

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS**

**Karolina Januškaitė**

**INFORMACINIŲ PRANEŠIMŲ POVEIKIS DIDŽIŲJŲ**  
**FARMACIJOS KOMPANIJŲ AKCIJŲ KAINOMS**

Baigiamasis magistro projektas

**Vadovai**

Lekt. dr. Mindaugas Kavaliauskas

Prof. dr. Rytis Krušinskas

**KAUNAS, 2017**

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS**

**INFORMACINIŲ PRANEŠIMŲ POVEIKIS DIDŽIŲJŲ  
FARMACIJOS KOMPANIJŲ AKCIJŲ KAINOMS**

Baigiamasis magistro projektas

**Didžiųjų verslo duomenų analitika (kodas 621G12002)**

**Vadovai**

Lekt. dr. Mindaugas Kavaliauskas

Prof. dr. Rytis Krušinskas

**Recenzantai**

Doc. dr. Tomas Ruzgas

Doc. dr. Rasa Norvaišienė

**Projektą atliko**

Karolina Januškaitė

**KAUNAS, 2017**



KAUNO TECHNOLOGIJOS UNIVERSITETAS

Matematikos ir gamtos mokslų fakultetas

(Fakultetas)

Karolina Januškaitė

(Studento vardas, pavardė)

Didžiųjų verslo duomenų analitika (kodas 621G12002)

(Studijų programos pavadinimas, kodas)

Baigiamojo projekto „Informacinių pranešimų poveikis didžiųjų farmacijos kompanijų akcijų kainoms“

### AKADEMINIO SĄŽININGUMO DEKLARACIJA

20 17 m. gegužės 25 d.  
Kaunas

Patvirtinu, kad mano **Karolinos Januškaitės** baigiamasis projektas tema „Informacinių pranešimų poveikis didžiųjų farmacijos kompanijų akcijų kainoms“ yra parašytas visiškai savarankiškai, o visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

\_\_\_\_\_  
(vardą ir pavardę įrašyti ranka)

\_\_\_\_\_  
(parašas)

# TURINYS

SANTRUMPOS .....	10
IŽANGA.....	11
LITERATŪROS APŽVALGA .....	12
1. AKCIJŲ KAINAS LEMIANTYS VEIKSNIAI .....	12
2. AKCIJŲ KAINOS KITIMO IR PROGNOZAVIMO METODIKŲ ANALIZĖ .....	16
2.1. Techninė akcijų kainų analizė.....	17
2.2. Laiko eilučių analizė .....	20
2.3. Sistemų mokymosi ir duomenų tyrybos analizė .....	24
3. TEKSTO TYRYBOS ANALIZĖ .....	26
3.1. Tekstinės informacijos apdorojimas .....	27
3.2. Sentimentų analizė .....	28
3.3. Tekstinės informacijos poveikio nustatymas .....	29
MEDŽIAGOS IR TYRIMŲ METODAI .....	31
4. PRADINIAI DUOMENYS IR JŲ APDOROJIMAS .....	31
4.1. Didžiųjų farmacijos kompanijų akcijų laiko eilutės .....	31
4.2. Informacinių pranešimų atrinkimas ir gavyba .....	32
4.3. Teksto sentimentų vertinimas informaciniuose pranešimuose .....	34
5. TYRIMO METODAI .....	35
5.1. Akcijų laiko eilučių analizė .....	35
5.2. Latentinio Dirichlet'o pasiskirstymo modelio taikymas tekstinei informacijai.....	35
5.3. ARIMA modelio taikymas naujienų poveikio akcijų kainoms analizei .....	38
TYRIMŲ REZULTATAI IR JŲ APTARIMAS.....	40
6. TYRIMO REZULTATAI .....	40
6.1. Informacinių pranešimų pagal sentimentus įtakos akcijų kainoms vertinimas .....	42
6.2. Informacinių pranešimų pagal poliariškumą įtakos akcijų kainoms vertinimas.....	53
IŠVADOS.....	64
LITERATŪROS SĄRAŠAS.....	66

PRIEDAI .....	71
1 PRIEDAS .....	72
2 PRIEDAS .....	73
3 PRIEDAS .....	75
4 PRIEDAS .....	76
5 PRIEDAS .....	78
6 PRIEDAS .....	81

Januškaitė, Karolina. Informacinių pranešimų poveikis didžiųjų farmacijos kompanijų akcijų kainoms. Magistro baigiamasis projektas / vadovai Lekt. dr. Mindaugas Kavaliauskas, Prof. dr. Rytis Krušinskas; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Mokslo kryptis ir sritis: Fiziniai mokslai, Matematika (01 P)

Reikšminiai žodžiai: teksto tyryba, laiko eilutės, ARIMA, sentimentai, poliariškumas, akcijų kaina.

Kaunas, 2017. 85 p.

## **SANTRAUKA**

Didžiųjų farmacijos kompanijų akcijų kainų analizės aktualumas pastebimas dėl keletą pastarųjų metų stebimo ženkliaus šių akcijų kainų augimo. Nuo 2014 metų per didesnę nei metų laikotarpį šio sektoriaus įmonių akcijų kainos išaugo apie 3 kartus. Investuotojai, prieš įsigydamiesi akcijas, atlieka įvairiapusę analizę apie kompanijas, tačiau pagrindinė šios analizės dalis yra susijusi su tekstinės informacijos analizavimu. Todėl šiame darbe, taikant teksto tyrybai tinkamus metodus, siekiama nustatyti ar egzistuoja informacinių pranešimų poveikis akcijų kainoms.

Tyrimui buvo pasirinktas 2014 – 2016 metų laikotarpis dėl pastebimo akcijų kainų augimo, didžiosios farmacijos kompanijos tyrimui pasirinktos pagal 2016 metų pelno dydį bei rinkos įtaką. Informaciniai pranešimai buvo vertinami atsižvelgiant į jų pozityvumą ir negatyvumą – tai buvo atlikta vertinant straipsnių žodžius pagal jų sentimentus ir poliariškumą.

Atlikus tyrimą pastebėta, jog tik „AstraZeneca“ bei „Bayer“ kompanijų atveju egzistuoja informacinių pranešimų poveikis akcijų kainoms. Nustatyta, jog likusių kompanijų akcijų kainai informaciniai pranešimai statistiškai reikšmingo poveikio neturi. Taip pat buvo pastebėta, jog „Bayer“ kompanijos atveju informacinių pranešimų poveikio statistinis reikšmingumas skiriasi priklausomai nuo naujienų pozityvumo ir negatyvumo vertinimo: informacinius pranešimus vertinant pagal sentimentus nustatyta, jog informacinių pranešimų poveikis „Bayer“ akcijų kainoms yra statistiškai reikšmingas, taip pat nustatyta, jog viena diena pavėlintų pranešimų poveikis taip pat statistiškai reikšmingas. Tačiau vertinant naujienas pagal poliariškumą nustatyta, jog informacinių pranešimų bei pavėlintų jų reikšmių kintamieji nėra statistiškai reikšmingi. Taip pat atliktu tyrimu patvirtinta, jog tyrime analizuotų didžiųjų farmacijos kompanijų akcijų kainų kitimui yra būdingas atsitiktinis kitimas (angl. random walk).

Januškaitė, Karolina. THE IMPACT OF INFORMATION ON MAJOR PHARMACEUTICS COMPANIES' SHARE PRICES Master's thesis in Business Big Data Analytics / supervisor assoc. Lect. dr. Mindaugas Kavaliauskas, Prof. dr. Rytis Krušinskas. The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Research area and field: Natural Sciences, Mathematics (01 P)

Key words: text mining, time series, ARIMA, sentiments, polarity, stock price.

Kaunas, 2017. 85 p.

## SUMMARY

The relevance of major pharmaceuticals companies' stock prices is noticeable due to major rise in stock prices in several previous years. From 2014 in almost a year stock prices in this sector has risen about 3 times. Before buying stocks, investors make a thorough analysis about the companies, but main part of this analysis consist of analyzing textual information. Therefore, in this master thesis, with the appropriate methods for text mining, is intended to define whether there exists news impact to stock prices.

For this research timeframe of 2014 – 2016 years has been chosen due to noticeable rise in stock prices, major pharmaceuticals companies have been chosen by the amount of revenue in 2016 and by impact in the market. Information signals were evaluated by their positivity and negativity – it was done by evaluating each word in an article by its sentiment and polarity.

After the research, it has been noticed, that only in “AstraZeneca” and “Bayer” companies' cases news impact on stock prices exists. It has been ascertained, that for the rest of the companies' information signals' impact on stock prices is not statistically significant. It has been also noticed, that in “Bayer” company's case news impact statistical significance differs due to evaluation of news positivity and negativity: when information signals are evaluated by sentiments it has been estimated, that news impact for “Bayer” stock prices is statistically significant, also it has been identified that one day news' lag impact is statistically significant. However, evaluating information signals by polarity, news and their lagged values are not statistically significant. It was also confirmed with this research, that major pharmaceuticals companies', that have been analyzed, stock prices alteration is a random walk process.

## LENTELIŲ SĄRAŠAS

4.2.1. lentelė. Straipsnių pasiskirstymas pagal temą tiriamuoju laikotarpiu .....	33
5.2.1. lentelė. LDA modelio temos žodžiai bei jų pasiskirstymo straipsniuose tikimybės.....	37
5.2.2. lentelė. Straipsnių skaičius sentimentų laiko eilutės sudarymui .....	38
6.1.1. lentelė. Papildyto Dikio – Fulerio testo reikšmės akcijų kainų laiko eilutėms .....	41
6.1.1.1. lentelė. Sentimentų modelio „AstraZeneca“ kompanijai rezultatai .....	43
6.1.2.1. lentelė. Sentimentų modelio „Bayer“ kompanijai rezultatai .....	44
6.1.3.1. lentelė. Sentimentų modelio „Gilead“ kompanijai rezultatai.....	45
6.1.4.1. lentelė. Sentimentų modelio „GlaxoSmithKline“ kompanijai rezultatai .....	46
6.1.5.1. lentelė. Sentimentų modelio „Johnson & Johnson“ kompanijai rezultatai .....	47
6.1.6.1. lentelė. Sentimentų modelio „Merck & Co“ kompanijai rezultatai .....	48
6.1.7.1. lentelė. Sentimentų modelio „Novartis“ kompanijai rezultatai.....	49
6.1.8.1. lentelė. Sentimentų modelio „Pfizer“ kompanijai rezultatai .....	50
6.1.9.1. lentelė. Sentimentų modelio „Roche“ kompanijai rezultatai .....	51
6.1.10.1. lentelė. Sentimentų modelio „Sanofi“ kompanijai rezultatai .....	52
6.2.1.1. lentelė. Poliariškumo modelio „AstraZeneca“ kompanijai rezultatai .....	53
6.2.2.1. lentelė. Poliariškumo modelio „Bayer“ kompanijai rezultatai.....	54
6.2.3.1. lentelė. Poliariškumo modelio „Gilead“ kompanijai rezultatai.....	55
6.2.4.1. lentelė. Poliariškumo modelio „GlaxoSmithKline“ kompanijai rezultatai .....	56
6.2.5.1. lentelė. Poliariškumo modelio „Johnson & Johnson“ kompanijai rezultatai .....	57
6.2.6.1. lentelė. Poliariškumo modelio „Merck & Co“ kompanijai rezultatai .....	58
6.2.7.1. lentelė. Poliariškumo modelio „Novartis“ kompanijai rezultatai.....	59
6.2.8.1. lentelė. Poliariškumo modelio „Pfizer“ kompanijai rezultatai.....	60
6.2.9.1. lentelė. Poliariškumo modelio „Roche“ kompanijai rezultatai .....	61
6.2.10.1. lentelė. Poliariškumo modelio „Sanofi“ kompanijai rezultatai .....	62



