informatsiooni institutsionaalsete investorite aktsiaportfellide kohta kvartalite lõikes. Andmekogum hõlmas ajavahemikku mai 2013 kuni märts 2023. Lisaks 13F raportitele kasutati Fails-To-Deliver andmekogumit, mis on samuti SEC-i loodud andmekogum pakkumaks USA börside kohta lisainfot. Fails-To-Deliver andmekogumist saadi puuduolevad CUSIP-Ticker seosed.

Andmete eeltöötlus algas andmekogumite importimisega andmebaasi. Andmebaasi loodud tabelid võtsid arvesse andmekogumi dokumentatsioonis esitatud infot tabelite ja tabelis olevate atribuutide kohta. Pärast andmete importimist tuli andmebaasi sisestatud andmeid valideerida, et oleks täidetud kõik eeldused andmeanalüüsiga alustamiseks. See tähendas veendumist, et andmed on soovitud formaadis, korrastatud ja ilma vigadeta.

Lisaks 13F raportite andmekogumile koguti finantsportaalist Yahoo Finance andmeid aktsiate ajalooliste hindade ja nende sektorite kohta. Selleks oli vaja saada raporteeritud aktsiate tickerid, mille kohta 13F raportite andmekogumis info puudus, kuid mida oli võimalik leida Fails-To-Deliver andmekogumist. Tänu Fails-To-Deliver-st leitud CUSIP-Ticker seosele oli võimalik lihtsamalt kasutada vabavarana kättesaadavat Pythoni teeki `Yahoo_fin`, kust tickeri abil on lihtne leida iga aktsia kohta infot selle ajalooliste hindade või sektori kohta [25]. Pärast 13F raportite ja Fails-To-Deliver andmekogumi importimist andmebaasi eksporditi andmebaasist unikaalsed tickerid, mille abil Yahoo Finance finantsportaalist vajaminevad hinnad ja sektorid koguti ja andmebaasi lisati. Eelneva protsessi käigus kasutati erineval moel Pythoni skripte, mis samuti kogusid ka aktsiate info Yahoo Finance'ist ja importisid kogutud andmed andmebaasi.

Kokkuvõtvalt koosnes andmete kogumine ja eeltöötlus erinevate andmekogumite integreerimisest, andmete valideerimisest ja lisainformatsiooni kogumisest Yahoo Finance portaalist [20]. Selle protsessi tulemusel valmis andmebaas vajaminevate andmetega, mis oli valmis edasiseks analüüsiks.

3.2 Andmebaasi struktuur

Andmebaasi struktuur koosneb 13F raportite andmekogumist, Yahoo Finance-st saadud aktsiate sektoritest ja ajaloolistest hindadest ning andmeanalüüsi päringute efektiivsemat käivitamist toetavatest vahetabelitest.

3.2.1 13F raportite andmekogum

13F raportite andmekogumi tabelid annavad põhjaliku ülevaate 13F aruannetest, sisaldades teavet insitutsionaalsete investorite portfelliidest, otsestest ja kaudsetest juhtidest ning allkirjastajatest. Iga andmekogumi tabel omab primaarvõtit nimega „ACCESSION_NUMBER“, mis seostab üle tabelite esitatud investori kirjed. 13F raportite andmekogum koosneb järgmistest tabelitest:

Tabeli nimi	Kirjeldus
SUBMISSION	Tabel sisaldab üldist infot esitatud dokumendi ja selle esitaja kohta.
COVERPAGE	Tabelist leiab rohkem detaile aruande ja selle esitaja kohta. On võimalik näha, kas tegemist täiendava või parandusi sisaldav aruandega. Lisaks on välja toodud institutsionaalse investori nimi, asukoht ja vajadusel investori poolt jagatud lisainformatsioon esitatu kohta.
OTHERMANAGER	Tabel sisaldab seotud juhtide loendit, kes omavad esitaja portfelli varasid ning kuuluvad raporteeritavate hulka.
SIGNATURE	Tabel sisaldab allkirjastaja nime, ametinimetust, telefoninumbrit, allkirja, linna, osariiki või riiki ning allkirja kuupäeva.
SUMMARYPAGE	Tabel sisaldab kokkuvõtlikut infot, sealhulgas teiste kaasatud juhtide arvu, tabeli kirjete kogusummat ja tabeli väärtuse kogusummat. Samuti näitab, kas konfidentsiaalsed varad on välja jäetud.

OTHERMANAGER 2	Tabel sisaldab seotud juhtide loendit, kes on esitatud investori portfelliga seotud, koos nende järjekorranumbri, CIK-i, 13F failinumbri ja nimega.
INFOTABLE	Tabel sisaldab esitatud portfelli andmeid ja nendes olevaid varasid, sealhulgas emitendi nime, klassi pealkirja, CUSIP numbrit, turuväärtust, aktsiate arvu, varaklassi tüüpi, optiooni tüüpi (PUT/CALL), investeerimisotsuste tegemise õigust, teiste juhtidega jagatud otsustusõigust ning hääleõiguse jagamist.

Tabel 2. 13F raportite andmekogumi tabelid ja nende kirjeldused

3.2.2 Täiendavad andmed

Käesolevas töös kasutatakse täiendavaid andmeid, mis pärinevad kolmest allikast: SEC-i koostatud Fails-To-Deliver andmekogum, Yahoo Finance portaalist kogutud lisainformatsioon raporteeritud aktsiate kohta ja lisainformatsioon dataroma nimetatud superinvestorite kohta [19] [20] [32].

Tabeli nimi	Kirjeldus
CNS	Fails-To-Deliver andmekogumi tabel, mis sisaldab informatsiooni nurjunud aktsiatehingute kohta, kuid käesoleva töö kontekstis on antud tabelist oluline CUSIP-SYMBOL vaheline seos. Tabel koosneb atribuutidest SETTLEMENT_DATE, CUSIP, SYMBOL, QUANTITY, DESCRIPTION, PRICE.
PRICES	Yahoo Finance portaalist kogutud andmed aktsiate ajalooliste hindade kohta. Tabel koosneb atribuutidest SYMBOL, DATE ja CLOSE.
SYMBOL_SECTOR	Yahoo Finance portaalist kogutud andmed aktsiate sektorite kohta. Tabel koosneb atribuutidest SYMBOL ja SECTOR.

SUPERINVESTORS	Dataroma 2023 aasta veebruari nimekirja alusel on loodud tabel SUPERINVESTORS, mis defineerib investorid, kelle portfelle käesolevas töös uuritakse. Tabel koosneb atribuutidest FILINGMANAGER_NAME, CIK ja SUPERINVESTOR (Väärtuse 0 korral välistame investori skoobist. Väärtuse 1 korral loeme valitud investori superinvestoriks).
----------------	---

Tabel 3. Andmebaasi tabelid täiendavate andmete kohta

3.3 Andmeanalüüs

Käesoleva peatüki eesmärk on luua andmebaasile edastatavad päringud, millega on võimalik saada kvartaalseid ülevaateid investorite portfellidest. See hõlmab erinevaid võimalusi investeerimistrendide uurimisest ja muude aspektide analüüsimisest, mis võiksid olla väikeinvestoritele või finantsmaailma huvilistele heaks sisendiks oma investeerimisotsuste tegemisel või finantsmaailma paremaks mõistmiseks.

Esimeste päringute motivatsioon tulenes LHV finantsportaali 25. mai 2022 väljaandest „USA superinvestorite hiljutised tehingud“, kus kajastatud tabelid tulenevad dataroma loodud 13F raportite analüüsist [33] [34]. Eesmärk oli luua soodsad tingimused eksperimendiks, kus GPT mudeli loodud kokkuvõtlike lõike oleks võimalik võrrelda LHV väljaandes olevate lõikudega.

Teostatud päringud on loodud sarnaselt, et oleks võimalik muuta nendes olevat perioodi. LHV artikli tabelitega võrdlemiseks me kasutame kuupäeva 31. märts 2022, kuid teiste perioodide kohta käivitamiseks on *hard-coded* väärtuse asemel kasutatud muutujat, mida kasutajal on võimalik valida. Muutuja defineeritakse *Python*-i koodis, mis käivitab päringuid ning on liidestatud GPT mudeliga.

3.3.1 LHV artikli tabel „Top omatud aktsiad“

LHV artiklis viidatud dataroma loodud tabel „Top omatud aktsiad“ toob esile 10 kõige populaarsemat aktsiat valitud superinvestorite seas. Tabelil on veerud, mis kirjeldavad iga rea kohta aktsia nime, sümbolit, mitu investorit aktsiat oma portfellis omavad, milline investor omab suurimat osakaalu valitud aktsiaga ja milline on „YTD“ ehk Year to Date,

mis näitab aktsia hinnamuutust aasta algusest kuni valitud päevani. LHV artikli tabelis esitatud andmete põhjal on näha, et kasutatud on valitud päeva puhul artikli avalikustamisele lähedast päeva. Oluline on mainida, et artiklis kajastatud superinvestorite all mõeldakse institutsionaalseid investoreid, mille nimekirja on koostanud dataroma. Kahjuks ei leidu infot, mis kajastaks dataroma investorite nimekirja 25. mai 2022 seisuga, mistõttu käesolevas töös valiti superinvestoriteks dataroma esitatud nimekiri veebruari 2023 seisuga [32].

3.3.2 „Top omatud aktsiad“ päringu tehniline teostus

Tabeli "Top omatud aktsiad" tulemuste saavutamiseks kasutatav SQL päring on loodud mitmete tabelite liitmise ning erinevate arvutuste abil, et esitada superinvestorite kümme enimomatud aktsiat valitud perioodil.

Esiteks alustatakse *infotable* tabelist ning liidetakse see teiste asjakohaste tabelitega: *coverpage*, *submission*, *cns* ja *superinvestors*. See liitmine on vajalik, et pääseda ligi kõikidele vajalikele andmetele, mida „Top omatud aktsiad“ tabeli loomiseks vaja läheb. Päringu skoop ehk analüüsitava investori portfelli on määratud tabeliga *superinvestors*. Päringu tulemused rühmitatakse CUSIP-i järgi, mis on unikaalne identifikaator iga aktsia jaoks, et näidata iga aktsia kohta ainult ühte rida.

Päring arvutab iga aktsia kohta kolm olulist väärtust:

- ***investor_count***, mis näitab investorite arvu, kes omavad aktsiat. See arv ühtlasi tõlgendab ka aktsia populaarsust valitud investorite seas.
- ***Largest share of the portfolio***, mis selgitab, milline investor omab suurimat osakaalu valitud aktsiast oma portfellis. Selle väärtuse leidmiseks kasutatakse alampäringut, mis arvutab iga superinvestori portfelli osakaalu valitud aktsia kohta ja valib sellest suurima.
- ***YTD***, mis näitab aktsia hinnamuutust alates valitud perioodi aasta algusest kuni valitud kuupäevani. Selle arvutamiseks kasutatakse alampäringut, mis leiab aktsia hinna valitud perioodil ja jagab selle eelmise aasta lõpu hinnaga. Seejärel lahutatakse sellest tulemusest üks ning korrutatakse saadud väärtus 100-ga, et saada hinnamuutus protsentides.

Lõpuks järjestatakse tulemused investorite arvu (*investor_count*) järgi kahanevas järjekorras ning piiratakse tulemusi esimese 10 reaga.

3.3.3 LHV artikli tabel „Portfellig suurima osakaaluga“

LHV artiklis viidatud tabel "Portfellig suurima osakaaluga" toob esile 10 aktsiat, mis moodustavad valitud superinvestorite portfelliges suurima osakaalu. Tabeli veerud kirjeldavad iga rea kohta väärtpaberi nime, sümbolit, hinnamuutust, osakaalu protsenti portfellig ja investorite arvu, kes aktsiat omavad.

3.3.4 „Portfellig suurima osakaaluga“ päringu tehniline teostus

Tabeli "Portfellig suurima osakaaluga" tulemuste saavutamiseks kasutatav SQL päring on loodud mitmete tabelite liitmise ning erinevate arvutuste abil, et esitada investorite portfelliges suurima osakaalu moodustavad aktsiad. Antud päringuga leiame investorid, kes hoiavad suurt osa oma portfellig ühes aktsias või aktsiad, millest suurem osa turul kuulub ühele investorile.