## PAVEIKSLŲ SĄRAŠAS

2.2.1 pav. Trijų sluoksnių tiesioginis atgalinio mokymo neuroninis tinklas.....	23
2.2.2 pav. Wang hibridinio modelio loginė schema.....	24
5.2.1 pav. LDA modelio loginė schema.....	36
6.1. pav. Grynųjų akcijų kainų laiko eilučių grafinis vaizdas .....	40
6.2. pav. „Johnson & Johnson“ autokoreliacijos ir dalinės autokoreliacijos grafinis vaizdas .....	41
6.3. pav. „Pfizer“ autokoreliacijos ir dalinės autokoreliacijos grafinis vaizdas .....	42
6.1.1. pav. „AstraZeneca“ sentimentų modelio paklaidų pasiskirstymo grafikas.....	43
6.1.2.1. pav. „Bayer“ sentimentų modelio paklaidų pasiskirstymo grafikas .....	44
6.1.3.1. pav. „Gilead“ sentimentų modelio paklaidų pasiskirstymo grafikas .....	45
6.1.4.1. pav. „GlaxoSmithKline“ sentimentų modelio paklaidų pasiskirstymo grafikas .....	46
6.1.5.1. pav. „Johnson & Johnson“ sentimentų modelio paklaidų pasiskirstymo grafikas.....	47
6.1.6.1. pav. „Merck & Co“ sentimentų modelio paklaidų pasiskirstymo grafikas .....	48
6.1.7.1. pav. „Novartis“ sentimentų modelio paklaidų pasiskirstymo grafikas .....	49
6.1.8.1. pav. „Pfizer“ sentimentų modelio paklaidų pasiskirstymo grafikas .....	50
6.1.9.1. pav. „Roche“ sentimentų modelio paklaidų pasiskirstymo grafikas .....	51
6.1.10.1. pav. „Sanofi“ sentimentų modelio paklaidų pasiskirstymo grafikas.....	52
6.2.1.1. pav. „AstraZeneca“ poliariškumo modelio paklaidų pasiskirstymo grafikas .....	53
6.2.2.1. pav. „Bayer“ poliariškumo modelio paklaidų pasiskirstymo grafikas.....	54
6.2.3.1. pav. „Gilead“ poliariškumo modelio paklaidų pasiskirstymo grafikas.....	55
6.2.4.1. pav. „GlaxoSmithKline“ poliariškumo modelio paklaidų pasiskirstymo grafikas .....	56
6.2.5.1. pav. „Johnson & Johnson“ poliariškumo modelio paklaidų pasiskirstymo grafikas .....	57
6.2.6.1. pav. „Merck & Co“ poliariškumo modelio paklaidų pasiskirstymo grafikas .....	58
6.2.7.1. pav. „Novartis“ poliariškumo modelio paklaidų pasiskirstymo grafikas.....	59
6.2.8.1. pav. „Pfizer“ poliariškumo modelio paklaidų pasiskirstymo grafikas .....	60
6.2.9.1. pav. „Roche“ poliariškumo modelio paklaidų pasiskirstymo grafikas .....	61
6.2.10.1. pav. „Sanofi“ poliariškumo modelio paklaidų pasiskirstymo grafikas .....	62

## **SANTRUMPOS**

AKT – atsitiktinio kitimo teorija

ARIMA – autoregresinis integruotas slenkančių vidurkių modelis

ANN – dirbtiniai neuroniniai tinklai

ERH – efektyvioji rinkos hipotezė

GARCH – apibendrintas autoregresinis sąlyginio heteroskedastiškumo modelis

LDA – Latentinis Dirichlet'o pasiskirstymo modelis

RMSE – liekanų vidutinės kvadratinės paklaidos

SVM – atraminių vektorių mašinos modelis

## IŽANGA

Jau ilgą laiką investavimas į įvairius vertybinius popierius yra aktualus būdas didinti savo turimą materialų kapitalą. Priklausomai nuo kiekvieno norinčio investuoti rizikos toleravimo, investuotojai renkasi ar skirti lėšas obligacijų įsigijimui, ar investuoti į akcijų biržoje listinguojamų kompanijų akcijas. Prieš pasirenkant vieną iš šių galimybių, investuotojai stengiasi įvertinti ekonominę situaciją, atsižvelgti į kompanijų, kurių akcijas svarstoma įsigyti, finansinę situaciją. Vienas iš plačiausiai naudojamų metodų yra naujienų ir kitos tekstinės informacijos analizė, padedanti nustatyti planuojamos investicijos saugumą, galimą riziką bei norimos finansinės grąžos galimybes.

Keletą pastarųjų metų yra pastebimas sveikatos priežiūros sektoriaus augimas, kuris sulaukė didelio investuotojų dėmesio. Aktualu pastebėti, jog nuo 2014 metų iki 2016 metų vidurio šio sektoriaus akcijų kainos išaugo beveik tris kartus. Neretai pastebima, jog akcijų biržoje gali pasitaikyti spekuliatyvaus akcijų įsigijimo ar pardavimo, siekiant koreguoti akcijų kainą rinkoje. Akcijų kainos taip pat gali kisti dėl įvairių spaudoje pasirodančių pranešimų apie pasikeitimus kompanijos viduje, finansinės situacijos pokyčius ir kt. Dėl šios priežasties aktualu nustatyti, ar analizuojant informacinius pranešimus apie šiam sektoriui priklausančias kompanijas buvo galima įžvelgti tokį akcijų kainų augimo potencialą, t. y. norima nustatyti ar informaciniai pranešimai galėjo turėti poveikį akcijų kainoms.

**Darbo tema:** informacinių pranešimų poveikio didžiųjų farmacijos kompanijų akcijų kainoms analizė.

**Darbo tikslas:** nustatyti ar informaciniai pranešimai turi poveikį didžiųjų farmacijos kompanijų akcijų kainoms.

### **Darbo uždaviniai:**

- Nustatyti akcijų kainos analizei taikomas metodus;
- Išsiaiškinti teksto tyrybai taikomus metodus;
- Atrinkti informacinius pranešimus, kurie gali daryti įtaką didžiųjų farmacijos kompanijų akcijoms;
- Įvertinti informacinių pranešimų pozityvumą ir negatyvumą;
- Sudaryti modelį, kuriuo bus tikrinamas informacinių pranešimų poveikis kiekvienos iš didžiųjų farmacijos kompanijų atveju.

## LITERATŪROS APŽVALGA

Informacija nuo neatmenamų laikų buvo svarbi žmonijai – aktualios informacijos žinojimas įgalina asmenį priimti sprendimą, atsižvelgus į galimas aplinkybes, kurias jis sužino gavęs informaciją. Akmens amžiuje reikiamos informacijos žinojimas buvo aktualus išgyvenimui ir fizinių poreikių patenkinimui - žinojimas apie galimus maisto išteklius, kaip juos gauti ir pan. Tačiau šiandien informacijos svarba kiekvieno žmogaus gyvenime yra daug didesnė nei kada nors anksčiau – informacija reikalinga ne tik išlikimui, bet ir gyvenimo kokybės gerinimui, įvairių (pvz. išsilavinimo, finansinių ir t. t.) sprendimų priėmimui. Taip pat svarbu atsižvelgti, jog bėgant laikui keitėsi ir informacijos perdavimo bei saugojimo būdai.

Anksčiau didžioji dalis turimos informacijos buvo perduodama žodžiu ir niekur nesaugoma, tačiau taip buvo iki rašto sukūrimo. Šiandien didžioji informacijos dalis yra saugoma raštu, tačiau sparčiai vystantis informacinėms technologijoms, plečiantis interneto tinklų aprėpčiai, informacijos saugojimas bei perdavimas dažniausiai vyksta internetu, o informacija pasiekiamą internetiniuose puslapiuose. Savaime suprantama, jog palengvėjus informacijos srautų perdavimui bei jos saugojimui, prieinamos informacijos kiekis ženkliai išaugo tiek, jog jį, be papildomų pagalbinių priemonių (pvz., programų, leidžiančių filtruoti turinį) apdoroti yra neįmanoma. Todėl esant tokiam dideliame informacijos kiekiui, gauti reikiamą informaciją, reikalingą sprendimų priėmimui, yra aktualu ir labai vertinga.

Kaip jau minėta anksčiau, informacija reikalinga priimant įvairius finansinius sprendimus, pavyzdžiui, asmeniui, turinčiam laisvų piniginių lėšų yra aktuali informacija apie galimybes investuoti turimas lėšas ir už tai gauti atlygį kuria nors materialia forma – pinigais ar turtu. Viena iš galimybių investuoti laisvas lėšas yra įmonių akcijų pirkimas. Tačiau, kaip pastebi Keller'is [1], investavimas į akcijas yra rizikingas, kadangi į akcijas investuojantys asmenys yra linkę prisiimti didesnę riziką nei kiti investuotojai. Galima rizika yra susijusi su daugybe veiksnių, veikiančių akcijos kainą. Akcijų kainoms poveikį gali daryti ekonominiai, politiniai ir kiti faktoriai. Todėl verta plačiau apžvelgti investavimo į akcijas ypatumus bei aptarti veiksnius, darančius įtaką akcijų kainoms.

### 1. AKCIJŲ KAINAS LEMIANTYS VEIKSNIAI

Finansų rinkoje esama daugybės įvairių vertybinių popierių bei išvestinių finansinių priemonių, į kurias galima investuoti turimas lėšas. Vertybiniai popieriai skaidomi į dvi pagrindines kategorijas: skolos ir nuosavybės. Pagrindiniai skolos vertybiniai popieriai yra obligacijos, nuosavybės – akcijos. Investuotojų pasirinkimas, kokio tipo vertybinius popierius pirkti, yra nulemtas dviejų pagrindinių veiksnių, lemiančių investicijos patrauklumą: investicijos

grąžos bei jos rizikos. Remiantis Kumar'u [2], akcijos riziką sudaro dvi dedamosios: sisteminė ir nesisteminė rizika. Sisteminė rizika veikia daugelį, biržoje listinguojamų, akcijų ir negali būti diversifikuojama. Sisteminė rizika pasireiškia kaip bendrojo vidaus produkto, infliacijos ir kitų makroekonominių veiksnių poveikis akcijų kainoms. Nesisteminė rizika laikomi įvairūs neprognozuojami pokyčiai, kurie daro įtaką tik ribotam skaičiui akcijų, pavyzdžiui, darbuotojų streikai, apyvartinių lėšų trūkumas ir kt. Šio tipo rizika gali būti diversifikuojama, todėl šios rizikos sumažinimas gali sąlygoti didesnę investicinę grąžą.

Vertinant akcijų kainos prognozavimą, ilgą laiką buvo remiamasi dvejomis teorijomis – efektyviosios rinkos hipoteze (angl. Efficient Market Hypothesis) bei atsitiktinio kitimo teorija (angl. Random Walk Theory). Efektyvioji rinkos hipotezė (toliau ERH) buvo pasiūlyta Fama'os [3]. ERH teorijoje teigiama, jog akcijos kaina atspindi visą rinkoje esančią informaciją. Rinkoje pasirodžius naujai informacijai, galinčiai veikti konkrečios akcijos kainą, rinka iš karto prisitaiko prie pasikeitusių sąlygų ir naujos informacijos įtaka atsispindi akcijos kainoje (pvz. jei nauja informacija yra neigiama, akcijos kaina krenta). Todėl ERH teorijoje teigiama, jog akcijos kainos visuomet yra teisingos ir rinka yra efektyvi, t. y. niekas negali gauti papildomos naudos iš akcijų kainos pokyčio. ERH teigiama, jog rinkos efektyvumas priklauso nuo informacijos, kuri, kaip teigė Fama'as [3], turi būti atspindėta akcijos kainoje. Remiantis Elton'u [4] rinkos efektyvumas skirstomas į tris efektyvumo formas:

- Silpna efektyvumo forma. Teigiama, jog esamu momentu akcijos kainą atspindi visą aktualią informaciją. Todėl manoma, jog praeities akcijos kainos taip pat kiekvienu laiko momentu atspindėjo visą aktualią informaciją ir praeities akcijos kainos neturi poveikio dabartinei akcijos kainai. Tai taip pat reiškia, jog praeities bei dabartinė informacija nėra susijusios, kadangi esant šiam sąryšiui, nauja informacija nėra nauja [5]. Todėl galima sakyti, kad šios rinkos efektyvumo forma yra teigiama, jog akcijos kainų kitimas yra visiškai atsitiktinis procesas. Šiuo požiūriu EHR teorija atitinka AKT, kadangi abi teorijos pažymi, jog akcijos kainos kitimas yra atsitiktinis procesas ir kainos kitimas negali būti prognozuojamas, remiantis praeities informacija.
- Pusiau stipri forma. Šiuo atveju teigiama, jog dabartinė akcijos kaina yra nulemta istorinių kainų, tačiau ją taip pat veikia visa viešai prieinama informacija. Tokia efektyvumo forma remiasi paklausos ir pasiūlos dėsniais [5]. Manoma, jog akcijos kaina visuomet yra pasiūlos ir paklausos pusiausvyros taške. Pasirodžius naujai informacijai, rinka prisitaiko ir akcijos kaina pasislenka į naują pusiausvyros tašką, kuriame akcijos kaina atspindi visą, įskaitant ir naują, informaciją.
- Stipri forma – esant tokio efektyvumo rinkai, teigiama, jog akcijų kainą veikia visa galima – vieša ir privati – informacija. Manoma, jog privati informacija gali sąlygoti galimybę

gauti papildomą grąžą iš akcijų, kadangi privati, mažam investuotojų ratui žinoma, informacija gali padėti priimti investavimo sprendimus turint tam tikrą pranašumą prieš kitus investuotojus. Remiantis Majouji [5], jei esama investuotojų, turinčių informaciją apie tai, jog tam tikros akcijos kaina yra žemesnė, nei turėtų būti, t. y. investuotojų turima informacija yra žinoma, jog akcijos kaina turėtų kilti, žinantys šią informaciją investuotojai perka konkrečią akciją rinkoje esančia kaina. Akcijos perkamos iki tol, kol privati informacija tampa vieša ir žinoma visiems investuotojams ir akcijos kaina pakyla iki pusiausvyros, t. y. rinkoje nusistovi teisinga, visą viešą informaciją atspindinti kaina.

Taigi remiantis ERH, teigiama, jog rinka visuomet yra efektyvi, t. y. akcijos kainos visuomet teisingos ir atspindi atitinkamą informaciją. Dėl šios priežasties manoma, jog akcijų kainos prognozavimas nėra įmanomas. Tačiau esama šios teorijos skeptikų, kurie bandė teoriją paneigti. Malkiel'as [6], atlikdamas kitų autorių darbų analizę, pastebėjo, jog akcijų skaidymas (angl. stock split) dažniausiai yra siejamas su laukiamais didesniais dividendais bei teigiamais įmonės vadovybės lūkesčiais ateičiai, o tai atitinkamai gali lemti akcijos kainos padidėjimą. Todėl iš anksto žinant, jog bus atliekamas skaidymas, kaip ir stiprios efektyvumo formos rinkos atveju žinant išskirtinę informaciją, įsigijus tokią akciją prieš paskelbiant skaidymą, galima gauti papildomą grąžą. Tačiau jei ši informacija nežinoma iš anksto, Malkiel'o [6] teigimu, paskelbus akcijų skaidymą rinka prisitaiko iš karto. Timmermann'as [7] taip pat pastebėjo, jog didelė dalis mokslininkų abejoja ERH teorijos patikimumu, todėl kiekvienas stengiasi sukurti savo modelį siekiant gauti grąžos iš akcijų prognozavimo.

Nemažos dalies mokslininkų noras patikrinti šias teorijas asmeniškai bei išsiaiškinti ar rinka visais atvejais yra efektyvi paskatino tirti akcijų prognozavimo galimybes plačiau. Atliktais Malkiel'as [8], Huang'as [9] tyrimais nustatyta, jog akcijų kainą galima numatyti. Vis dėl to, prognozuoti akcijos kainą ateičiai yra labai sudėtinga, kadangi finansų rinka, taip pat ir akcijų birža, yra veikiamą daugybės išorinių veiksnių – ekonominių, politinių, investuotojų lūkesčių ir kt. Taip pat šis poveikis dažniausiai nėra tiesinis ir tiesioginis, todėl jį sunkiau įvertinti. Timmermann'as [7] pastebėjo, jog didžioji dalis atliktų tyrimų remiasi praeities akcijų kainų analize, tačiau jis taip pat paminėta tai, jog tokios analizės dažniausiai turi tik labai mažą pasisekimą. Dėl šios priežasties analizė, įtraukiant informaciją apie išorinius veiksnus, yra vis aktualesnė.

Nemaža dalis autorių, Adam'as [10], Muzammal'as [11], Jareno [12] išskiria makroekonominius rodiklius, kurie yra pagrindiniai veiksniai, darantys didžiausią įtaką akcijų kainai. Remiantis Adam'u [10], pramonės gamybos pokyčiai daro teigiamą įtaką tikėtinaai akcijų grąžai – grąža didėja, tačiau infliacija sąlygoja akcijų grąžos sumažėjimą. Toks infliacijos poveikis paaiškinamas tuo, jog didėjanti infliacija skatina perkamosios galios mažėjimą, taigi ir šiuo atveju, jei infliacija auga, gaunama grąža iš akcijų sumažėja. Muzammal'u [11] išskiria ir kitų, darančių

poveikį akcijoms, makroekonominių veiksnių – bendrąjį vidaus produktą, palūkanų normas, užimtumą ir kt. Šių veiksnių įtaką lygiai taip pat patvirtina ir Jareno [12]. Tačiau savo tyrime jis taip pat patvirtino makroekonominių veiksnių poveikį akcijų kainoms, remdamasis akcijų biržos bei ekonomikos ciklų tarpusavio palyginimu – buvo pastebėta, jog akcijų biržos ciklas juda pagal ekonomikos ciklą, tačiau su 6-12 mėnesių atsilikimu. Taigi galima teigti, jog makroekonominių veiksnių poveikis egzistuoja, tačiau yra skirtingas priklausomai nuo veiksnių bei ekonomikos, kurioje veikia akcijas listinguojančios įmonės.

Muzammal'o [11] darbe taip pat buvo tiriamas įmonės finansinių rodiklių poveikis jų akcijų kainoms. Muzammal'as [11] vertino įmonės likvidumo rodiklį (angl. liquidity), finansinį svertą (angl. leverage), pelningumą, pelno vienai akcijai (angl. Earning Per Share) augimą, rinkos kapitalizacijos dydį, dividendus. Atliktu tyrimu nustatyta, jog net 65 proc. akcijos kainos variacijos gali būti paaiškinta pastaraisiais įmonės finansiniais rodikliais.

Kaip jau minėta anksčiau, esama autorių darbų, kurie savo tyrimuose analizavo išorinių, makroekonominių ir mikroekonominių veiksnių įtaką. Tačiau augant interneto, bei naujienų ir įvairios informacijos prieinamumui tekstinė informacija tapo dar vienu analizės šaltiniu, siekiant nustatyti kas veikia akcijų kainas. Tekstinės informacijos įtaka akcijų kainoms tampa vis aktualesniu tyrimų objektu, kadangi tekstinės informacijos kiekis bei prieinamumas yra ženkliai didesnis nei skaitinės įvairių rodiklių informacijos. Tekstinė informacija, naudojama akcijų kainos analizei, vyrauja nuo naujienų apie ekonominių ir finansinių rodiklių pokyčius ([13], [14]), atskiro sektoriaus ar šakos naujienas ([15], [16]) iki sentimentus atspindinčių naujienų ([17], [18], [19]). Svarbu pastebėti, jog informacinių naujienų analizė yra sudėtingesnė, kadangi tekstinė informacija dažnai neturi griežtos struktūros, priešingai nei skaitinė [20]. Tačiau informacinis pranešimas ar naujiena apie tam tikrą rodiklį gali atskleisti daugiau informacijos nei skaitinė rodiklio reikšmė. Dėl šios priežasties tekstinė analizė yra vis populiareesnė.