Tehnilist teostust alustatakse *infotable* tabelist ning liidetakse seejärel teiste asjakohaste tabelitega, milleks on *coverpage*, *submission*, *cns*, *superinvestors* ja *largest_share*. Eelmise peatüki päringust erineb käesoleva päring selle poolest, et päringu efektiivsemaks käivitamiseks on loodud vahetabel *largest_share*, kuhu on loodud kokkuvõtlik tabel investorite portfelliges ja protsendist, palju iga portfellig olev aktsia investori portfellig moodustab.

Sarnaselt eelneva „Top omatud aktsiad“ päringule on skoobiks ainult valitud investorid. Tulemused rühmitatakse CUSIP-i järgi, järjestatakse vahetabelist *largest_share* saadud protsendi järgi kahanevalt ning piiratakse tulemusi 10 reaga.

Päringu arvutatud neli peamist ja olulisemat väärtust:

- *investor_count*, mis on teostatud sarnaselt „Top omatud aktsiad“ päringule.
- *% of portfolio*, mis näitab protsenti, palju valitud aktsia superinvestori portfellig moodustab. Tegemist on väärtusega, mis on eelnevalt vahetabelis *largest_share* arvutatud.

- *portfolio_manager* on alampäringuga leitud atribuut, mis kuvab portfelli halduri nime, kes omab suurimat osakaalu valitud aktsiat oma portfellis.
- *YTD*, mis on teostatud sarnaselt „Top omatud aktsiad“ päringule.

3.3.5 LHV artikli tabel „Top ostud 2022. aasta esimeses kvartalis“

LHV artiklis viidatud tabel "Top ostud 2022. aasta esimeses kvartalis" esitab 10 enimostetud aktsiat valitud investorite poolt 2022. aasta esimese kvartali jooksul. Tabeli veerud kirjeldavad iga rea kohta väärtpaberi nime, sümbolit, hinnamuutust, ostetud aktsiate arvu ja investorite arvu, kes aktsiat omavad.

3.3.6 „Top ostud 2022. aasta esimeses kvartalis“ päringu tehniline teostus

Tabeli "Top ostud 2022. aasta esimeses kvartalis" tulemuste saavutamiseks kasutatav SQL päringu peamine loogika peitub selles, et võrreldakse investori valitud perioodi portfellis olevate aktsiate koguväärtust ning võrreldakse seda eelneva kvartali koguväärtusega. Kui valitud perioodi aktsiate koguväärtus, millest on lahutatud hinnamuutus alates eelmisest kvartalis, on suurenenud võrreldes eelneva perioodiga, siis järeldatakse, et investor on aktsiat juurde ostnud. Hinnamuutusega arvestamine on oluline, sest kui aktsia väärtus on kvartaliga kasvanud ja seda arvutuses võrreldes eelmise perioodiga maha ei lahutata, siis päringu silmis oleks tegemist ostuga.

Päringu peamine arvutus tehakse alampäringu abil, millega arvutatakse aktsiate ostude, müükide ja muutumatute investeeringute hulk valitud perioodil. Pärast aktsia ostude arvu leidmist lisatakse vahetabeli *symbol_nameofissuer* abil lõpptulemuses kuvamiseks aktsia nimi ja summeeritakse alampäringus leitud ostude read. Atribuudi *YTD* teostus on tehtud sarnaselt eelnevatele päringutele. Tabeli tulemused on järjestatud ostjate arvu järgi kahanevalt ning piiratud esimese kümne reaga.

3.3.7 Sektorite osakaalud investorite portfellides aegreana

Käesoleva töö praktilise väärtuse tõstmiseks ja GPT mudeliga eksperimenteerimiseks on loodud päring, millega on võimalik kuvada aegreana valitud investorite portfellides olevaid osakaale aktsiatele omaste sektorite lõikes. Päring annab ülevaate investorite investeerimistrendidest ja nende muutustest ajas.

Tehniline teostus hõlmab mitut abistava eesmärgiga loodud vahetabelit. Kaks peamist vahetabelit on *cik_symbol_sector_value* ja *grouped_quarters*.

- *cik_symbol_sector_value* tabel on loodud, et salvestada iga aktsia valitud aktsiate koguväärtus investorite portfellides kvartalite lõikes. Tabel on ühendatud teiste tabelitega, nagu *cusip_symbol*, *symbol_sector*, *grouped_quarters*, ja *prices* (keskmise hinna arvutamiseks).

cik_symbol_sector_value tabeli loomiseks andmebaasis kasutatakse järgmist SQL päringut:

```
create table cik_symbol_sector_value as
SELECT
    curr.periodofreport, curr.cik, cs.symbol, ss.sector,
    sum(curr.sharesamt)*p.avg_price selected_shares_total_val
FROM grouped_quarters curr
JOIN cusip_symbol cs on curr.cusip = cs.cusip
join symbol_sector ss on cs.symbol = ss.symbol
join
    (SELECT symbol, AVG(close) as avg_price
    FROM prices
    WHERE date > '2022-09-30' AND date < '2022-12-31'
    GROUP BY symbol)
    p on p.symbol = cs.symbol
WHERE 1=1
AND curr.PERIODOFREPORT = '2022-12-31'
group by curr.periodofreport, curr.cik, cs.symbol, ss.sector;
```

Joonis 2. SQL päring vahetabeli *cik_symbol_sector_value* loomiseks 2022 aasta neljanda kvartali kohta

Vahetabel *grouped_quarters* on loodud, et kombineerida ja grupeerida andmeid tabelitest *INFOTABLE*, *SUBMISSION* ja *SUPERINVESTORS*. Tabeli eesmärk on luua andmestruktuur, mis sisaldab perioodi, investorite CIK numbreid, aktsiate CUSIP- numbreid ja nende koguhulka. *INFOTABLE*-s esitatud read võivad ühe investori portfellis olevat aktsiat kajastada mitmel real, mistõttu *grouped_quarters* vahetabel aitab grupeerida sarnased näited ning muuta sektorite osakaalude päring efektiivsemaks (käivitusaeg lüheneb).

Vahetabeli *grouped_quarters* loomiseks andmebaasis kasutatakse järgmist SQL päringut:

```

create table grouped_quarters as
Select
  SUB.periodofreport,
  sub.cik,
  it.cusip,
  sum(IT.SSHPRNAMT) sharesamt
FROM INFOTABLE IT
JOIN SUBMISSION SUB ON IT.ACESSION_NUMBER = SUB.ACESSION_NUMBER
JOIN SUPERINVESTORS SUPER on SUB.CIK = SUPER.CIK
AND SUPER.SUPERINVESTOR = 1
WHERE 1=1
GROUP BY IT.CUSIP, SUB.CIK, SUB.periodofreport;

```

Joonis 3. SQL päring vahetabeli grouped_quarters loomiseks kõikide kvartalite kohta

Järgnevas näites, kus päriti sektorite osakaalude andmed perioodi 2021 kolmas kvartal kuni 2022 teine kvartal, kasutati SQL päringut:

```

SELECT
  t1.PERIODOFREPORT Selected_Period,
  t1.sector,
  ROUND(SUM(t1.selected_shares_total_value) / t2.all_sector_total * 100.00, 3) Selected_Period_Percentage
FROM cik_symbol_sector_value t1
LEFT JOIN (
  SELECT
    periodofreport,
    SUM(selected_shares_total_value) all_sector_total
  FROM
    cik_symbol_sector_value
  GROUP BY 1
) t2 ON t1.periodofreport = t2.periodofreport
WHERE t1.periodofreport IN ('2021-12-31', '2022-06-30', '2021-09-30', '2022-03-31')
GROUP BY 1,2;

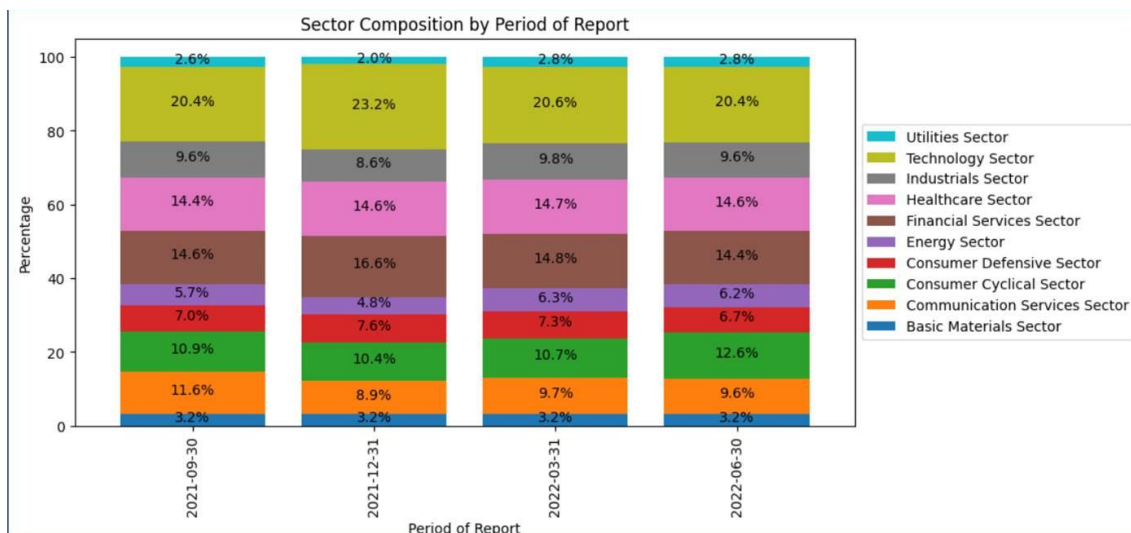
```

Joonis 4. SQL päring sektorite osakaalude andmete pärimisest

Pythoni koodilõiguga kasutati Pandas ja Matplotlib teekide võimalusi, et luua 100% virnastatud tulpdiagramm eelnevalt kogutud andmetest. Pandas teek võimaldas andmete hõlpsat laadimist, töötlemist ja ümberkorraldamist, muutes andmed graafikus

kasutamiseks sobivaks. Matplotlib teek võimaldas diagrammi loomist, kujundamist ning selle interaktiivsete elementide, nagu legendi ja protsendimärgiste, lisamist [29] [28].

Pandas ja Matplotlib teekide kasutamine andis paindlikkuse andmete töötlemisel ja visualiseerimisel, muutes keerukate diagrammide loomise lihtsaks ja efektiivseks.



Joonis 5. Investorite portfelli osakaalud sektorite lõikes aegreana kasutades Pandast ja Matplotlibi

3.3.8 Investorite portfelli suurused

Selleks, et mõista analüüsitava investorite portfelli suuruseid, on loodud päringud, millega on võimalik näha iga valitud investorite portfelli suurust kui ka kõikide investorite keskmist portfelli väärtust.