Informaciniai pranešimai akcijos kainą gali veikti nebūtinai tiesiogiai – pranešimai apie ekonominės ar kitos situacijos pokyčius dažniausiai sąlygoja investuotojų nuomonės formavimąsi, o tai savo ruožtu veikia investuotojų apsisprendimą pirkti (parduoti) ar laikyti turimas akcijas. Svarbu pastebėti tai, jog informaciniai pranešimai gali turėti skirtingą poveikį investuotojų nuomonės bei elgesio formavimui skirtingu laikotarpiu. Investuotojams, kuriems priimtinas trumpalaikis investavimas, t. y. akcijos įsigijamos su tikslu jas greitai parduoti pakilus kainai, informaciniai pranešimai labai svarbūs. Įvairios naujienos, kurios gali paveikti turimų akcijų kainą, gali būti vertinamos kaip signalai: jei informacija teigiama – tai signalas, jog akcijos kaina ateityje turi kilti, priešingai, jei naujienos neigiamos – tai signalas apie galimą akcijos kainos kritimą. Atsižvelgiant į dabartinį informacijos srautą, kasdien esama daugybės teigiamų bei neigiamų

signalų, todėl šiuo atveju pagrindinis pasirinkimas – pirkti, parduoti ar laikyti akcijas – tenka investuotojui.

Investuotojų nuomonę, literatūroje ([21], [22], [23]) dažnai įvardinamą investuotojų sentimentu (angl. investor sentiment), nulemia ne tik esama informacija ar pranešimai apie pasikeitimus rinkose, tačiau taip pat ir kiekvieno investuotojo asmeninės savybės (pvz. noras rizikuoti, investavimo patirtis, nuojauta ir kt.). Investuotojų sprendimų svarba akcijų kainoms pastebėta jau seniai. Finansų rinkose yra galimos dvi situacijos: vertybinių popierių kainos auga ar tikimasi, jog jos augs, arba kainos krenta. Remiantis Finansų terminų žodyne [24] pateiktomis sąvokomis, tokios rinkoje egzistuojančios situacijos atitinkamai vadinamos bulių bei lokių rinkomis. Situacija, kai rinkoje akcijų kainos krenta – vadinama lokių rinka. Akcijų kainų kritimą gali iššaukti investuotojų neigiamos ateities perspektyvos dėl kurių investuotojai parduoda turimas akcijas. Priešingai, jei akcijų kainos kyla – rinka priklauso buliams. Aktualu tai, jog pagal požiūrį į rinkos perspektyvas – ar tikimasi akcijų kainų augimo, ar kritimo – investuotojai taip pat priskiriami buliams ir lokiams. Šiuo skaidymu yra patvirtinama, jog investuotojų nuomonė turi didelę reikšmę akcijų kainos bei kiekio pokyčiams rinkoje.

Taigi, kaip jau buvo minėta anksčiau, tekstinės informacijos duomenys dažniausiai neturi aiškos struktūros. Todėl priklausomai nuo to, ar akcijos kainą veikiančius veiksnius aprašantys duomenys turi struktūrą ar ne, galima išskirti dvi duomenų grupes: duomenys, turintys struktūrą bei duomenys, kurie neturi struktūros ar jų struktūra nėra griežta. Pirmajai grupei galima priskirti mikroekonominius ir makroekonominius veiksnius. Tokie rodikliai, kurie turi kiekybinę išraišką, beveik visuomet pateikiami sugrupuoti pagal stulpelius ar eilutes, yra saugomi duomenų bazėse, todėl Beal'o [25] teigimu tokie duomenys turi juos aprašančią struktūrą. Remiantis Tanwar'u [26] informaciniai pranešimai ir naujienos internete, socialinių tinklų pranešimai (sentimentai) atitinkamai turi dalinę struktūrą ir neturi jokios struktūros. Todėl tokie duomenys gali būti priskiriami antrai grupei ir įvardinami kaip kokybiniai, kadangi jų negalima pamatuoti. Dėl savo sudėtingumo, antrosios grupės veiksnių – informacinių pranešimų ir naujienų – poveikis akcijų kainoms yra mažiau ištirtas nei mikroekonominių ar makroekonominių veiksnių. Tačiau svarbu pastebėti, jog atlikti tyrimai ([15], [16]), kuriuose analizuojamas informacijos bei jos signalų poveikis akcijų kainai, turėjo pasisėkimą, todėl aktualu šios grupės veiksnius analizuoti plačiau. Siekiant sėkmingai atlikti tokią analizę, toliau pateikiama tokio pobūdžio tyrimų analizė bei jų sėkmingumo vertinimas.

## **2. AKCIJŲ KAINOS KITIMO IR PROGNOZAVIMO METODIKŲ ANALIZĖ**

Teorijoje pateikiamos dvi pagrindinės akcijų kainos analizės kryptys: techninė bei fundamentalioji. Pagrindinis šių analizės krypčių skirtumas yra tai, jog techninė analizė remiasi



tik praeities akcijos kainos analize, o atliekant fundamentaliąją analizę taip pat atsižvelgiama į galimus išorinius veiksnius, darančius įtaką akcijos kainos pokyčiams. Taip pat verta pastebėti, jog esama autorių ([27], [28]) pritariančių Hellstrom'o [29] apibrėžtoms akcijų kainos analizės ir prognozavimo metodikoms, kurios yra skaidomos į:

1. Techninę analizę;
2. Laiko eilučių prognozavimą;
3. Sistemų mokymąsi ir duomenų tyrybą (angl. Machine Learning and Data Mining).

Šios akcijų kainos prognozavimo metodikos tarpusavyje yra labai stipriai susijusios, tačiau jos taip pat turi ir nemažai skirtumų. Svarbu tai, jog kaip buvo minėta anksčiau, akcijų kainos analizei ir prognozavimui naudojama informacija gali būti dviejų tipų: turinti struktūrą (kiekybinę) ir turinti dalinę ar neturinti struktūros (kokybinę) informacija. Siekiant suprasti šių metodikų tarpusavio panašumus bei esminius skirtumus, verta kiekvieną iš jų aptarti plačiau bei apžvelgti anksčiau atliktų tyrimų sėkmingumą taikant kiekvieną iš šių metodikų.

## 2.1. Techninė akcijų kainų analizė

Techninė akcijų kainos analizė, pasak Edwards'o [30], apibrėžiama kaip mokslas apie savaimę akcijų rinkoje vykstančius veiksmus. Ši analizė skirta užfiksuoti akcijų kainos, kiekio kitimą ir jį pateikti grafiškai. Turint tiriamų rodiklių pokyčių grafiką yra prognozuojamas rodiklių kitimas ateityje – numatoma labiausiai tikėtina tendencija (angl. trend). Šiuo atveju svarbu pabrėžti, jog siekiama prognozuoti ne tikslią akcijos kainą, tačiau tik numatyti būdingus dėsningumus. Techninė analizė remiasi istorinių akcijos kainų analize – manoma, jog praeities dėsningumai turi tendenciją pasikartoti, todėl atsižvelgiant į esamą kainą siekiama pagal istorinius duomenis rasti panašią situaciją ir remiantis ja numatyti kitimo tendenciją ateičiai. Šiai analizės metodikai taip pat priskiriama Elioto bangos teorija (angl. Elliot wave theory), kurios pagrindinė esmė – daugybės akcijų istorinės kainos nubrėžtos viename grafike. Tai leidžia matyti visos rinkos tendenciją.

Vis dėl to, techninės analizės efektyvumas ir svarba dažnai vertinami prieštarinai. Hudson'as [31] savo atliktu tyrimu siekė nustatyti, ar prekybos akcijomis taisyklių taikymas yra efektyvesnis nei „pirk ir laikyk“ (angl. buy-and-hold) strategija, kuri remiasi įprastu akcijų įsigijimu ir jų laikymu. Tyrimas buvo atliekamas sudarius akcijų indekso FT-30 (angl. Financial Times Industrial Ordinary Index) laiko eilutę nuo 1935 m. liepos mėn. iki 1994 m. sausio mėn. Šis indeksas yra apskaičiuojamas pagal 30 Jungtinės Karalystės pramonės šakų kompanijų akcijų kainas. Indekso skaičiavimą Hudsonas atliko pagal formulę

$$FT30_t = FT30_{t-1} \sqrt[30]{\frac{S_{t,1}}{S_{t-1,1}} \times \frac{S_{t,2}}{S_{t-1,2}} \times \dots \times \frac{S_{t,30}}{S_{t-1,30}}} , \quad (1)$$

kurioje  $t$  žymi atitinkamą laiko momentą, o  $S_{t,1}$  žymi atitinkamos kompanijos akcijos kainą laiko momentu  $t$ . Svarbu pastebėti, kad sudarytą laiko eilutę Hudson'as [31] išskaidė dalimis pagal aktualius ekonominius ir politinius veiksnius į keturis laikotarpius:

- Laikotarpį nuo 1935 m. iki 1951 m. siekiant atspindėti karo metus bei galimą jų įtaką;
- Laikotarpį nuo 1951 m. iki 1966 m. atsižvelgiant į po karo prasidėjusį ekonominio bumo tarpsnį;
- Laikotarpį nuo 1966 m. iki 1981 m. atspindėti politinį bei ekonominį neapibrėžtumą Jungtinėje Karalystėje;
- Laikotarpį nuo 1981 m. iki 1994 m. siekiant įvertinti M. Thather'ės politinio valdymo įtaką.

Analizei Hudson'as [31] taikė dvi techninės prekybos taisykles (angl. technical trading rules): kintamojo slenkančio vidurkio (angl. variable moving average) ir fiksuoto slenkančio vidurkio (angl. fixed moving average). Pirmoji taisyklė teigia, jog akcijos pirkimas (pardavimas) turi būti atliekamas tuo momentu, kai trumpojo laikotarpio slenkančio vidurkio kreivė yra aukščiau (žemiau) ilgojo laikotarpio slenkančio vidurkio kreivės. Remiantis antra taisykle, akciją reikia pirkti (parduoti), kai trumpojo laikotarpio slenkančio vidurkio kreivė kerta ilgojo laikotarpio slenkančio vidurkio kreivę iš apačios.

Remiantis gautais rezultatais, Hudson'as [31] teigia, jog atvejai, kai grąža, gaunama taikant taisykles, labai skyrėsi nuo „pirk ir laikyk“ grąžos. Taisyklių kuriama grąža daugeliu atvejų buvo neigiama, todėl abejojama ar taisyklių taikymas gali būti laikomas pelninga investavimo strategija. Panaši techninės prekybos taisyklių taikymo problematika buvo pastebėta Jensen'o [32] tyrime. Šiame tyrime taip pat buvo lyginamas taisyklių taikymo efektyvumas su anksčiau minėta „pirk ir laikyk“ strategija. Tačiau Jensen'as [32] naudojo kiek kitokią techninę taisyklę. Šio tyrimo atveju buvo apskaičiuojama kiek reikia investuoti į atitinkamą akciją bei kada ją reikia parduoti. Pirmiausia buvo apskaičiuotas kainos santykis (angl. ratio) pagal formulę

$$PR_{jt} = P_{jt} / \bar{P}_{jt}, \quad (2)$$

kai  $\bar{P}_{jt}$  yra  $j$ -tosios akcijos vidutinė kaina tiriamuoju 27 savaičių laikotarpiu iki  $t$  momento jį įskaitant. Taip pat nustatomas kiekis  $X$  procentais, kurį reikia investuoti į  $X$  proc. akcijų, kurių  $t$  laiko momentu kainos santykis  $PR_{jt}$  yra didžiausias. Tiriamu periodu kainos santykis yra perskaičiuojamas visoms akcijoms kiekvieną savaitę ir akcijos, kurių „išmetimo reitingas“ (angl. „cast out rank“) viršija  $K^1$  reikšmę, akcija parduodama ir visos gautos lėšos investuojamos iš naujo į akciją, kurios kainos santykis tuo metu yra didžiausias. Analizei buvo pasirinktos visos Niujorko akcijų biržoje esančių akcijų kainos, dividendai ir komisiniai tiriamuoju laikotarpiu nuo 1926 m.

---

<sup>1</sup> Tyrime šio kintamojo skaičiavimo metodika nepateikiama.

sausio iki 1966 m. kovo mėn. Apžvelgdamas savo tyrimo rezultatus, Jensen'as [32] pabrėžia, jog neįvertinus komisinių mokesčių, taisyklių taikymu paremto investavimo grąža 1,4 proc. didesnė nei „pirk ir laikyk“ strategijos taikymo atveju. Tačiau įvertinus komisinius, taisyklių kuriama grąža yra lygi ar mažesnė nei „pirk ir laikyk“ atveju.

Chong'o [33] atlikta analizė tik iš dalies patvirtina techninės analizės taisyklių efektyvumo stoką. Kaip ir Hudson'as [31], Chong'as [33] tyrė FT-30 indeksą taikant tokį patį tyrimo laikotarpį. Tyrimui taikytos dvi techninės prekybos taisyklės: santykinio stiprumo indeksas (RSI) ir slenkančio vidurkio konvergencija - divergencija (MACD). RSI apskaičiavimui taikyta formulė

$$RSI_t(n) = \frac{\sum_{i=0}^{n-1} (P_{t-1} - P_{t-i-1}) 1_{\{P_{t-i} > P_{t-i-1}\}}}{\sum_{i=0}^{n-1} |P_{t-i} - P_{t-i-1}|} \times 100, \quad (3)$$

kai  $RSI_t$  yra santykinio stiprumo indeksas  $t$  momentu, o  $P_t$  yra indekso reikšmė  $t$  momentu,  $n$  žymi RSI periodų skaičių,  $1_{\{\cdot\}}$  yra rodiklinė funkcija, kuri lygi vienetui, kai skliausteliuose esanti nelygybė yra teisinga, ir lygi nuliui jei nelygybė neteisinga. MACD skaičiavimas yra paremtas skirtumu tarp ilgo ir trumpo eksponentinių slenkančių vidurkių. Chong'as MACD skaičiavimui taikė formulę

$$EMA_t = \left[ \frac{2}{n} \times (P_t - EMA_{t-1}) \right] + EMA_{t-1}, \quad (4)$$

kai  $EMA_t$  yra eksponentinis slenkantis vidurkis momentu  $t$ ,  $n$  yra EMA periodų skaičius. Pradine EMA reikšme tyrimo autoriai pasirinko 12 dienų periodą trumpajam EMA ir 26 dienų periodą ilgajam, kadangi, kaip teigiama straipsnyje, tai yra dažniausiai naudojamos periodų reikšmės EMA skaičiavimams. MACD taisyklės atveju pirkimo signalas iššaukiamas, kai MACD kreivė kerta nulinę ašį iš apačios, o pardavimo signalas, kai kreivė kerta nulinę ašį iš viršaus. Tyrimu buvo nustatyta, jog taikant RSI ir MACD taisykles gauta grąža viršijo „pirk ir laikyk“ strategijos grąžą, todėl tyrimu teigiama, jog šių taisyklių taikymas yra sėkmingas ir atsparus tyrimo periodo trukmei, kadangi lyginant šių strategijų rezultatus atskirais periodais su visu tyrimui pasirinktu laikotarpiu, gauti rezultatai yra labai stabilūs.

Taip pat esama ir kitų autorių tyrimų, kuriuose šios metodikos taisyklių taikymas turėjo pasisekimą. Brock'as [34], Mills'as [35] kaip ir Chong'as [33] techninę analizę taikė siekdami prognozuoti ne atskirų akcijų tendencijas, tačiau tyrimui naudojo indeksus – Brock'o [34] tyrime analizuotas Dow Jones pramonės indeksas, Mills'as [35] tyrė FT30 indeksą. Pastarieji tyrėjai taip pat taikė pagrindines, jau anksčiau aptartas, prekybos taisykles nustatant pirkimo ir pardavimo signalus. Šių tyrimų gauti rezultatai patvirtino, jog techninės analizės taikymas buvo efektyvesnis nei „pirk ir laikyk“ strategija – tyrimuose taikant taisykles investicinė grąža ženkliai viršija grąžą, gautą taikant strategiją.

Vis dėl to, aktualu pastebėti, jog techninės analizės sėkmingo taikymo atvejų literatūroje aptinkama mažiau, nei teigiančių, kad taisyklių taikymas nėra efektyvus. Taip pat svarbu tai, jog

sėkmingo taikymo atvejai dažnai kritikuojami teigiant, jog tuo metu gauta didesnė grąža taikant taisyklės galėjo būti visiškas atsitiktinumas, o tai tik patvirtintų faktą apie techninės analizės neefektyvumą. Taip pat pastebėta, jog labai didelė dalis investuotojų pasikliauna akcijų biržos tendencijomis, kurios yra sudarytos taikant techninės analizės metodus. Dėl šios priežasties, kaip teigia Frankel'is [36] tokiais atvejais, kai rinkoje pasitaiko nebūdingų akcijų kainų pokyčių, formuojant kainos kitimo tendenciją tokie atsitiktinumas nėra eliminuojami. Tai reiškia, jog atsitiktinis akcijos kainos pokytis yra vertinamas kaip įprastas dėsningumas akcijų biržoje ir tai sąlygoja spekuliacinių burbulų susidarymą.

Shiller'is [37], vertindamas techninės analizės taikymo naudingumą, šį klausimą tiria kiek kitaip. Autoriaus tyrimai orientuoti į investavimą, kai rinkoje susidaro spekuliaciniai burbulai. Tačiau vienas esminių autoriaus pastebėjimų yra tai, jog akcijų biržoje labai svarbus elementas yra investuotojų psichologija. Tai stipriai pasireiškia prieš finansines recesijas ir jų metu, nes vystantis finansiniam burbului, dauguma investuotojų bus linkę pervertinti rinkoje esančias akcijas ir taip padidins būsimos recesijos mastą. Taip pat aktualus ir investuotojų pesimizmas, kadangi nesitikint finansinio pagerėjimo netolimoje ateityje, recesija gali užsitęsti. Autoriaus teigimu, padidėjęs akcijų biržos kintamumas (angl. stock market volatility) buvo neracionalaus investuotojų elgesio pasekmė. Shiller'is [37] yra linkęs manyti, jog 1987 m. techninė analizė ir esami jos trūkumai (galimybė formuoti spekuliaciniams burbulams) galėjo lemti akcijų biržos krizę.

Taigi, apibendrinant verta pastebėti, jog daugelis tyrėjų skeptiškai vertina techninių taisyklių taikymo efektyvumą, kadangi atliktų tyrimų rezultatai įrodo, jog renkantis tarp dviejų alternatyvų – techninės analizės taisyklių ir paprastos „pirk ir laikyk“ strategijos taikymo, dažniausiai investuotojo tikėtina grąža būtų didesnė, jei akcijos, jas įsigijus, būtų tiesiog laikomos. Vis dėl to, svarbu nepamiršti, jog techninės taisyklės stengiasi aptikti tiesinę akcijų kainos priklausomybę. Kaip jau buvo minėta anksčiau, akcijų kainas gali lemti daug įvairių veiksnių, kurių poveikis gali būti netiesioginis, todėl tiesinės priklausomybės nustatymas šiuo atveju gali nebūti sėkmingas. Dėl šios priežasties verta apžvelgti kitas akcijų kainos prognozavimo metodikas, siekiant nustatyti metodiką, geriausiai tinkančią akcijų kainos analizei ir prognozavimui.

## **2.2. Laiko eilučių analizė**

Laiko eilučių prognozavimas yra viena plačiausiai taikomų metodikų prognozuojant akcijų kainą. Laiko eilute, pagal Stabingienės [38] apibrėžimą yra laikoma „analizuojamo atsitiktinio dydžio  $\zeta_t$  stebėjimų, gautų laiko momentais  $t=1, \dots, T$  eilutė  $Z_1, Z_2, \dots, Z_T$ “. Lyginant laiko eilučių prognozavimo metodiką ir anksčiau aptartą techninę analizę, svarbu paminėti, jog abi metodikas sieja tai, jog akcijų kainos analizei naudojamų duomenų struktūra yra tokia pati – finansinės laiko

eilutės. Techninė analizė siekiama tik įžvelgti kainų kitimo tendenciją, tačiau akcijos kaina nėra prognozuojama, priešingai nei taikant laiko eilučių prognozavimą.

Remiantis Tay [39] finansinėse laiko eilutėse pastebima nemažai triukšmo (angl. noise), t. y. jos yra nestacionarios bei chaotiškos. Šie požymiai finansinėms laiko eilutėms būdingi, kadangi jos yra veikiamos daugybės faktorių: ekonominių, politinių ir kt. Dėl šios priežasties, atliekant akcijų kainos analizę bei prognozę, taikant laiko eilučių prognozavimo metodiką, į akcijos kainos prognozę naudinga įtraukti ir papildomų išorinių veiksnių, norint nustatyti ar išorėje esama makroekonominių ir mikroekonominių veiksnių, galinčių daryti įtaką akcijos kainai.

Laiko eilučių prognozavimui gali būti taikomi vienmačiai ir daugiamačiai modeliai. Šiuo atveju modelio pasirinkimą lemia išorinių veiksnių naudojimas atliekant finansinės laiko eilutės prognozę. Jei yra įtraukiami papildomi kintamieji, pavyzdžiui kitų akcijų laiko eilutės (siekiant įvertinti skirtingų akcijų kainų tarpusavio sąveiką), mikroekonominių ar makroekonominių rodiklių laiko eilutės, tokiais atvejais yra naudojami daugiamačiai modeliai. Rusu [40] teigimu, laiko eilučių modeliai geriausiai tinka prognozavimui, kai susiduriama su tokiomis situacijomis:

- Nėra žinoma pakankamai informacijos apie veiksnį, galinčius daryti įtaką laiko eilutei, kurią norima prognozuoti;
- Didelis duomenų kiekis;
- Siekiama atlikti trumpo laikotarpio prognozę.