Päring, millega leida valitud perioodi kohta superinvestoriteks nimetatud investorite portfelli keskmine väärtus:

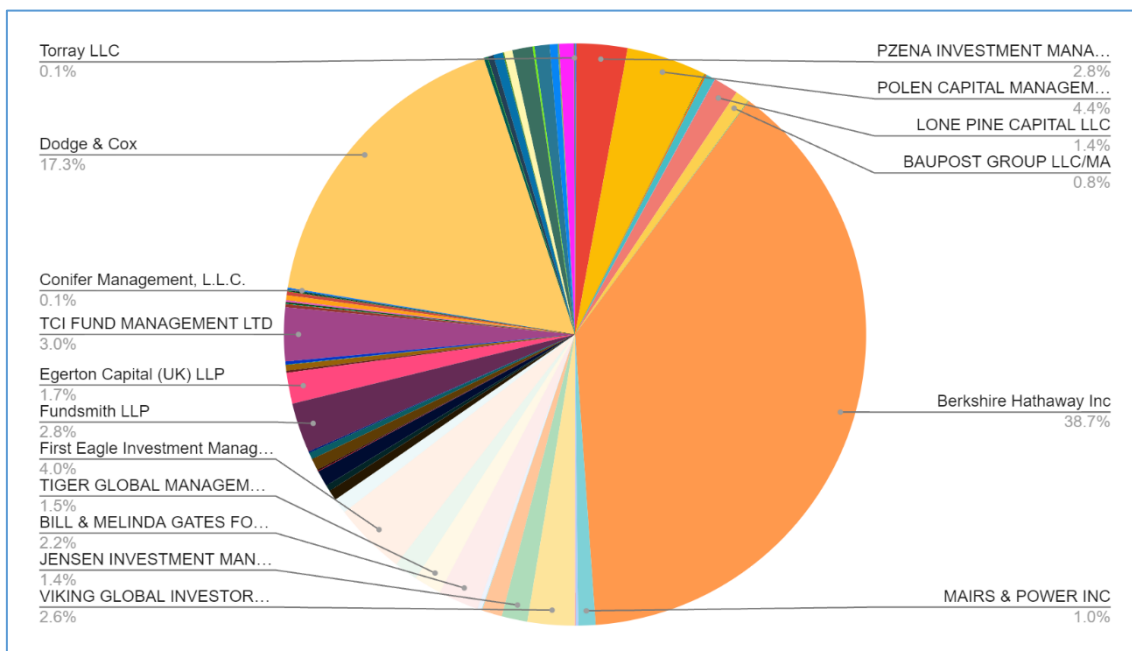

```

SELECT PERIODOFREPORT
      ,COUNT(*) SELECTED_INVESTORS_COUNT
      ,AVG(PORTFOLIO_VALUE_BILLIONS)
FROM
(
  SELECT CSSV.PERIODOFREPORT
        ,CSSV.CIK
        ,CN.FILINGMANAGER_NAME
        ,ROUND( SUM( SELECTED_SHARES_TOTAL_VALUE) / 1000000000.00,2 ) PORTFOLIO_VALUE_BILLIONS
  FROM CIK_SYMBOL_SECTOR_VALUE CSSV
  JOIN SUPERINVESTORS SUPER
  ON SUPER.CIK = CSSV.CIK AND SUPER.SUPERINVESTOR = 1
  LEFT JOIN CIK_NAMES CN
  ON CN.CIK = CSSV.CIK
  WHERE 1 = 1
  AND PERIODOFREPORT = '2022-06-30'
  GROUP BY 1
         ,2
         ,3
) A
GROUP BY 1;

```

Joonis 6. SQL päring valitud investorite portfelli keskise väärtuse leidmiseks

Kui soovitakse terviklikku ülevaadet kõikide valitud investorite portfelli väärtustest, siis tuleks kasutada eelneval joonisel olevat alampäringut.



Joonis 7. Valitud investorite portfelli suuruste jaotus sektordiagrammil

3.3.9 Aktsiate osakaalud sektorite lõikes

Käesoleva töö raames on loodud päring, millega on võimalik analüüsida investorite eelistusi sektorite lõikes ning tuvastada aktsiad, mis on muutunud populaarsemaks või ebasoosituks, võrreldes nende osakaalu terves sektoris eelneva kvartaliga. Päringu eesmärk on aidata mõista, millised aktsiad olid valitud perioodil investorite jaoks atraktiivsed või ebasoositud, pakkudes väärtuslikku teavet investorite käitumise ja investeerimistrendide kohta.

	Previous_Period	Selected_Period	sector	symbol	Previous_Period_Stock_Percentage_In_Sector	Selected_Period_Stock_Percentage_In_Sector	diff3
1	2021-12-31	2022-03-31	Healthcare Sector	DNA	7.465	0.328	-7.137
2	2021-12-31	2022-03-31	Communication Services Sector	CHTR	18.268	13.538	-4.73
3	2021-12-31	2022-03-31	Basic Materials Sector	AVTR	6.589	2.34	-4.249
4	2021-12-31	2022-03-31	Energy Sector	HES	2.611	0.04	-2.571
5	2021-12-31	2022-03-31	Industrials Sector	NDSN	2.418	0.001	-2.417
6	2021-12-31	2022-03-31	Healthcare Sector	HUM	2.526	0.135	-2.391
7	2021-12-31	2022-03-31	Communication Services Sector	TWLO	2.261	0.083	-2.178
8	2021-12-31	2022-03-31	Basic Materials Sector	ECL	10.341	8.168	-2.173
9	2021-12-31	2022-03-31	Consumer Defensive Sector	EL	9.474	7.409	-2.065
10	2021-12-31	2022-03-31	Utilities Sector	PCG	26.011	23.95	-2.061

Joonis 8. Päringu tulemus aktsiate osakaalude muutustest sektorites

Päringu tulemustes on keskne veerg diff3, mis kujutab aktsia osakaalu muutust sektoris eelneva kvartaliga võrreldes. Kui diff3 väärtus on positiivne, tähendab see, et aktsia osakaal sektoris on suurenenud, viidates sellele, et investorid on seda aktsiat juurde ostnud. Vastupidi, kui diff3 väärtus on negatiivne, näitab see, et aktsia osakaal sektoris on vähenenud, mis võib viidata sellele, et investorid on pigem eelistanud aktsiat müüa.

Sorteerides päringu tulemusi diff3 veeru alusel, on võimalik leida, millised aktsiad on investorite seas populaarsemad või ebasoositud. Seega annab loodud päring meile praktilise vahendi investorite eelistuste analüüsimiseks sektorite lõikes ning võimaldab meil tuvastada aktsiad, mida on eelistatud osta või müüa.

3.4 Liidestamine GPT mudeliga

GPT mudeli (*Generative Pre-trained Transformer*) kasutamine nõuab tehnilist teostust, et liidestada GPT mudel kasutajaliidese, andmebaasi ja muude süsteemi komponentidega. Käesolevas peatükis kirjeldatakse, kuidas loodi tehniline teostus, mis võimaldab GPT

modeli kasutamist kasutajatele suunatud rakenduses, mis võimaldab andmebaasist saadud informatsiooni analüüsida ja visualiseerida.

3.4.1 GPT mudeli API

GPT mudeli API on vajalik, et kasutada GPT mudeli funktsionaalsust rakenduses. API tagab juurdepääsu GPT mudelile ja võimaldab selle tulemusi kasutada läbi HTTP päringute. Käesoleva töö raames kasutati GPT mudeli API-ga suhtlemiseks Pythoni teeki nimega "requests", mis võimaldas käivitada POST meetodit, et saada mudelilt vajalikke tulemusi [35]. Seejärel loodi Pythoni koodis "header" muutuja, mis sisaldas vajalikke võtmeid, ning määrati GPT mudeli alammodell, millega sooviti ühendust võtta. Valitud alammodelliks oli "text-davinci-003" [26].

Järgnev koodilõik illustreerib, kuidas kasutati "requests" teeki GPT mudeli API-ga suhtlemiseks:

```
import requests

url = 'https://api.openai.com/v1/completions'
headers = {
    'Content-Type': 'application/json',
    'Authorization': 'Bearer YOUR_API_KEY',
}
data = {
    'prompt': "Your input for GPT",
    'max_tokens': 3000,
    'temperature': 0.5,
    'frequency_penalty': 0.0,
    'presence_penalty': 0.0,
    'model': 'text-davinci-003'
}

response = requests.post(url, headers=headers, json=data)
gpt_answer = response.json()['choices'][0]['text']
```

Joonis 9. Illustreeriv koodilõik GPT mudeliga liidestamisest [26]

Illustreerivast koodilõigust on näha, et päringu saatmisel lisatakse vajalikud päised (headers), milles määratakse sisutüüp ja autentimisvõti, ning json formaadis andmed, milles täpsustatakse päringu parameetrid, näiteks "prompt", "max_tokens", "temperature", "frequency_penalty", "presence_penalty" ja valitud GPT mudeli alammodell.

API-st saadud vastus parsitakse JSON formaadist, et eraldada soovitud tekstivastus, mida seejärel kasutatakse edasi rakenduses.

3.4.2 GPT mudeli API-ga kasutatud parameetrid

Mudelilt soovitud vastuste saamiseks on oluline kasutada ülesandele vastavaid parameetreid, mis omavad genereeritud vastustele märkimisväärset mõju. GPT mudeli API võimaldab parameetrite defineerimist, mida käesoleva peatükiga määratletakse [36].

Järgnev tabel on GPT mudeli API parameetritest ja nende kirjeldustest:

Parameeter	Kirjeldus
prompt	Lähtetekst, mille põhjal GPT mudel genereerib täiendava teksti.
max_tokens	Maksimaalne arv <i>tokeneid</i> , mida GPT mudel peaks genereerima. <i>Token</i> on GPT mudelites kasutatav ühik, mille abil teksti jaotatakse. <i>Tokeni</i> abil on võimalik piirata mudeli poolt kasutatavat mälu ja võimekust.
temperature	Loovuse reguleerimiseks kasutatav parameeter. Mida suurem väärtus, seda loovam on tulemus.
frequency_penalty	Parameeter piiramaks sagedaste <i>tokenite</i> kasutamist, kus suuremad väärtused vähendavad <i>tokeni</i> sagedust terve töödeldava teksti lõikes.
presence_penalty	Parameeter piiramaks korduvate <i>tokenite</i> kasutamist, kus suuremad väärtused vähendavad konkreetsete <i>tokeni</i> kordumise mõju. Korduvuse puhul erinevalt sagedusest keskendutakse rohkem sellele, kui tihti üksik element ilmub teksti alamosades ja kui lähedal need esinemised on üksteisele.

model	Määrab, millist GPT mudeli alammodellit kasutada teksti genereerimiseks.
-------	--

Tabel 4. GPT mudeli API parameetrid [26]

Eelnevas tabelis toodud parameetrid on valitud eelkõige, kuna need on OpenAI poolt soovitatud ja on leitud, et nende abil on võimalik tagada tasakaal loovuse, täpsuse ja sidususe vahel. Lisaks on need parameetrid sobivad magistritöös läbiviidavate eksperimentide jaoks, kuna need loovad eeldused genereerida kvaliteetset ja usaldusväärset teksti.

Kasutatud parameetrid ja nende väärtused on valitud järgmistel põhjustel:

- Prompt - lähtetekstina kasutatakse SQL-i päringu tulemusi ja tabelis olevaid veergusi kirjeldavat lõiku, et suunata mudel genereerima soovitud sisuga teksti.
- Max_tokens - Maksimaalne arv *tokeneid* on seatud 3000-le, et piirata genereeritud teksti pikkust.
- Temperature - Temperatuuri väärtus 0.5 tagab hea tasakaalu loovuse ja sidususe vahel, ilma et tekst muutuks liiga ettearvamatuks või kaootiliseks.
- Frequency_penalty - 0.0 väärtusega ei karistata sagedaste *tokenite* kasutamist, mis võimaldab mudelil paremini keskenduda olulisele sisule.
- Presence_penalty - 0.0 väärtusega ei karistata korduvate *tokenite* kasutamist, mis võimaldab mudelil luua loomulikuma ja sidusama teksti.
- Model - Valitud alammodelliks on text-davinci-003, mis on üks võimsamaid GPT mudeli versioone, pakkudes suurepärasest täpsusest ja loovust genereeritud tekstis. Samuti valik tulenes sellest, et text-davinci-003 ei ole *chat*-tüüpi mudel, mis tagab vastused, kus ei tule formaati ega üleliigseid detaile eemaldada.

3.5 Rakenduse poolt loodud tulemuste võrdlemine LHV artikli lõikudega

3.5.1 „Top omatud aktsiad“ tabelite võrdlemine

LHV artiklis kajastatud „Top omatud aktsiad“ tabel, millega oma rakenduse tulemusi võrdleme:

Aksia	Sümbol	Kokku investoreid	Portfelist suurima osakaaluga	Hinnamuutus (YTD)
Meta Platforms	FB	31	Pat Dorsey	-46,1%
Alphabet	GOOGL	30	AltaRock Partners	-26,8%
Alphabet C	GOOG	27	TCI Fund Management	-26,8%
Microsoft	MSFT	25	AltaRock Partners	-22,5%
Amazon	AMZN	23	Josh Tarasoff	-37,6%
Visa	V	22	TCI Fund Management	-6,2%
Mastercard	MA	18	Valley Forge Capital	-4,9%
Wells Fargo	WFC	18	Charles Munger	-8,9%
Berkshire Hathaway B	BRK.B	17	Bill & Melinda Gates	+3,4%
United Health Group	UNH	16	Sequoia Fund	-0,6%

Allikas: dataroma

Joonis 10. Kuvatõmmis LHV artikli "Top omatud aktsiad" tabelist [33]

Rakenduses kasutatava „Top omatud aktsiad“ päringu tulemused 2022. aasta esimese kvartali kohta olid järgnevad:

stock_name	symbol	investor_count	Largest share of the portfolio	ytd
META PLATFORMS INC	META	26	Dorsey Asset Management, LLC	-42.43%
MICROSOFT CORP	MSFT	26	ALTAROCK PARTNERS LLC	-19.16%
ALPHABET INC	GOOGL	25	ALTAROCK PARTNERS LLC	-21.46%
VISA INC	V	24	TCI FUND MANAGEMENT LTD	-2.09%
ALPHABET INC	GOOG	23	TCI FUND MANAGEMENT LTD	-21.18%
AMAZON COM INC	AMZN	21	GREENLEA LANE CAPITAL MANAGEMENT, LLC	-27.9%
MASTERCARD INCORPORATED	MA	20	Valley Forge Capital Management, LP	-0.4%
BERKSHIRE HATHAWAY INC DEL	BRKB	20	BILL & MELINDA GATES FOUNDATION TRUST	5.68%
JOHNSON & JOHNSON	JNJ	16	Tweedy, Browne Co LLC	4.95%
UNITEDHEALTH GROUP INC	UNH	15	Ruane, Cunniff & Goldfarb L.P.	-1.07%

Tabel 5. "Top omatud aktsiad" päringu tulemused 2022. aasta esimese kvartali kohta

Veeru „investor_count“ tulemused erinesid käesoleva töö „Top omatud aktsiad“ tehnilise teostuse punktis väljatoodud põhjusel ehk kahjuks ei leidu infot dataroma superinvestorite nimekirjast, mida kasutati 2022. esimese kvartali tulemuste analüüsimisel. Sellest tulenevalt me näeme, et tabeli esimese 10 aktsia sekka ei mahtunud ka aktsia „Wells Fargo“.

Veeru „Largest share of the portfolio“ tulemustes erinevusi ei olnud, kuigi investoreid on nimetatud erinevalt. Kõige raskem on märgata seost aktsia United Health Group puhul, kus LHV artiklis on suurimat osakaalu hoidvaks investoriks nimetatud Sequoia Fund ja rakenduse loodud tabelis Ruane, Cunniff & Goldfarb L.P. Sequoia Fund on Ruane, Cunniff & Goldfarb L.P. poolt hallatav investeerimisfond, millest tulenevalt saame öelda, et mõlemal juhul on mõeldud sama investorit.

Veeru „YTD“ tulemused erinevad vähesel määral ning üks põhjustest on info puudumine kuupäeva kohta, mida LHV tabeli YTD arvutuses kasutati. Sellest tulenevalt rakenduses olev lahendus kasutab antud näite puhul aktsia 2022 aasta mai keskmist sulgemishinda.

3.5.2 „Top omatud aktsiad“ lõikude võrdlemine

LHV artiklis kajastatud „Top omatud aktsiad“ lõik, millega oma rakenduse tulemusi võrdleme:

Populaarsete investorite poolt enimomatud aktsiad ühtivad enamjaolt S&P 500 indeksi esikümnega, mistõttu selles osas kvartalite lõikes suuri muutusi ei saa toimuda. Aluseks võetud 78st tuntumatest USA riskifondide juhtidest koguni 31 ehk 48% omavad esimese kvartali seisuga Meta Platformsi aktsiat, mille puhul suurimat osakaalu portfelist omab Pat Dorsey juhtitud Dorsey Asset Management.

Dorsey on Meta Platformsi aktsiaga ostu- või müügitehinguid teinud igas kvartalis alates selle soetamisest 2017. aastal ning viimaste kvartalite põhjal võib öelda, et Dorsey tunnetab Meta aktsia hinnaliikumist üsnagi hästi. 2021. aasta kolmandas kvartalis, kui Meta aktsia oma tipu tegi, võttis ta osa positsioonist kasumina välja, samas kui eelmise aasta neljandas ja selle aasta esimeses kvartalis on needsamad aktsiad odavamalt tagasi ostenud.

Viimases kvartalis, kui Meta aktsia tegi neljanda kvartali tulemuste peale järsu kukkumise, täiendas Dorsey oma Meta positsiooni aga eriti suurelt (aktsiate arvu muutus +31% võrreldes eelneva kvartaliga).

Kuna Alphabetil on börsil noteeritud kahe erineva klassi aktsiad, võib öelda, et suurima investorite arvuga ettevõtte on hoopis Alphabet, olgugi, et mõned investorid omavad korraga mõlema klassi aktsiaid. Reaalselt aga annavad hääleõigusega GOOGL ja ilma hääleõiguseta aktsia GOOG ettevõttesse võrdse osaluse, ehkki viimane võib A-klassi aktsia suhtes börsil kaubelda mõningase allahindlusega.

Joonis 11. Kuvatõmmis LHV artikli "Top omatud aktsiad" lõigust [33]

GPT mudelilt saadud vastus ehk kokkuvõtlik lõik illustreerimaks rakenduse edastatud tabeli „Top omatud aktsiad“ tulemusi:

Kvartali 13F raportite analüüs näitas, et superinvestorite portfellid sisaldasid erinevate US börsil kaubeldavate aktsiate, nagu META PLATFORMS INC, MICROSOFT CORP, ALPHABET INC, VISA INC, AMAZON COM INC, MASTERCARD INCORPORATED, BERKSHIRE HATHAWAY INC DEL, JOHNSON & JOHNSON ja UNITEDHEALTH GROUP INC, aktsiaid. Enim esindatud aktsiaid omandas Dorsey Asset Management, LLC, ALTAROCK PARTNERS LLC, TCI FUND MANAGEMENT LTD, GREENLEA LANE CAPITAL MANAGEMENT, LLC, Valley Forge Capital Management, LP, BILL & MELINDA GATES FOUNDATION TRUST ja Tweedy, Browne Co LLC. Aktsiate YTD hinnamuutused olid vahemikus -42.43% ja 5.68%.

Joonis 12. Kuvatõmmis GPT mudeli loodud "Top omatud aktsiad" lõigust

Võrreldes GPT mudeli loodud lõiku ja LHV artikli lõiku, on mõlemad lõigud keskendunud superinvestorite portfellides sisalduvate aktsiate analüüsile, kuid sisaldavad erinevat tüüpi informatsiooni. GPT mudeli lõik annab kiire ja kokkuvõtliku ülevaate superinvestorite poolt eelistatud aktsiatest ja aasta algusest kuni mai 2022 toimunud

hinnamuutustest, samas kui LHV artikli lõik annab lisaks sellele ka täiendavat konteksti, nagu S&P 500 indeksi esikümne võrdlus ja Pat Dorsey juhtumi analüüs Meta Platforms Inc aktsia ostu- ja müügitehingutes.

LHV artikli lõik sisaldab ka põhjalikumat informatsiooni Alphabedi erinevate aktsiaklasside kohta, mis ei ole GPT mudeli lõigus esitatud. GPT mudeli loodud lõik põhineb ainult sisendandmetel, mis on talle ette antud, samas kui LHV artikli lõik pakub rohkem taustainfot ja analüüsi, mis on finantsuudiste lugejatele kõrgema praktilise väärtusega.

3.5.3 „Portfellist suurima osakaaluga“ tabelite võrdlemine

LHV artiklis kajastatud „Portfellist suurima osakaaluga“ tabel, millega oma rakenduse tulemusi võrdleme:

Väärtpaber	Sümbol	% portfellist	Kokku investoreid	Hinnamuutus (YTD)
Micron Technology	MU	99,8	9	-28,4%
St. Joe	JOE	84,6	1	-7,5%
Icahn Enterprises	IEP	61,0	1	+11,6%
Berkshire Hathaway B	BRK.B	51,2	17	+3,4%
Atlas Corp	ATCO	48,8	2	-9,6%
Berkshire Hathaway A	BRK.A	47,3	12	+2,9%
Hillman Solutions	HLMN	46,9	1	+1,2%
IAC/InterActive	IOC	45,8	2	-39,0%
Bank of America	BAC	44,6	15	-19,5%
Apple	AAPL	42,8	10	-20,7%

Allikas: dataroma

Joonis 13. Kuvatõmmis LHV artikli "Top omatud aktsiad" tabelist [33]

Rakenduses kasutatava „Portfellist suurima osakaaluga“ päringu tulemused 2022. aasta esimese kvartali kohta olid järgnevad:

stock_name	symbol	investor_count	% of portfolio	portfolio_manager	ytd
ST JOE CO	JOE	1	84.64%	FAIRHOLME CAPITAL MANAGEMENT LLC	-3.0%
Icahn Enterprises LP	IEP	1	61.04%	ICAHN CARL C	3.55%
ATLAS CORP	ATCO	2	48.76%	Fairfax Financial Holdings Ltd/ Can	-7.97%
IAC INTERACTIVECORP	IAC	1	45.75%	ShawSpring Partners LLC	-34.74%
EQUITABLE HLDGS INC	EQH	3	32.6%	Conifer Management, L.L.C.	-7.26%
BRC INC	BRCC	1	32.54%	Engaged Capital LLC	-44.77%
HILTON GRAND VACATIONS INC	HGV	1	26.9%	CAS Investment Partners, LLC	-12.2%
Ferguson Plc New	FERG	1	26.09%	TRIAN FUND MANAGEMENT, L.P.	-33.35%
Invesco Ltd.	IVZ	3	23.52%	TRIAN FUND MANAGEMENT, L.P.	-15.99%
EVOLENT HEALTH INC	EVH	1	22.46%	Engaged Capital LLC	1.66%

Tabel 6. "Portfelli suurima osakaaluga" päringu tulemused 2022. aasta esimese kvartali kohta
Tulenevalt superinvestorite nimekirjade erinevusest on „Portfelli suurima osakaaluga“ tabeli puhul erinevusi rohkem märgata kui „Top omatud aktsiad“ tabeli puhul. Tulemused ühtivad aktsiate St Joe Co, Icahn Enterprises LP, Atlas Corp ja IAC Interactive puhul.

3.5.4 „Portfelli suurima osakaaluga“ lõikude võrdlemine

LHV artiklis kajastatud „Portfelli suurima osakaaluga“ lõik, millega oma rakenduse tulemusi võrdleme:

Tuntud investorite suurimaid panuseid vaadates nähtub, et tehnoloogiaaktsiad on üsnagi ebapopulaarne valik, piirdudes vaid kahe *blue-chipi*, Micron Technologiesi ja Apple'iga. Esimesse on märkimisväärse panuse teinud omanimelist riskifondi haldav Mohnish Pabrai, kes usub pooljuhtide valdkonna suundumustesse ning on mitmes intervjuus oma panust Micronisse seletanud sellega, et Micron on üks vähestest pooljuhtide ettevõtetest, kellel on dünaamilise muutmälu (DRAM) tootmise võimekus (eriti nõutud komponent tehisintellektis).

Kuna aga 13F raport, mille alusel andmed on kogutud, nõuab ainult USA väärtpaberitega seotud tehingute raporteerimist, ei ole Pabrai investeringutest täit pilti võimalik saada, sest mõned neist ei ole USA börsidel noteeritud. Kuigi viimase aruande põhjal moodustab Micron Technology Pabrai portfelli justkui 99,8%, on tegelik osakaal tõenäoliselt väiksem. Samas on Pabrai tõdenud, et ei karda hoida suurt osa varadest vaid ühes aktsias, kui on veendunud selle kasumipotentsiaalis. Microni aktsiat ostis Pabrai viimati veidi juurde 2021. aasta neljandas kvartalis, kuid viimases kvartalis püsis positsioon muutumatuna.