Atsižvelgiant į Rusu [40] išvardintas galimas problemas, informacijos apie išorinius veiksnį trūkumas pastebimas taikant daugiamačius modelius, kadangi vienmačiai modeliai neįtraukia papildomų veiksnių. Vertinant didelio duomenų kiekio situaciją, verta pastebėti, jog ji gali būti dvejopa: vienmačio ir daugiamačio modelių atveju gali būti naudojama labai ilga finansinių duomenų laiko eilutė arba daugiamačio modelio atveju norima į modelį įtraukti daug papildomus veiksnį aprašančių finansinių laiko eilučių. Esant aukščiau aprašytoms sąlygoms, finansinių laiko eilučių modelių taikymas turi stiprų pranašumą lyginant su technine akcijų kainos tendencijos analize.

Remiantis Hyndman'u [41], analizuojant laiko eilutes galima pastebėti joms būdingus šablonus (angl. pattern). Išskiriamos trys šablonų kategorijos: tendencija (angl. trend), sezoniškumas ir cikliškumas. Tendencijos šablonas tiriant akcijų kainos kitimą, kaip ir techninės analizės atveju, yra siejamas su ilgojo laikotarpio kainų kilimu ar kritimu. Sezoniškumą galima pastebėti tik jei tiriama laiko eilutė priklauso nuo sezoninių faktorių, pavyzdžiui metų ar paros laiko. Cikliškumo šabloną tiriant akcijų kainą galima pastebėti jei esama ryškaus kainos kilimo ar kritimo, kuris neturi apibrėžto periodiškumo. Tokie kainos pokyčiai gali būti sąlygojami ekonominės situacijos ar verslo ciklų. Vis dėl to, Rusu [40] būdingiems laiko eilučių šablonams taip pat priskiria ir atsitiktinio kitimo (angl. random change) šabloną. Taigi, prognozuojant

finansines laiko eilutes šablonų nustatymas duomenyse leidžia įvertinti prognozuojamo veiksnio ypatybes, pavyzdžiui jei pastebimas sezoniškumas, nustačius jo dažnį, galima rasti veiksnus, kurie sąlygoja jo atsiradimą.

Prognozuojant finansines laiko eilutes, galima išskirti keletą dažniausiai taikomų metodų [42]: eksponentinis suglodinimas, autoregresiniai integruotų slenkančių vidurkių modeliai (ARIMA), apibendrinti autoregresiniai sąlyginio heteroskedastiškumo modeliai (GARCH). Aktualu atsižvelgti į tai, jog GARCH modelis yra autoregresinio sąlyginio heteroskedastiškumo (ARCH) modelio apibendrinimas. Svarbu pastebėti, jog išvardinti metodai yra tinkami nustatyti tiesinei laiko eilutės priklausomybei, todėl nėra tinkami prognozuoti netiesines priklausomybes. Tačiau literatūroje esama atvejų, kai tiesinę priklausomybę prognozuojantys metodai yra taikomi kartu su netiesiniais ([43], [44]). Dažniausiai netiesinėms priklausomybėms nustatyti yra naudojami dirbtinių neuroninių tinklų (angl. artificial neural network) modeliai, taip pat SVM (angl. support vector machine) modeliai. Šie modeliai yra priskiriami dirbtinio intelekto (angl. artificial intelligence) metodams, kurie taikomi duomenų tyryboje. Lu [45] teigimu SVM metodo rizikos mažinimo struktūra leidžia geriau apibendrinti duomenis nei tradicinių neuroninių tinklų atveju.

Remiantis atliktais tyrimais, tiesinių ir netiesinių metodų derinimas tarpusavyje yra sėkmingesnis nei kiekvieno iš metodų taikymas atskirai. Pai [43] atliktame tyrime buvo derinamas ARIMA modelis prognozuoti tiesinę priklausomybę ir SVM modelis netiesinės priklausomybės analizei. Standartinė ARIMA modelio išraiška, kuri taip pat buvo taikoma ir Pai [43] tyrime, pateikiama formule

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (5)$$

kai  $y_t$  yra tikroji reikšmė,  $\varepsilon_t$  yra atsitiktinė paklaida momentu  $t$ ,  $\phi_i$  ir  $\theta_j$  yra koeficientai, o  $p$  ir  $q$  yra atitinkamai autoregresiją ir slenkantį vidurkį atspindinčios daugianario dalys. SVM modelio regresijos funkcija [43] aprašoma formule

$$y = w\phi(x) + b, \quad (6)$$

kai  $\phi(x)$  yra laikoma ypatybe (angl. feature), kuri yra susiejama netiesiškai iš įvesties erdvės (angl. input space)  $x$ . Koeficientų  $w$  ir  $b$  reikšmės yra nustatomos minimizuojant funkcijas

$$R(C) = C \frac{1}{N} \sum_{i=1}^N L_\varepsilon(d_i, y_i) + \frac{1}{2} \|w\|^2, \quad (7)$$

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon & |d - y| \geq \varepsilon \\ 0 & \text{kiti} \end{cases}, \quad (8)$$

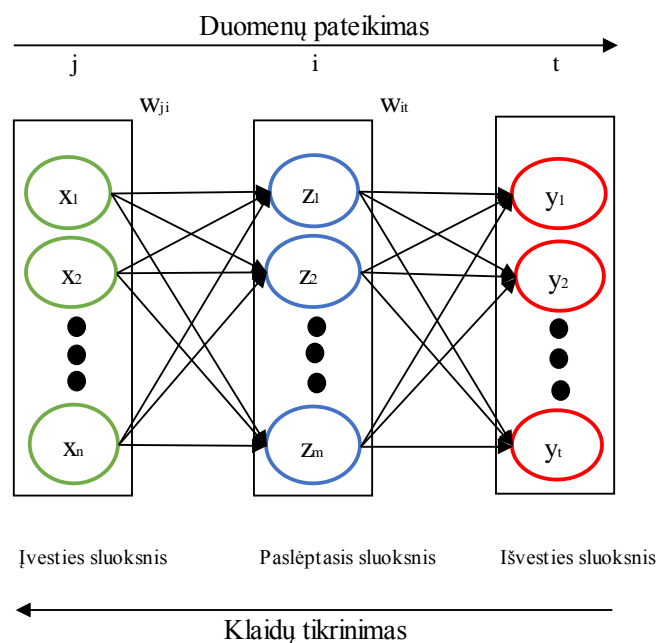
kai  $C$  ir  $\varepsilon$  yra numatyti parametrai,  $d_i$  žymi tikrąją akcijos kainą  $i$ -tuoju periodu. Formulės (7) pirmoji dalis iki sudėties ženklo žymi empirinę klaidą, o antroji funkcijos dalis skirta matuoti funkcijos plokštumą (angl. flatness). Tyrime taikyto hibridinio modelio struktūra rėmėsi tuo, jog pirmiausia buvo nustatoma tiesinė priklausomybė ARIMA modeliu ir gautos paklaidos buvo

naudojamos SVM modeliui, siekiant SVM modeliu sumažinti paklaidų funkciją, sudarytą taikant ARIMA modelį. Tik atlikus šį žingsnį, buvo koreguojami SVM modelio parametrai ( $\epsilon$ ,  $C$ ,  $\sigma$ ). Tyrimo metu gauti rezultatai patvirtino, jog hibridinio modelio prognozė (sudaryto ARIMA ir SVM modelių pagrindu) yra tikslesnė nei atskira ARIMA ar SVM modelio prognozė. Tačiau pastebima, jog hibridinio modelio prognozė buvo mažiau tiksli nei tikėtasi ir tam įtakos turėjo atskirų modelio dalių parametrų parinkimas.

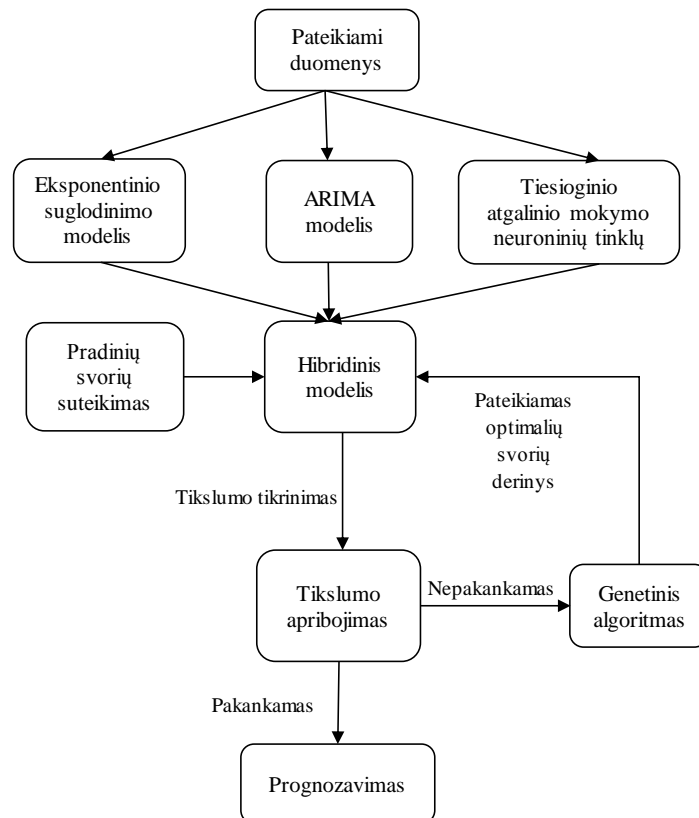
Wang [44] tyrimo pagrindinis uždavinys atliekant kainų prognozavimą buvo suderinti eksponentinio suglodinimo, ARIMA bei tiesioginiai atgalinio mokymo neuroniniai tinklai (angl. back propagation neural network). Eksponentinio suglodinimo prognozei numatyti taikomas matematinis išvedimas

$$F_{t+1} = \alpha x_t + (1-\alpha)F_t = F_t + \alpha (x_t - F_t) = F_t + \alpha e_t, \quad (9)$$

kai  $x_t$  ir  $F_t$  atitinkamai žymi tikrąją ir prognozuojamą reikšmę momentu  $t$ ;  $\alpha$  žymi eksponentinio suglodinimo parametą, esantį tarp 0 ir 1, kuriuo koreguojamas svorio pokyčio įvertis (angl. weight change rate);  $e_t$  žymima paklaida  $t$  momentu. ARIMA modelio išraiška pateikta (5) formule. Wang [44] tyrime taikytų neuroninių tinklų schema pateikiama paveiksle (2.2.1 pav.), tyrimui taikyto hibridinio modelio loginė schema taip pat pateikiama paveiksle (2.2.2 pav.).



**2.2.1 pav.** Trijų sluoksnių tiesioginis atgalinio mokymo neuroninis tinklas



**2.2.2 pav.** Wang hibridinio modelio loginė schema

Kaip ir Pai [43] atveju, Wang [44] nustatė, jog šio tyrimo hibridinis modelis turi mažiausią absoliutinę paklaidą lyginant su atskirų, anksčiau išvardintų, modelių rezultatais. Taip pat vertinant modelio atsparumą naujiems duomenims (angl. robustness), hibridinis modelis buvo atspariausias iš kitų, tyrime taiktų modelių.