Joonis 14. Kuvatõmmis LHV artikli "Portfelli suurima osakaaluga" lõigust [33]

GPT mudelilt saadud vastus ehk kokkuvõtlik lõik illustreerimaks rakenduse edastatud tabeli „Portfellist suurima osakaaluga“ tulemusi:

Tabeli andmetel omasid 72 superinvestori portfellis valitud aktsiaid, milleks olid ST JOE CO (JOE), Icahn Enterprises LP (IEP), ATLAS CORP (ATCO), IAC INTERACTIVECORP (IAC), EQUITABLE HLDGS INC (EQH), BRC INC (BRCC), HILTON GRAND VACATIONS INC (HGV), Ferguson Plc New (FERG) ja Invesco Ltd. (IVZ). Suurima osakaalu oma portfellis omas investor FAIRHOLME CAPITAL MANAGEMENT LLC 84,64 protsendiga ST JOE CO aktsiast ning valitud aktsia YTD hinnamuutus alates aasta algusest kuni "2022-03-31" ulatus -3.0% lõpuni.

Joonis 15. Kuvatõmmis GPT mudeli loodud "Portfellist suurima osakaaluga" lõigust

GPT mudeli lõik sisaldab taaskord ainult sisendandmete tõlgendamist ehk esitatakse tabelis olevaid andmeid kokkuvõtlikul viisil. LHV artikli lõigus seevastu lisatakse konteksti ning lisaks aktsiate sektorite mainimisele analüüsitakse põhjalikumalt investori Mohnish Pabrai portfelli ja selles toimunud muutusi.

3.5.5 „Top ostud“ tabelite võrdlemine

LHV artiklis kajastatud „Top ostud“ tabel, millega oma rakenduse tulemusi võrdleme:

Top ostud 2022. aasta esimeses kvartalis

Väärtpaber	Sümbol	Ostjaid	Kokku investoreid	Hinnamuutus (YTD)
Meta Platforms	FB	18	31	-46,1%
Microsoft	MSFT	12	25	-22,5%
Alphabet	GOOGL	10	30	-26,8%
Amazon	AMZN	8	23	-37,6%
Visa	V	7	22	-6,2%
Moody's	MCO	7	12	-24,7%
Paypal Holdings	PYPL	6	7	-58,3%
Citigroup	C	6	11	-11,2%
S&P Global	SPGI	6	10	-25,5%
Starbucks	SBUX	6	8	-37,1%

Allikas: dataroma

Joonis 16. Kuvatõmmis LHV artikli "Top ostud" tabelist [33]

Rakenduses kasutatava „Top ostud“ päringu tulemused 2022. aasta esimese kvartali kohta olid järgnevad:

Aktzia_Nimi	symbol	Ostjaid	Kokku_investoreid	YTD
VISA INC	V	14	25	-2.09%
MASTERCARD INCORPORATED	MA	13	17	-0.4%
BERKSHIRE HATHAWAY INC	BRKB	8	19	5.68%
META PLATFORMS INC	META	6	25	-42.43%
BRISTOL-MYERS SQUIBB CO	BMY	6	8	21.01%
APPLE INC	AAPL	6	13	-16.18%
ANTHEM INC	ELV	6	8	9.94%
AMERICAN EXPRESS CO	AXP	6	12	3.19%
WELLS FARGO CO NEW	WFC	5	13	-4.61%
SCHWAB CHARLES CORP	SCHW	5	9	-16.65%

Tabel 7. "Top ostud" päringu tulemused 2022. aasta esimese kvartali kohta

Päringu „Top ostud“ tulemused sõltuvad määratud superinvestorite nimekirjast, millest tulenevad ka erinevused võrreldes LHV artiklis kajastatud tabeliga. Päringu tulemused on sorteeritud veeru „Ostjaid“ järgi kahanevalt. Ühtivate aktsiate nagu Meta Platforms Inc ja Visa Inc veeru „YTD“ tulemusi analüüsid näeme, et hinnamuutust on kajastatud üheselt LHV artiklis oleva tabeliga ehk milliste aktsiate hinnad tõusid ja milliste aktsiate hinnad langesid.

3.5.6 „Top ostud“ lõikude võrdlemine

LHV artiklis kajastatud „Top ostud“ lõik, millega oma rakenduse tulemusi võrdleme:

Kõige rohkem osteti esimeses kvartalis juurde Meta Platformsi aktsiat, kuid ostupoolel oldi ka teiste tuntud nimede suhtes, eriti mis puudutab finantsettevõtteid. Meta Platforms on ka viimase kahe kvartali arvestuses enimostetum aktsia, sealjuures viimase kvartali 18st ostust kaks hõlmasid esmakordset positsiooni avamist (Michael Burry ja Stephen Mandel).

Kui tabeli esimene ots ühtib taaskord suuresti S&P 500 järjekorraga, siis kuuendale kohale sattus S&P 500 nimekirjas 159. kohal olev reitinguagentuur Moody's, mis on ühtlasi ka tabeli nimekirjas turuväärtuselt väikseim. Tabelisse mahtus ka teine turuanalüüse koostav ettevõtte S&P Global, mille ostjad ühtisid esimeses kvartalis suures osas Moody'se ostjatega (4 investorit ostsid samal ajal nii Moody'se kui ka S&P Globali aktsiaid).

Joonis 17. Kuvatõmmis LHV artikli "Top ostud" lõigust [33]

GPT mudelilt saadud vastus ehk kokkuvõtlik lõik illustreerimaks rakenduse edastatud tabeli „Top ostud“ tulemusi:

Kvartal "2022-03-31" finantsmaailmas oli põnev, sest mitmed suured investorid olid liikunud aktiivselt turule ja investeerinud erinevatesse aktsiatesse. Järgnevalt on toodud tabel, mis näitab 10 enim ostetud aktsiat investorite järgi kahanevalt sorteerituna. Tabeli veerust Ostjaid näeme, et VISA INC (V) oli enim ostetud aktsia 14 investoriga, samas kui MASTERCARD INCORPORATED (MA) oli teisel kohal 13 investoriga. Kokku investoreid veergust näeme, et 25 investoril oli VISA INC portfellis ja 17 investoril MASTERCARD INCORPORATED. YTD veergust näeme, et VISA INC ja MASTERCARD INCORPORATED vähenesid vastavalt -2.09% ja -0.4%. Samuti on tabelis BERKSHIRE HATHAWAY INC (BRKB), mis tõusis 5.68%, META PLATFORMS INC (META), mis langes -42.43%, BRISTOL-MYERS SQUIBB CO (BMY), mis tõusis 21.01%, APPLE INC (AAPL), mis langes -16.18%, ANTHEM INC (ELV), mis tõusis 9.94%, AMERICAN EXPRESS CO (AXP), mis tõusis 3.19%, WELLS FARGO CO NEW (WFC), mille hinnamuutus oli -4.61% ja SCHWAB CHARLES CORP (SCHW), mille hinnamuutus oli -16.65%.

Joonis 18. Kuvatõmmis GPT mudeli loodud "Top ostud" lõigust

Tulenevalt tabelite erinevusest, ei saa valeks pidada, et Meta Platformsi aktsia näitel oleks lõik valesti koostatud. Peamiselt erinevad lõigud selle poolest, et LHV artikkel sisaldab rohkem informatsiooni, mida tabel edasi ei anna. Näiteks tabeli tulemusi on LHV artiklis võrreldud S&P500 järjekorraga. GPT mudeli loodud lõigu puhul on näha, et peamine fookus on olnud tabeli andmete peegeldamine ning peale investorite aktiivse tegevuse ei ole finantsmaailma kohta rohkem midagi lisatud.

3.6 Rakenduse poolt loodud lõikude ja sisendandmete võrdlemine

Käesoleva peatüki raames tehti kvalitatiivne analüüs, et võrrelda 50 loodud lõiku ja sisendandmete vastavust erinevate perioodide ja päringute kohta. Analüüs hõlmas mudeli poolt loodud lõikude ja algsete sisendandmete süvitsi võrdlemist, et hinnata mudeli võimet genereerida asjakohaseid ja täpseid lõike, mis peegeldavad üheselt mudelile edastatud sisendandmeid.

Lisaks sisendandmete vastavusele kontrolliti, kas GPT mudel on lisanud lõikudesse tausta, konteksti või loonud järeldusi, mida sisendandmetega ei esitatud.

Kvalitatiivse analüüsi tulemustest ilmnes, et 50-st analüüsitud lõigust 34 sisaldasid finantsmaailma olukorda kirjeldavat sisu. Väited finantsmaailma kohta analüüsitud perioodidel olid pigem üldised, kuid esines ka näiteid, kus mainiti Covid-19 pandeemiat või majandusfaasi.

Lisaks esines 26 lõigu puhul järeldusi, nagu investorite meelestatus aktsiaturu suhtes, kuid järeldused olid pigem üldised ega olnud seostatud mõne spetsiifilise fakti või näitega, mida sisendandmetena ei edastatud. GPT mudelile edastatud sisendandmeid kajastati lõikudes peamiselt korrektselt - 43 juhul 50-st ei esinenud vigu tabelites olevate andmete kajastamisel.

Siiski esines raskusi keerukamate päringutulemuste selgitamisel, nagu "Portfellist suurima osakaaluga" või "Top omatud aktsiad". Nende puhul oli GPT mudelil kõige enam probleeme tabeli sisu mõistmisega, hoolimata sellest, et esitatud *prompt* sisaldas detailset kirjeldust veergudest ja tabeli tulemustest. 16 analüüsitud keerukama päringu näitel esines 6 juhul vigu, kus veerge oli tõlgendatud valesti.

Näide, kus sisendina esitatud komakohaga 'YTD' väärtust on kasutatud lõigus protsendina, kus koma on puudu:

Tegemist oli "2022-03-31" kvartali 13F raportite analüüsiga, mille põhjal selgus, et 72 superinvestorit omavad US börsil kaubeldavaid aktsiaid. Kõige suurema osakaalu oma portfellis omab investor Fairholme Capital Management LLC 84.64% ulatuses ST JOE CO aktsiast, mille YTD hinnamuutus on -3%. Icahn Carl C omab 61.04% ulatuses IEP aktsiat, mille YTD hinnamuutus on 3.55%. Fairfax Financial Holdings Ltd/ Can omab 48.76% ulatuses ATCO aktsiat, mille YTD hinnamuutus on -7.97%. Engaged Capital LLC omab 32.54% BRCC aktsiat ja 22.46% EVH aktsiat ning nende vastavad YTD hinnamuutused olid -44.77%, 1.66%. Conifer Management, L.L.C omab 32.6% EQH aktsiat ja selle YTD hinnamuutus oli -7.26%. TRIAN FUND MANAGEMENT, L.P omab 26.09% FERG ja 23.52% IVZ aktsiat ning nende vastavad YTD hinnamuutused olid -33.35%, -15.99%.

stock_name	symbol	investor_count	% of portfolio	portfolio_manager	ytd
ST JOE CO	JOE	1	84.64%	FAIRHOLME CAPITAL MANAGEMENT LLC	-3.0%
Icahn Enterprises LP	IEP	1	61.04%	ICAHN CARL C	3.55%
ATLAS CORP	ATCO	2	48.76%	Fairfax Financial Holdings Ltd/ Can	-7.97%
IAC INTERACTIVECORP	IAC	1	45.75%	ShawSpring Partners LLC	-34.74%
EQUITABLE HLDGS INC	EQH	3	32.6%	Conifer Management, L.L.C.	-7.26%
BRC INC	BRCC	1	32.54%	Engaged Capital LLC	-44.77%
HILTON GRAND VACATIONS INC	HGV	1	26.9%	CAS Investment Partners, LLC	-12.2%
Ferguson Plc New	FERG	1	26.09%	TRIAN FUND MANAGEMENT, L.P.	-33.35%
Invesco Ltd.	IVZ	3	23.52%	TRIAN FUND MANAGEMENT, L.P.	-15.99%
EVOLANT HEALTH INC	EVH	1	22.46%	Engaged Capital LLC	1.66%

Joonis 19. Kuvatõmmis sisendandme mittekorrektsest kasutamisest lõigus

Analüüsitud 50 lõigu ja esitatud sisendandmete põhjal saame järeldada, et suures osas GPT mudel kajastab loodud lõikudes esitatud sisendandmeid üheselt ning ei lisa omalt poolt tausta, konteksti ega järeldusi.