Taigi, galima pastebėti, jog akcijų kainos laiko eilučių prognozavimą galima atskirti į dvi kategorijas – tiesinę priklausomybę vertinantys metodai (eksponentinis suglodinimas, ARIMA, GARCH ir kt.) bei netiesinę priklausomybę tiriantys algoritmai (SVM, ANN, genetiniai algoritmai ir kt.). Vis dėl to, naujausiais tyrimais pastebima, jog šių atskirų modelių derinimas tarpusavyje sudarant hibridinį modelį leidžia gauti geresnius rezultatus nei taikant modelius atskirai, kadangi finansinėms laiko eilutėms būdingą tiesinę ir netiesinę priklausomybę hibridiniai modeliai nustato tiksliau, kadangi atskiri modeliai vertina kiekvieną iš šių priklausomybių atskirai.

### **2.3. Sistemų mokymosi ir duomenų tyrybos analizė**

Kaip jau buvo minėta ankstesniuose skyriuose, informacijos kiekis, prieinamas didžiajai daugumai asmenų, auga eksponentiškai, todėl darosi vis sunkiau be papildomų įrankių išanalizuoti visą prieinamą informaciją ir atskirti aktualią informaciją nuo nereikšmingos. Duomenų tyryba (angl. data mining) Han [46] teigimu, iš tiesų yra įžvalgos gavimas iš duomenų (angl. knowledge



mining from data). Witten [47] nuomone, duomenų tyryba yra skirta atrasti pasikartojančius šablonus (angl. pattern) duomenyse, kurių svarba bei išskirtinumas suteiktų galimybę gauti ekonominę naudą jais pasinaudojus prognozuojant ateities tendencijas. Šiems požiūriams neprieštarauja Kantardzic [48] savo tvirtinimu, jog duomenų tyryba yra būdas išgauti naują, vertingą informaciją iš didelio duomenų kiekio. Aktualu tai, jog visi autoriai pabrėžia duomenų tyrybos metodų taikymo galimybes dideliems informacijos srautams. Tokie srautai dažnai įvardijami kaip didieji duomenys (angl. big data). Šių duomenų tyryba ir toliau augant informacijos srautams tampa neatsiejama įvairių veiksmų analizės dalimi.

Literatūroje esama daugybės pavyzdžių, jog duomenų tyryba gali būti taikoma patiems įvairiausiems informacijos šaltiniams: medicininiais ([49], [50]), banko ([51]), [52]), akcijų ([16], [18], [19]), agrokultūriniais ([53], [54]) duomenims bei tyrimams atlikti. Apibendrinant galima pastebėti, jog duomenų tyrybos metodai gali būti taikomi duomenims, turintiems struktūrą, tačiau taip pat ir duomenims, kurie neturi aiškios struktūros – yra nestruktūrizuoti (angl. unstructured). Vis dėl to, svarbu pabrėžti, jog duomenų tyryba yra duomenų analizės procesas, kurį sudaro keletas pagrindinių dalių ([46], [48]):

- Duomenų valymas – iš visos informacijos pašalinama bevertė informacija – triukšmas (angl. noise) – ir išskirtys;
- Duomenų dimensijos mažinimas - atrenkama analizei reikšminga informacija;
- Duomenų modeliavimas – šablonų paieška duomenyse, veiksmų tarpusavio sąveikos paieška;
- Gautų rezultatų interpretavimas – vertinami rezultatai, remiantis jais priimamos išvados.

Šiam analizės procesui įgyvendinti yra naudojami sistemų mokymuisi priskiriami metodai bei algoritmai. Sistemų mokymosi algoritams yra priskiriami ANN, SVM, genetiniai algoritmai ir daug kitų metodų. Vis dėl to svarbu pastebėti, jog sistemų mokymąsi, taip pat ir jam priklausančius algoritmus galima skaidyti į dvi pagrindines kategorijas: sistemų mokymąsi su mokytoju (angl. supervised learning) ir be mokytojo (angl. unsupervised learning). Remiantis Zhou [55], jei tiriamų duomenų rinkinyje įvesties elementai turi atsako kintamąjį, kuriam priskiriama žymė (angl. label) su teisinga išvestimi, tada gali būti taikomas sistemos mokymasis su mokytoju. Priešingai, jei duomenų rinkinys neturi informacijos apie teisingą išvestį, atliekamas mokymasis be mokytojo.

Analizuojant akcijų kainos kitimo prognozavimo uždavinį tyrusius darbus, galima pastebėti, jog šiai analizei naudojami ankstesniuose skyriuose minėti kiekybiniai bei kokybiniai duomenys. Didelė dalis akcijų kainos kitimą prognozavusių tyrimų ([56], [57], [58], [59], [60]), savo analizei naudojo kiekybinius duomenis (akcijų, indeksų kainas ir kt.). Tokio pasirinkimo priežastimi

galima įvardinti tai, jog šie duomenys dažniausiai turi struktūrą, todėl jų tyryba bei šablonų įžvalga yra paprastesnė ir lengviau įgyvendinama nei duomenų, neturinčių aiškos struktūros. Vis dėl to esama akcijų kainos analizės tyrimų, analizavusių kokybinių duomenų poveikį akcijų kainoms ([16], [18], [61]). Aktualu pastebėti, jog tokių tyrimų pagrindinis analizės šaltinis yra tekstinė informacija. Tokio pobūdžio tyrimai įvardijami kaip atskira duomenų tyrybos kryptis – teksto tyryba (angl. text mining). Todėl šią akcijų kainos analizės bei prognozavimo kryptį verta apžvelgti plačiau bei įvertinti tekstinės informacijos taikymo analizei privalumus ir trūkumus.

### **3. TEKSTO TYRYBOS ANALIZĖ**

Kaip jau buvo minėta anksčiau, informacijos srautams augant eksponentiškai, informacijos apdorojimas bei vertingų žinių gavimas yra sudėtingas ir kompleksiškas procesas. Todėl ilgą laiką norint gauti aktualios informacijos iš laikraščių naujienų, straipsnių bei įvairių asmenų nuomonių, vienu ar kitu klausimu, reikėjo kiekvienam susidomėjusiam asmeniui tokią tekstinę informaciją perskaityti pačiam. Atsižvelgiant į tai, jog toks informacijos gavimas buvo labai imlus laikui, tekstinės informacijos analizė buvo vienas iš rečiausiai taikomų metodų siekiant įvertinti akcijos kainų kitimą bei remiantis tekstine informacija numatyti galimas akcijų kainų kitimo tendencijas.

Vis dėl to, atsiradus metodams bei algoritmams (pavyzdžiui SVM, neuroniniams tinklams, LDA), galintiems apdoroti tekstinę informaciją, ši analizės kryptis yra vis populiarsnė. Literatūroje [62], [63], [64], [65] teksto tyrybos procesas dažniausiai skaidomas į keletą pagrindinių dalių, kurios turi būti įgyvendintos sėkmingai teksto analizei. Pirmiausiai turi būti atrenkami informacijos šaltiniai, kuriuose esanti tekstinė informacija gali būti aktuali priklausomam kintamajam, kurio kaitos tendenciją norima įvertinti.

Turint įvairios tekstinės informacijos, aktualu išrinkti ypatybes (angl. features), kurios geriausiai atspindi turimą tekstinę informaciją kiekvienam šaltiniui (dokumentui, straipsniui ir kt.). Šios ypatybės gali būti žodžiai ar žodžių junginiai, kuriais galima aiškiai nusakyti teksto turinį. Atliekant teksto analizę sentimentų nustatymui, taip pat aktualu įvertinti sentimentus (teigiamus ar neigiamus) atspindinčius žodžius. Tolimesniame etape yra atrenkamos pačios reprezentatyviausios ypatybės – siekiant sumažinti dimensijų skaičių atrenkamos populiariausios ypatybės ir vertinamas jų buvimas kiekviename informacijos šaltinyje.

Paskutiniame žingsnyje atliekama tekstinės informacijos ir priklausomo kintamojo sąveikos analizė. Svarbu pastebėti, jog ši analizė anksčiau atliktuose tyrimuose varijuoja priklausomai nuo pasirinkto taikyti algoritmo (neuroninių tinklų, SVM, LDA ir kt.). Kiekvienas iš teksto tyrybos etapų yra aptariamas plačiau tolesniuose poskyriuose.

### 3.1. Tekstinės informacijos apdorojimas

Kaip jau minėta skyriaus pradžioje, pirmiausia svarbu pasirinkti tinkamus šaltinius teksto analizei. Anksčiau atliktuose tyrimuose pastebimi įvairūs tekstinės informacijos šaltiniai: Sun [18], Oliveira [19] savo analizei naudojo socialinio tinklo Twitter informacinius pranešimus susijusius su tiriamomis akcijomis, Geva [61], Li [66] tyrimams buvo naudojamos naujienos iš įvairių žinių internetinių puslapių (Reuters, PR Newswire ir kt.). Remiantis atliktais tyrimais, galima išskirti keletą svarbių teksto apdorojimo žingsnių siekiant efektyviai atlikti teksto analizę.

Losiewicz [67] pastebi, jog žodžiai anglų kalboje yra atskiriami tuščiomis erdvėmis (angl. white spaces), kuriomis gali būti įprasti maži tarpai tarp žodžių, tuščios eilutės ir kt. Todėl svarbu paminėti, jog prieš pradedant teksto analizę, pirmiausia turi būti pašalinami įvairios apimties tarpai tarp žodžių, o likę teksto žodžiai, kiekvienas atskirai, interpretuojami kaip atskiri teksto vienetai – leksemos (angl. tokens). Svarbu paminėti, jog skyrybos ženklai taip pat turi būti pašalinami. Turint tekstą, išskaidytą į leksemas, aktualu atrinkti ypatybes, kurios geriausiai reprezentuotų kiekvieno tekstinio šaltinio informaciją.

Geva [61] pastebėjimu, tam dažniausiai naudojamas žodžių vektorius (angl. bag-of-words). Savo atliktame tyrime Geva [61] taip pat naudojo žodžių vektorių, tačiau prieš taikant šį metodą buvo atliktas dažnai pasikartojančių bereikšmių žodžių valymas iš teksto (angl. „stop-list“ filtering) bei giminingų žodžių nustatymas pagal žodžio kamieną (angl. stemming). Atlikus šiuos žingsnius, Geva [61] sudarė žodžių vektorių. Žodžių vektoriaus pagrindinė esmė yra nustatyti žodžių pasikartojimo tekste dažnumą. Kaip pastebi Losiewicz [67], žodžių pasikartojimo dažnio tekste nustatymas yra naudingas nustatant teksto prasmę. Jis taip pat pastebi, jog žodžių dažnumą galima vertinti tiesiog pagal atskirus žodžius, pagal tos pačios šaknies ar to paties kamieno žodžių pasikartojimo dažnį, n-gramas ir t. t.