3.7 Rakenduse poolt loodud lõikude hindamine GPT mudeli poolt

Käesoleva töö raames loodi iga loodud päringu kohta 30 kokkuvõtlikku lõiku, mis hõlmasid erinevaid perioode. Kokku saadeti GPT mudelile hinnangu andmiseks ja hinnangu põhjendamiseks 180 lõiku koos lõiku kirjeldava tabeliga ehk vastava päringu tulemustega. GPT mudeli ülesandeks oli hinnata 10-palli süsteemis loodud lõiku selle vastavuses tabeliga ja sobivust uudisartiklis lõiguna kasutamiseks. Lisaks numbrilisele hinnangule küsiti GPT mudelilt selgitust põhjendamaks numbrilist väärtust.

Tulemuste analüüsimisel selgus, et GPT mudel on hinnanud kõrgemalt lõike, mis kajastavad võimalikult palju tabelis olevaid andmeid. Kriitiliselt oli hinnatud lõiku, kui selles puudusid andmed, mida tabelis ei kajastatud. GPT mudeli hinnangutest selgus, et näiteid oli ühest äärmusest teise ehk teatud olukordades hinnati lõike liiga kriitiliselt ning teatud olukordades vastupidiselt.

Järgnevalt on tabeli kujul esitatud kvantitatiivse analüüsi tulemused:

Päringu nimi	Arv	Keskmine hinne	Minimaalne hinne	Maksimaalne hinne
Aktsiate osakaalud (asc)	15	9.4	5	10
Aktsiate osakaalud (desc)	15	9.87	9	10
Investorite portfellid	30	8.43	5	10
Portfelligist suurima osakaaluga	30	7.9	5	10
Sektorite osakaalud	30	9.37	5	10
Top omatud aktsiad	30	9.33	5	10
Top ostud	30	8.8	5	10
Kõik päringud kokku	180	8.91	5	10

Tabel 8. GPT mudeli hinnangute tulemused

Keelemudeli hinnangud näitavad, et loodud päringute tulemuste põhjal genereeritud lõigud olid üldiselt kõrgelt hinnatud. Madalaim hinne oli 5 ja kõrgeim 10, mis viitab sellele, et GPT mudeli arvates oli enamikul juhtudel loodud sisukad ja asjakohased lõigud. Keskmine hinne kõikidele päringutele oli 8.91, mis näitab, et GPT mudel hindas oma loodud lõike üldiselt positiivselt.

GPT mudelilt hinnangute küsimine on hea selle poolest, et saab häid ideid ja mõtteid, mida uudisartikli lõik võiks sisaldada. Samas tuleb olla ettevaatlik selle hindamise meetodi

kasutamisel, kuna GPT mudel ei pruugi alati anda kõige usaldusväärsemaid tulemusi. Kuigi GPT mudel suudab tihti pakkuda asjakohaseid ja sisukaid lõike, on vaja hinnangute valideerimise osas veel tööd teha, enne kui neid täielikult GPT mudelile usaldada saab.

3.8 Automatiseeritud uudisartiklite loomine GPT mudeliga

Automatiseeritud uudisartikli tehniline teostus hõlmab:

1. Andmeanalüüsi käigus loodud päringuid
2. Pythoni skripti, mis:
 - a. Käivitab loodud päringud andmebaasil
 - b. Edastab koos kirjeldava teksti ja ülesandega päringute tulemused GPT mudelile kasutades OpenAI loodud API-t
 - c. Salvestab GPT mudelilt saadud vastused
 - d. Loob uudisartikli ehk HTML faili, kus on kasutatud päringute tulemustena saadud tabeleid ja lisatud nende kohta GPT mudeli loodud kokkuvõtlikud lõigud

Automatiseeritud tähendab, et enne valitud perioodi uudisartikli koostamist on loodud päringud tabelite jaoks, mida artiklis kasutatakse. Seejärel käivitatakse kasutatavad päringud ning saadakse andmebaasist päringute tulemused, mis kajastavad 13F raportitest loodud kokkuvõtlike finantstulemusi. Loodud tulemuste selgitamiseks esitatakse päringute tulemused läbi OpenAI API GPT mudelile, et saada tulemusi selgitavad lõigud. Vastusena saadud lõigud koos tabelitega kombineeritakse ühtsesse HTML faili, mis moodustabki uudisartikli.

Kvartali 13F raportite analüüsitulemustest selgus, et 72st investorist valitud 8 aktsiat omasid 26-24 investorit. Suurema osakaaluga investorite seas domineerisid ALTAROCK PARTNERS LLC ja TCI FUND MANAGEMENT LTD. Aktsia META PLATFORMS INC sümboliga META YTD aasta algusest kuni mai 2022 oli -42,43%, Microsoft Corpi MSFT aktsia YTD -19,16%, Alphabet Inc GOOGL aktsia YTD -21,46%, Visa Inc V aktsia YTD -2,09%, Alphabet Inc GOOG aktsia YTD -21,18%, Amazon Com Inc AMZN aktsia YTD -27,9%, Mastercard Incorporated MA aktsia YTD -0,4%, Berkshire Hathaway Inc Del BRKB aktsia YTD 5,68%, Johnson & Johnson JNJ aktsia YTD 4,95% ja Unitedhealth Group Inc UNH aktsia YTD -1,07%. Finantsmaailmas on viimaseks aastaks olnud suur volatiilsus ning see on selgitatav tabelis toodud väiksemate ja suuremate tulemustega.

stock_name	symbol	investor_count	Largest share of the portfolio	ytd
META PLATFORMS INC	META	26	Dorsey Asset Management, LLC	-42.43%
MICROSOFT CORP	MSFT	26	ALTAROCK PARTNERS LLC	-19.16%
ALPHABET INC	GOOGL	25	ALTAROCK PARTNERS LLC	-21.46%
VISA INC	V	24	TCI FUND MANAGEMENT LTD	-2.09%
ALPHABET INC	GOOG	23	TCI FUND MANAGEMENT LTD	-21.18%
AMAZON COM INC	AMZN	21	GREENLEA LANE CAPITAL MANAGEMENT, LLC	-27.9%
MASTERCARD INCORPORATED	MA	20	Valley Forge Capital Management, LP	-0.4%
BERKSHIRE HATHAWAY INC DEL	BRKB	20	BILL & MELINDA GATES FOUNDATION TRUST	5.68%
JOHNSON & JOHNSON	JNJ	16	Tweedy, Browne Co LLC	4.95%
UNITEDHEALTH GROUP INC	UNH	15	Ruane, Cunniff & Goldfarb L.P.	-1.07%

Finantsmaailmas on 2022-03-31 kvartalit iseloomustanud suur volatiilsus, mis on põhjustatud järskudest hinnamuutustest. Tabelis olevad andmed näitavad, et investorid on oma portfelligesse võtnud erinevaid aktsiaid, mille YTD hinnamuutused on erinevad. Näiteks Fairholme Capital Management LLC omab 84.64% oma portfelist ST JOE CO aktsiatest, mille YTD hinnamuutus oli -3.0%. Samuti omab Icahn Carl C 61.04% IEP aktsiatest ja selle YTD hinnamuutus oli 3.55%. Fairfax Financial Holdings Ltd/ Can omab 48.76% ATCO aktsiatest ja selle YTD hinnamuutus oli -7.97%.

stock_name	symbol	investor_count	% of portfolio	portfolio_manager	ytd
ST JOE CO	JOE	1	84.64%	FAIRHOLME CAPITAL MANAGEMENT LLC	-3.0%
Icahn Enterprises LP	IEP	1	61.04%	ICAHN CARL C	3.55%
ATLAS CORP	ATCO	2	48.76%	Fairfax Financial Holdings Ltd/ Can	-7.97%
IAC INTERACTIVECORP	IAC	1	45.75%	ShawSpring Partners LLC	-34.74%
EQUITABLE HLDGS INC	EQH	3	32.6%	Conifer Management, L.L.C.	-7.26%
BRC INC	BRCC	1	32.54%	Engaged Capital LLC	-44.77%
HILTON GRAND VACATIONS INC	HGV	1	26.9%	CAS Investment Partners, LLC	-12.2%
Ferguson Plc New	FERG	1	26.09%	TRIAN FUND MANAGEMENT L P	-33.35%

Joonis 20. Kuvatõmmis Pythoni scripti loodud HTML tüüpi uudisartiklist

4 Analüüs ja järeldused

Magistritöö tulemused põhinevad kahe peamise uurimismeetodi kasutamisel: *desk research* ja *prompt engineering*.

Desk researchi käigus koguti ja analüüsiti olemasolevat teavet 13F raportite ja finantsuudiste artiklite kohta, et luua andmebaas vajalike andmetega. Selles etapis keskenduti andmete kogumisele, eeltötlusele, teadmiste omandamisele ja analüüsile. See hõlmas olulise teabe eraldamist, finantsartiklite struktuuri uurimist ning 13F raportite andmete analüüsi.

Prompt engineeringu abil loodi GPT mudeli jaoks sisendandmete eeltöötlus, mis aitas mudelil paremini mõista sisendandmeid ning anda asjakohaseid ja kvaliteetseid väljundeid. Seejärel esitati promptid GPT mudelile, mis genereeris uudisartiklid. Protsessi käigus tehti mitmeid iteratsioone, et täiustada prompte ja parandada mudeli tulemusi. Lõpuks hinnati genereeritud uudisartiklite kvaliteeti - võrreldi neid finantsanalüütikute poolt loodud artiklitega, teostati kvalitatiivset analüüsi sisendandmetele ja loodud lõikudele ning kasutati tehisintellekti loodud lõikude hindamiseks.

Kokkuvõtvalt saavutati magistritöö tulemused, kasutades *desk researchi* ja *prompt engineeringu* meetodeid. Andmebaas loodi vajalike andmetega, töödeldi päringute tulemused, liidestati GPT mudeliga ning formuleeriti saadud kokkuvõtlikud lõigud uudisartiklikliks. Töö tulemusena loodi süsteem, mis automatiseerib 13F raportitest põhinevate finantsuudiste artiklite loomise protsessi, kasutades OpenAI GPT mudelit ja andmeanalüüsi meetodeid.

4.1 Tehnilise teostuse põhjendus

Tehnilise teostuse valik käesoleva töö jaoks tugines mitmetel kaalutlustel. Esiteks, 13F raportitest andmete eeltöötlus ja andmeanalüüs olid töö ühed mahukamad osad. Nende osade läbiviimiseks oli vaja kulutada märkimisväärne hulk aega ja ressursse, et saavutada kvaliteetseid tulemusi.

Kuigi alternatiiv ehk mudeli peenhäälestamine oleks võinud parandada GPT mudeli sooritust, oleks see samuti endast hõlmanud andmeanalüüsi ja andmebaasi loomist.

Mudeli peenhäälestamine oleks suurendanud projekti keerukust ja mahtu, mistõttu oleks olnud keerulisem jõuda praktiliste näidete ja uudisartiklite loomiseni.

Andmete eeltöötlus ja andmeanalüüs, mis olid tehtud töö käigus, võimaldasid töö autoril omandada väärtuslikke teadmisi 13F raportite sisust, struktuurist ja finantsuudiste loomise protsessist. See omakorda aitas kaasa parema ülevaate saamisele, kuidas kasutada OpenAI GPT mudelit uudisartiklite loomiseks.

Prompt engineeringu kasutamine võimaldas autoril keskenduda GPT mudeli sisendandmete optimeerimisele ja saavutada rahuldavad tulemused, ilma et oleks tehtud mudeli peenhäälestamist. See lähenemine osutus käesoleva magistritöö puhul valituks, kuna see võimaldas jõuda praktiliste tulemusteni kiiremini ja vähemate ressurssidega.

Arvestades töö mahukust, ressursse ja ajapiiranguid, otsustati tehniline teostus lahendada, keskendudes andmete eeltötlusele, andmeanalüüsile ning *prompt engineeringule*, loobudes mudeli peenhäälestamisest.

4.2 Järeldused GPT mudeli kasutamisest

Käesoleva peatüki raames tuuakse välja mitmed olulised aspektid ja järeldused, mis aitavad kaasa GPT mudeli paremale kasutamisele ja tulemuste saavutamisele.

4.2.1 Sisendandmete põhjaliku kirjeldamise olulisus GPT mudeli kasutamisel

Sisendandmete põhjalik kirjeldamine on GPT mudeli kasutamisel oluline, et saavutada asjakohaseid ja kvaliteetseid väljundeid. Mudel põhineb suurel hulgal erinevatel tekstilistel andmetel, mis tähendab, et ilma selgete ja põhjalike juhisteta võib mudel tõlgendada sisendandmeid ebatäpselt või anda ebarelevantseid vastuseid.

Järgmised punktid toovad välja, miks sisendandmete põhjalik kirjeldamine on GPT mudeli kasutamisel oluline:

- Konteksti mõistmine - põhjaliku sisendandmete kirjelduse abil suudab mudel paremini mõista nõutavat konteksti, mis aitab kaasa täpsemate ja asjakohaste vastuste genereerimisele.

- Täpsemate tulemuste saamine - üksikasjalike kirjelduste kasutamine tagab, et GPT mudelil on piisavalt teavet sisendandmete kohta, mis omakorda võimaldab mudelil teha täpsemaid järeldusi ja genereerida täpsemaid tulemusi.
- Vähem eksitavaid vastuseid ja suurem usaldusväarsus - sisendandmete põhjaliku kirjeldamise abil saab vähendada võimalust, et mudel annab eksitavaid või mittetäielikke vastuseid ning suureneb usaldusväarsus, et mudel genereerib täpseid ja asjakohaseid vastuseid, mis vastaksid kasutaja ootustele. Põhjalik kirjeldamine aitab kaasa lõpliku väljundi kvaliteedi tagamisele.

Sisendandmete põhjaliku kirjeldamise olulisus GPT mudeli kasutamisel märkimisväärne, kuna see aitab mudelil paremini mõista kasutaja ootusi ja pakkuda täpsemaid, asjakohaseid ja kvaliteetseid tulemusi.

4.2.2 GPT API muutuja *max_tokens* mõju sisendandmete analüüsimisele

GPT API muutuja *max_tokens* määrab, kui palju tokeneid mudel saab maksimaalselt genereerida vastusena sisendandmetele. Samuti muutuja mängib olulist rolli sisendandmete analüüsimisel. Järgnevalt on toodud mõned olulised aspektid, mis rõhutavad *max_tokens* muutuja mõju sisendandmete analüüsimisele:

- Ebapiisav analüüs mudeli poolt - kui *max_tokens* muutuja väärtus on liiga väike võrreldes sisendandmete suurusega, võib mudel jätta mõned olulised osad sisendandmetest analüüsimata. See võib omakorda viia ebatäpsete või mittetäielike vastuste genereerimiseni.
- Väljundi pikkuse kontrollimine - *max_tokens* muutuja abil saab kontrollida GPT mudelilt saadud väljundi pikkust. See võimaldab tagada, et väljund ei ole liiga pikk ega liiga lühike, vaid vastab kasutaja ootustele ja vajadustele.
- Ressurssikasutuse optimeerimine - *max_tokens* muutuja abil saab piirata mudeli ressursikasutust, vähendades genereeritavate tokenite arvu. See aitab vältida ressursimahukate analüüside tegemist, mis võivad olla tarbetud või ebaefektiivsed.

GPT API muutuja *max_tokens* on oluline sisendandmete analüüsimisel, kuna see mõjutab mudeli võimekust analüüsida sisendandmeid ja genereerida asjakohaseid tulemusi.

Max_tokens väärtuse hoolikas valimine ja optimeerimine tagab, et GPT mudel suudab pakkuda täpseid ja kvaliteetseid tulemusi, mis vastavad kasutaja ootustele ja vajadustele.

4.2.3 GPT mudeli usaldusväarsus sisendandmete peegeldamisel

GPT mudeli usaldusväarsuse hindamiseks sisendandmete peegeldamisel viidi läbi kvalitatiivne analüüs, mille käigus analüüsiti 50 GPT mudeli loodud lõiku. Autor analüüsis GPT mudeli poolt loodud lõikude sisendandmete kasutust ja vastavust, et hinnata, kui hästi mudel suudab käsitleda ja kajastada esitatud andmeid.

Kvalitatiivse analüüsi käigus leiti 50 analüüsitud lõigust 7, kus sisendandmetes esitatut ei kajastatud loodud lõikudes üheselt. Näiteks esines juhte, kus päringutulemuses miinuskärgiga olevaid protsente kasutati lõigus positiivse väärtusena, mis tähendab, et loodud lõikudes jäid miinuskärgid lisamata. Lisaks esines üksikutel juhtudel tabelites olevate veergude mitte mõistmist, hoolimata sellest, et GPT mudelile oli esitatud sisendandmete kohta täpne kirjeldus, mida andmed esindavad. See näitab, et mudel võib mõnikord olla piiratud oma võimekuses mõista ja kajastada andmeid, isegi kui on sisendandmetega esitatud selged juhised.

Kvalitatiivse analüüsi tulemused näitavad, et kuigi GPT mudel on üldiselt usaldusväärne sisendandmete peegeldamisel, võib esineda üksikuid vigu ja piiranguid. Hoolikas sisendandmete ettevalmistamine ja *promptide* formuleerimine aitavad parandada mudeli täpsust, kuid kasutajad peaksid olema teadlikud võimalikest vigadest ja mudeli piirangutest.

4.2.4 GPT mudeli kasutamine loodud lõikude hindamiseks

Käesoleva töö raames uuriti, kuidas GPT mudel hindab enda loodud lõike. Selleks esitati GPT mudelile nii loodud lõik kui ka sisendandmed ning küsiti hinnangut. GPT mudelilt küsiti hinnanguid kõikide loodud päringute kohta ning igale päringule loodud 30 kokkuvõtliku lõigu kohta.

Kvantitatiivse analüüsi tulemused näitavad, et GPT mudel hindab enda loodud lõike keskmiselt hindegaga 8.91 10-palli süsteemis. Põhjendused numbrilistele hinnangutele jäävad pigem üldiseks, kuid teatud juhtudel on toodud ka üksikuid detaile, mis hinnet mõjutasid.

Loovuslik lähenemine tulemuste valideerimisele annab ülevaate GPT mudeli usaldusväärsusest loodud lõikude kvaliteedi osas. Samuti aitab see mõista, kuidas mudel suhestub oma loodud tekstidega ning milliseid aspekte peetakse oluliseks tekstide hindamisel.

4.2.5 Välised tegurid, mis mõjutavad andmete võrreldavust erinevates allikates

Andmete võrreldavust erinevates allikates mõjutavad mitmed välised tegurid, mis võivad põhjustada erinevusi analüüsitulemustes ja järeldustes. Mõned neist teguritest on:

- Andmeallikate erinevused - erinevad allikad võivad kasutada erinevaid andmekogumise, töötlemise ja esitamise meetodeid, mis võivad põhjustada analüüsi tulemustes erinevusi. Näiteks võivad erinevates allikates esitatud aktsiahinnad põhineda erinevatel ajahetkedel, erinevate valuutakursside kasutamisel või kui puudub info, kuidas aktsiahind on arvutatud. Sellised erinevused võivad tuleneda asjaolust, et kõikide detailide ja kasutatud meetodikate kohta puudub täielik teave, mis omakorda võivad mõjutada analüüsitulemusi.
- Andmete töötlemine ja eeltöötlus - andmete töötlemise ja eeltötluse käigus tehtud valikud võivad mõjutada andmete võrreldavust erinevates allikates. Näiteks erinevate andmebaaside loomisel võidakse kasutada erinevaid kriteeriume ja arvutusi, mis võivad põhjustada erinevusi analüüsitulemustes.
- Valimi erinevused - andmete võrreldavust võivad mõjutada ka valimi erinevused, kui uurimisobjektide valik erineb erinevates allikates. Näiteks võib magistritöö ja uudisartikli vaadeldavate investorite nimekiri erineda, mis võib mõjutada andmete võrreldavust.
- Mõõtmisvead ja ebatäpsused - kõikides andmeallikates võib esineda mõningaid mõõtmisvigu ja ebatäpsusi, mis võivad mõjutada andmete võrreldavust erinevates allikates.

On keeruline luua andmeanalüüsi ja päringuid, mille tulemused täielikult kattuksid finantsartiklites kajastatud tulemustega, kuna mitmed välised ja sisemised tegurid mõjutavad andmete võrreldavust erinevates allikates. Need tegurid võivad põhjustada erinevusi nii andmete kogumise, töötlemise, esitamise kui ka tõlgendamise osas.

Sellela arvestades on oluline olla tähelepanelik võimalike erinevuste ja vigade suhtes, et teha usaldusväärseid järeldusi. Andmete võrreldavuse hindamisel erinevates allikates tuleks keskenduda kvaliteedi tagamisele ning kriitilisele lähenemisele, et saavutada kõige täpsem ja usaldusväärsem tulemus.

4.3 Alternatiivid teostatud tööle

Käesoleva töö raames kasutati GPT mudelit andmeanalüüsi tulemuste põhjal uudisartiklite loomiseks. Siiski on olemas alternatiive, mida võinuks kaaluda, et saavutada sarnaseid tulemusi või veelgi paremat kvaliteeti loodud tekstides. Üks neist alternatiividest on GPT mudeli peenhäälestamine (*fine-tuning*), mida OpenAI arendajatele võimaldab [37].

13F raportite näitel võiks GPT mudeli peenhäälestamine tähendada, et mudelit treenitakse spetsiifiliste andmetega, mis on seotud finantsvaldkonna ja 13F raportitega. See võimaldaks mudelil paremini mõista finantstermineid, konteksti ja andmestruktuure, mis on seotud 13F raportitega. Peenhäälestamine hõlmaks andmete eeltötlust, andmeanalüüsi ning andmebaasi loomist, mis sisaldab erinevaid näiteid 13F raportitest ja nendega seotud uudisartiklitest. On tõestatud, et peenhäälestamine on erinevate ülesannete täitmiseks efektiivne lahendus, tagades kõrge kvaliteediga tulemused [38].

Selle protsessi käigus oleks võimalik ka integreerida spetsiifilisi reegleid ja eelistusi, et mudel suudaks paremini luua sisukaid ja täpseid uudisartikleid 13F raportite põhjal.

4.3.1 Olemasolevad lahendused

Olemasolevatest lahendustest, mis kasutab peenhäälestamise meetodit, on FinChat. Finchat on veebipõhine platvorm, mis pakub finantsteksti analüüsi pakkuvat *chat*-tüüpi lahendust, kasutades peenhäälestatud GPT mudelit. See tähendab, et nende poolt loodud tekstid on rohkem kohandatud finantsvaldkonna jaoks, pakkudes potentsiaalselt paremat kvaliteeti ja täpsust uudisartiklite loomisel [39].

Andmeanalüüsi valdkonnas on erinevaid lahendusi, mis pakuvad andmeanalüüsi 13F raportite näitel. Näiteks Whale Wisdom ja FormThirteen on mõlemad silmapaistvad näited rakendustest, mis teostavad keerukaid andmeanalüüse, pakkudes samas kasutajasõbralikku ja intuitiivset kasutajaliidest. Mõlema veebipõhise platvormi puhul on

tegemist lahendustega, mis pakuvad suurepärase eeskuju, kuidas finantsandmete analüüsi teostada. Ka käesoleva töö puhul võeti mainitud platvormi andmeanalüüsi osas eeskujuks [40] [41].

4.4 Edasine töö

Käesoleva magistr töö käigus loodud lahendus on hea tehniline *proof of concept*, mis näitab GPT mudeli võimekust luua uudisartikleid 13F raportite põhjal. Siiski on mitmeid potentsiaalseid suundi, kuidas arendada ja laiendada GPT mudeli kasutust erinevates valdkondades.

- GPT mudeli peenhäälestamine erinevate valdkondade jaoks - käesoleva töö põhjal saab uurida, kuidas GPT mudelit saab peenhäälestada erinevate valdkondade jaoks, et pakkuda täpsemaid ja sisukamaid tekstilahendusi. Selline lähenemine võimaldaks luua veelgi spetsiifilisemaid ja kvaliteetsemaid tekste, mis on kohandatud konkreetsete valdkondade eripäradele ja nõuetele.
- GPT mudeli kasutamine teistes valdkondades - 13F raportitele sarnaseid tekstilisi ülesandeid leiab ka teistest valdkondadest, nagu meditsiin, teadus, tehnoloogia jne. Käesoleva töö põhjal saab uurida, kuidas GPT mudelit saab kohaldada ja kasutada teistes valdkondades, et luua sisukaid ja täpseid tekste, mis on seotud nendes valdkondades leiduvate andmetega. Teisest küljest võib jätkata automatiseeritud uudisartiklite loomist, kuid saab ka muuta sisu valdkonda ehk loodud artiklid ei pea olema 13F raportitest, vaid hoopis sisuturundus nagu arvamused ja muude tekstide loomine.
- GPT mudeli kasutamine haridus- ja õppematerjalide loomisel - käesoleva töö põhjal võib uurida, kuidas GPT mudelit saab kasutada haridus- ja õppematerjalide loomiseks, mis on kohandatud erinevatele õppeainetele, vanuseastmetele ja õpistiilidele. Selline lähenemine võimaldaks pakkuda õpetajatele ja haridustöötajatele abivahendeid, mis aitavad neil luua kohandatud ja huvipakkuvaid õppematerjale, mis toetavad õpilaste arengut ja huvi õppimise vastu.

5 Kokkuvõte

Käesoleva magistritöö eesmärgiks oli uurida, kuidas automatiseerida 13F raportite analüüs ja uudisartikli loomine ning kuidas võrrelda tehisintellekti abil loodud uudisartikleid finantsanalüütiku poolt loodud uudisartiklitega ja hinnata nende usaldusväärsust ja kvaliteeti. Töös kasutati OpenAI GPT3.5 seeria mudelit Text-Davinci-003, mille abil loodi uudisartiklid 13F raportite analüüsimisest.

Esmalt viidi läbi *desk research* 13F raportitest ja nendega seotud uudisartiklitest, mille käigus koguti andmeid, uuriti uudisartiklite sisu, struktuuri ning teostati andmeanalüüs. Lähtudes võrreldavate uudisartiklite sisust, loodi päringud, mis sisuliselt vastaksid nende artiklite sisule. Seejärel rakendati *prompt engineering* meetodikaid, et töötada välja efektiivsed sisendid GPT mudelile.

GPT mudelit kasutati uudisartiklite ja praktiliste näidete genereerimiseks analüüsitud andmete põhjal. Töö tulemusi valideeriti LHV finantsväljaande näitel, teostati kvalitatiivne analüüs võrreldes tehisintellekti abil loodud lõikude vastavust sisendandmetega ning loovusliku meetodina kasutati GPT mudelit hindamaks enda loodud lõike. Leiti, et GPT mudel suudab luua sisukaid, asjakohaseid ja sisendandmetega vastavuses olevaid artikleid analüüsitud andmete põhjal.

Automatiseerimaks 13F raportite analüüsi ja uudisartikli loomist, liidestati andmepäringuid teostatav Pythoni kood koos Text-Davinci-003 API-ga, mis võimaldas luua kvaliteetseid ja asjakohaseid artikleid analüüsitud andmetest. Töö käigus leiti, et põhjalik sisendandmete kirjeldamine on oluline GPT mudeli kasutamisel, et saavutada ühtlane kvaliteet loodud lõikudes.

Töö tulemused näitavad, et GPT mudel on usaldusväärne 13F raportite analüüsitulemuste tõlgendamisel ning tulemuste põhjalik valideerimine aitab minimaliseerida vigu ja saavutada kõrge kvaliteediga automatiseeritud lahendus. Tegemist on hea praktilise näitega GPT mudeli kasutamisest ja liidestamisest andmeanalüüsi tulemustega. Edasiste potentsiaalsete tööde osas võib uurida GPT mudeli rakendamist teistes valdkondades, nagu ajakirjandus, sisuturundus, haridus- ja õppematerjalide loomine ning teiste andmeallikate analüüsimine või teostada lahendus läbi GPT mudeli peenhäälestamise.

Kasutatud kirjandus

- [1] R. Dale, „GPT-3: What’s it good for?“, *Natural Language Engineering*, pp. 113-118, Cambridge University Press, 2021.
- [2] Augustin Lecler, Loïc Duron, Philippe Soyer, „Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT“, *Diagnostic and Interventional Imaging*, 2023.
- [3] Rodolfo Ocampo, Josh Andres, Adrian Schmidt, Caroline Pegram, Justin Shave, Charlton Hill, Brendan Wright, Oliver Bown, „Using GPT-3 to Achieve Semantically Relevant Data Sonification for an Art Installation“, *Artificial Intelligence in Music, Sound, Art and Design*, Springer, Cham, 2023.
- [4] S. Biswas, „Role of chat GPT in Microbiology: According to chat GPT“, 2023. [Võrgumaterjal]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4405385. [Kasutatud 12 04 2023].
- [5] Nazif Aydın, O. Ayhan Erdem, „A Research On The New Generation Artificial Intelligence Technology Generative Pretraining Transformer 3“, *3rd International Informatics and Software Engineering Conference (IISEC)*, Ankara, Turkey, 2022.
- [6] A. Maalla, C. Zhuang, Q. Feng, „Research on Financial Data Analysis Based on Applied Deep Learning in Quantitative Trading“, *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 1781-1785, Chongqing, China, 2021.
- [7] M. Rizinski, H. Peshov, K. Mishev, L. T. Chitkushev, I. Vodenska, D. Trajanov, „Ethically Responsible Machine Learning in Fintech“, *IEEE Access*, kd. 10, pp. 97531-97554, 2022.
- [8] A. Mehrotra, S. Menon, „Second Round of FinTech - Trends and Challenges“, *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pp. 243-248, Dubai, United Arab Emirates, 2021.
- [9] S. Saravanan, K. Sudha, „GPT-3 Powered System for Content Generation and Transformation“, *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, pp. 514-519, Sonapat, India, 2022.
- [10] „Models - OpenAI API“, OpenAI, [Võrgumaterjal]. Available: <https://platform.openai.com/docs/models>. [Kasutatud 13 03 2023].
- [11] Damocles, „Retrieval Augmented Generation to Replace Google Search, Not ChatGPT“, *Metaroids*, 30 01 2023. [Võrgumaterjal]. Available: <https://metaroids.com/news/retrieval-augmented-generation-to-replace-google-search-not-chatgpt/>. [Kasutatud 23 02 2023].
- [12] Xuanting Chen, Junjie Ye, Can Zu, Nuo Xu, Rui Zheng, Minlong Peng, Jie Zhou, Tao Gui, Qi Zhang, Xuanjing Huang, *How Robust is GPT-3.5 to Predecessors? A*

Comprehensive Study on Language Understanding Tasks, Shanghai, China: Fudan University, 2023.

- [13] Cole E. Short, Jeremy C. Short, „The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation,“ *Journal of Business Venturing Insights*, kd. 19, nr e00388, 2023.
- [14] UNITED STATES SECURITIES AND EXCHANGE COMMISSION, „Form 13F,“ [Võrgumaterjal]. Available: <https://www.sec.gov/files/form13f.pdf>. [Kasutatud 20 02 2023].
- [15] „Financial Statement Data Sets,“ U.S. Securities and Exchange Commission, [Võrgumaterjal]. Available: <https://www.sec.gov/dera/data/financial-statement-data-sets>. [Kasutatud 20 02 2023].
- [16] „EDGAR advanced search,“ [Võrgumaterjal]. Available: <https://www.sec.gov/edgar/search/>. [Kasutatud 25 02 2023].
- [17] „About EDGAR,“ [Võrgumaterjal]. Available: <https://www.sec.gov/edgar/about>. [Kasutatud 02 03 2023].
- [18] „Accessing EDGAR Data,“ U.S. Securities and Exchange Commission, [Võrgumaterjal]. Available: <https://www.sec.gov/os/accessing-edgar-data>. [Kasutatud 25 02 2023].
- [19] „Fails-To-Deliver Data,“ U.S. Securities and Exchange Commission, [Võrgumaterjal]. Available: <https://www.sec.gov/data/foiadocsfailsdatahtm>. [Kasutatud 05 03 2023].
- [20] „Yahoo Finance,“ [Võrgumaterjal]. Available: <https://finance.yahoo.com/>. [Kasutatud 05 03 2023].
- [21] „SQLiteStudio,“ [Võrgumaterjal]. Available: <https://sqlitestudio.pl/about/>. [Kasutatud 28 04 2023].
- [22] „Python,“ [Võrgumaterjal]. Available: <https://www.python.org/about/>. [Kasutatud 29 04 2023].
- [23] „Visual Studio Code,“ [Võrgumaterjal]. Available: <https://code.visualstudio.com/docs>. [Kasutatud 02 05 2023].
- [24] „Yahoo_fin MIT license,“ [Võrgumaterjal]. Available: https://github.com/atreadw1492/yahoo_fin/blob/master/LICENSE.txt. [Kasutatud 03 03 2023].
- [25] „Yahoo_fin,“ [Võrgumaterjal]. Available: https://github.com/atreadw1492/yahoo_fin/. [Kasutatud 03 03 2023].
- [26] „API Reference - OpenAI API,“ OpenAI, [Võrgumaterjal]. Available: <https://platform.openai.com/docs/api-reference/completions>. [Kasutatud 25 03 2023].
- [27] „Pandas DataFrame,“ [Võrgumaterjal]. Available: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>. [Kasutatud 15 04 2023].
- [28] „Pandas,“ [Võrgumaterjal]. Available: <https://pandas.pydata.org/docs/#>. [Kasutatud 15 04 2023].
- [29] „Matplotlib,“ [Võrgumaterjal]. Available: <https://matplotlib.org/stable/index.html>. [Kasutatud 15 04 2023].

- [30] Chenglei Si, Zhe Gan, Zhengyuan Yang, Shuohang Wang, Jianfeng Wang, Jordan Boyd-Graber, Lijuan Wang, „Prompting GPT-3 To Be Reliable,“ University of Maryland, 2022.
- [31] Atsushi Mizumoto, Masaki Eguchi, „Exploring the potential of using an AI language model for automated essay scoring,“ *Research Methods in Applied Linguistics*, kd. 2, nr 2, 2023.
- [32] „Superinvestor list from Dataroma,“ [Võrgumaterjal]. Available: <https://www.dataroma.com/m/managers.php>. [Kasutatud 25 02 2023].
- [33] R. Tõnisson, „USA superinvestorite hiljutised tehingud,“ 2022. [Võrgumaterjal]. Available: <https://fp.lhv.ee/news/newsView?locale=et&newsId=5629604>. [Kasutatud 26 03 2023].
- [34] „Dataroma,“ [Võrgumaterjal]. Available: <https://www.dataroma.com/m/home.php>. [Kasutatud 20 03 2023].
- [35] „Python package Requests,“ [Võrgumaterjal]. Available: <https://pypi.org/project/requests/>. [Kasutatud 20 03 2023].
- [36] A. Cavin, „GPT-3 Parameters and Prompt Design,“ [Võrgumaterjal]. Available: <https://towardsdatascience.com/gpt-3-parameters-and-prompt-design-1a595dc5b405>. [Kasutatud 17 03 2023].
- [37] „GPT fine-tuning,“ OpenAI, [Võrgumaterjal]. Available: <https://platform.openai.com/docs/guides/fine-tuning>. [Kasutatud 25 04 2023].
- [38] Zhao Mandi, Pieter Abbeel, Stephen James, „On the Effectiveness of Fine-tuning Versus Meta-reinforcement Learning,“ University of California, Berkeley, 2022.
- [39] „AI Chat for Financial Analysts,“ [Võrgumaterjal]. Available: <https://finchat.io/>. [Kasutatud 28 04 2023].
- [40] „Whale Wisdom,“ [Võrgumaterjal]. Available: <https://whalewisdom.com/>. [Kasutatud 14 03 2023].
- [41] „FormThirteen,“ [Võrgumaterjal]. Available: <https://formthirteen.com/>. [Kasutatud 14 03 2023].

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Kevin Kiil

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „13F finantsraportite automatiseeritud uudiste genereerimine GPT mudeli abil“ mille juhendaja on Innar Liiv.
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

10.05.2023

¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.