Hagenau [65] tyrime buvo taip pat analizuojamos ir kitos dažniausių žodžių nustatymo metodikos. Tyrimo [65] ypatybės (angl. features) taip pat buvo nustatomos pagal n-gramas, daiktavardžių frazes bei dviejų žodžių kombinacijas (angl. 2-word combination). N-gramomis yra laikomas žodžių junginys, sudarytas iš n žodžių, pavyzdžiui, žodžių junginys, sudarytas iš dviejų žodžių yra vadinamas bigrama. Šiuo atveju tiriant žodžių junginį yra svarbu koks yra šalia esantis žodis bei žodžių tvarka tekste. Daiktavardžių frazėmis yra laikomos (dažniausiai dviejų žodžių) frazės, kurių pagrindinis žodis yra daiktavardis ar įvardis, kuris sakinyje yra šalia būdvardžio ar kitos kalbos dalies. Dviejų žodžių kombinacijos yra panašios į anksčiau aptartas žodžių tyrybos metodikas, tačiau šiuo atveju yra leidžiamas didesnis atstumas tarp kombinaciją sudarančių žodžių, t. y. tarp dviejų žodžių gali būti įsiterpę ir kitų žodžių, kurie nėra laikomi kombinacijos dalimi.

Kiekvieno tyrimo atveju pasirinkus tinkamą leksemas reprezentuojantį metodą: žodžių vektorių, n-gramas ar žodžių kombinacijas, aktualu įvertinti žodžių ar kombinacijos dažnumą tekste. Remiantis Nassirtoussi [64] galima išskirti keletą populiariausių ypatybių atspindėjimo metodų: informacijos padidėjimo (angl. information gain), Chi-kvadrato statistikos (angl. Chi-squared statistics), požymių atrinkimo dažnumo (angl. document frequency), tikslumo subalansavimo (angl. accuracy balanced) ir sąvokos atvirkštinio dažnio požymių atrinkimo (angl. term frequency-inverse document frequency). Vis dėl to, vienas paprasčiausių ypatybių atspindėjimo metodų yra dvinarė reprezentacija, kuria 0 ir 1 vertinamas ypatybės buvimas tiriamame tekste priskiriant 0, jei ypatybės nėra, ir 1, jei ypatybė yra tiriamame tekste.

Taigi apibendrinant galima pastebėti, jog pirmiausia reikalinga pasirinkti tinkamus informacijos šaltinius, atlikti nereikšmingos informacijos (bereikšmių žodžių, tarpų, skyrybos ženklų ir t. t.) pašalinimą, pasirinkti tinkamą ypatybių reprezentavimo metodą bei nustatyti ypatybių pasikartojimo tekste dažnumą. Šie žingsniai turi būti atlikti prieš pradedant tekstinės informacijos poveikio analizę.

### **3.2. Sentimentų analizė**

Atliekant sentimentų analizę, svarbu pastebėti, jog sentimentus tekste galima vertinti įvairiai, tačiau pastebima, jog sentimentai gali būti teigiami arba neigiami. Kaip ir įprastos teksto tyrybos atveju, taip ir siekiant nustatyti teksto sentimentus, pirmiausia reikalinga atlikti ankstesniame poskyryje aptartus žingsnius. Vienas iš galimų sentimentų analizės būdų yra sentimentų žodyno [68] taikymas tekstinės informacijos analizei. Tokiu atveju tiriamos tekstinės informacijos žodžiams suteikiama teigiama ar neigiama emocinė reikšmė, kurios analizė leidžia suprasti tiriamo teksto prasmę: ar tiriamą informaciją pasižymi pozityviomis žiniomis ar negatyviomis.

Tačiau tiriant sentimentus taip pat galima taikyti apklausų indeksus, kuriais pažymima teigiama ar neigiama informacija, kaip tai buvo atlikta Oliveiros [19] tyrime. Šio tyrimo metu buvo taikomi šie apklausų indeksai: Amerikos individualių investuotojų asociacijos indeksas (AAII), investuotojų įžvalgos indeksas (II), namų ūkio pasitikėjimo indeksas (USMC), įvairių akcijų biržų sentimentų indeksai (Sentix). AAII indeksas yra laikomas sentimentų įverčiu (angl. sentiment measure), kadangi yra sudaromas remiantis savaitiniu šios asociacijos narių akcijų biržos vertinimu – investuotojams pateikiama apklausa, kurioje jie turi įvertinti savo sentimentus (kuriai kategorijai jie priklauso: bulių, lokių ar neutraliai) akcijų biržos atžvilgiu šešių mėnesių periodui. II indeksas yra sudaromas remiantis 100 nepriklausomų naujienlaiškių, kurie pagal pateikiamas naujienas gali būti priskiriami buliams ar lokiams arba informacija gali būti laikoma kaip ankstesnių naujienų korekcija. USMC indeksas sudaromas apklausiant 500 atsitiktinai

parenkamų namų ūkių, o Sentix indeksas formuojamas pagal apklausas, kurias užpildo apie 3500 dalyvių.

Nguyen'as [69] atlikdamas sentimentų analizę, tyrė socialinių tinklų tekstinę informaciją. Tačiau šiuo atveju, kiekvienai informacinei žinutei teksto autorius galėjo priskirti vieną iš kategorijų, žyminčių atitinkamą sentimentą: būtina pirkti (angl. strong buy), pirkti, laikyti (angl. hold), parduoti, būtina parduoti (angl. strong sell). Nguyen'as [69] klasifikavo turimą tekstinę informaciją į šias sentimentų kategorijas ir tokiu būdu galėjo nustatyti investuotojų požiūrį į tekste išdėstomas problemas. Tačiau svarbu pastebėti, jog toks sentimentų skaidymas pagal atsako kintamąjį, t. y. iš anksto žinomą teksto žymę (angl. label), dažnai nėra įmanomas, kadangi didelė dalis tekstinės informacijos neturi iš anksto priskirtos žymės, kuria remiantis galima nusakyti tekste vyraujančius sentimentus.

Taigi apibendrinant galima pastebėti, jog teksto sentimentų analizė padeda suprasti teksto prasmę bei numatyti esamą (žinant išankstines teksto žymes) ar numanomą investuotojų požiūrį į esamą situaciją biržoje, akcijų kainų pokyčius ir kt.

### **3.3. Tekstinės informacijos poveikio nustatymas**

Tekstinė analizė gali būti atliekama įvairiai tekstei informacijai, tačiau kaip pastebėta ankstesniame poskyryje, tekstinė informacija gali turėti iš anksto priskirtą žymę (angl. label), kuria gali būti nusakomas tekstinės informacijos poveikis priklausomam kintamajam – akcijų kainai (pavyzdžiui, tiriant tekstinę informaciją, žinoma, jog ji paveikė akcijos kainos padidėjimą (sumažėjimą) ir tokiam tekstui priskiriama kainos augimą žyminti žymė) arba teksto sentimentų reikšmė (pavyzdžiui, iš anksto žinoma, jog tiriamas tekstas yra pozityvus ar negatyvus ir tekstui suteikiama atitinkama žymė). Iš anksto žinant tekstui suteiktas žymes, informacijos poveikis gali būti vertinamas taikant klasifikavimo metodus. Tokio pobūdžio tyrimus atliko Sun [18], Oliveira [19].

Tačiau tais atvejais, kai iš anksto negalima nustatyti teksto žymės, pagal kurią tekstinę informaciją būtų galima klasifikuoti, teksto tyrybai yra naudojami temų modeliai (angl. topic models). Tokią analizę savo tyrimuose taikė Geva [61], Li [66]. Tačiau svarbu pastebėti, jog šių tyrimų atveju pirmiausia buvo taikomas temų modeliavimas, t. y. atrinkus reprezentatyviausias ypatybes yra sudaromos temos, kurioms šios ypatybės priskiriamos, atitinkamai įvertinami šių žodžių sentimentai, tačiau vėliau visais atvejais stengiamasi atlikti automatinį ar iš dalies automatizuotą žymių (angl. label) nustatymą, jei norima prognozuoti akcijų kainos kitimo tendenciją remiantis tekste informacija.

Taigi remiantis atlikta literatūros analize, svarbu pastebėti, jog akcijų kainų kitimo tendencijų numatymas bei kainos prognozavimas yra aktualus uždavinys jau ilgą laiką. Priklausomai nuo

turimų informacijos išteklių: skaitinės mikroekonominių ir makroekonominių rodiklių informacijos, naujienų internete bei socialinių tinklų informacijos, buvo taikomi skirtingi metodai akcijos kainos analizei. Aktualu tai, jog tekstinės bei socialinių tinklų informacijos įtaka lyginant su statistine ekonominių rodiklių analize yra mažiau analizuota, taip pat jos poveikis akcijų kainai yra mažiau tirtas nei statistinių rodiklių. Dėl šios priežasties informacinių naujienų poveikio analizė akcijų kainoms yra aktuali tyrimo kryptis.

Kaip jau buvo minėta anksčiau, investavimas į akcijas yra laikomas rizikingu, kadangi akcijas gali veikti daugybė išorinių veiksnių. Vis dėl to, renkantis įmonę, kurios akcijų norima įsigyti, svarbu ne tik įmonės finansiniai rodikliai ir padėtis rinkoje, tačiau ir sektoriaus, kuriame veikia įmonė, plėtros potencialas. Pastaraisiais metais labiausiai savo augimu išsiskiria sveikatos priežiūros (angl. healthcare) sektorius. Remiantis Seth'u [70] sveikatos priežiūros sektoriaus įmonių akcijos gali būti laikomos viena iš saugiausių investicijų, kadangi istoriškai nauji atradimai bei vaistai sveikatos priežiūros srityje padėjo padaryti žmonių gyvenimo trukmę ilgesne.

Borzykowski's [71] tai pat pastebi, jog S&P 500 sveikatos priežiūros sektoriaus dalis (angl. S&P 500 health-care) nuo 2010 metų išaugo 162 procentais. Todėl jis taip pat teigia, jog ilgalaikis investavimas į sveikatos priežiūros sektoriaus įmones užtikrina saugią investicinę grąžą. Atsižvelgiant į Zucchi [72] įvardintas sveikatos priežiūros sektoriui palankias tendencijas, tokias kaip senstanti visuomenė, asmenų, sergančių chroniškais ligomis, ilgėjanti gyvenimo trukmė, technologinė pažanga, suasmeninta medicina ir kt., aktualu, jog sveikatos priežiūros įmonių paslaugų poreikis tik augs. Taip pat svarbu pastebėti, jog sveikatos priežiūros įmonių akcijų kainų augimui įtakos turėjo sparti biotechnologijų pažanga [71]. Todėl tolesniam tyrimui buvo pasirinkta analizuoti sveikatos priežiūros sektoriui priklausančių didžiausių farmacijos kompanijų akcijų kainas bei nustatyti informacinių naujienų poveikį pasirinktoms akcijoms.

# MEDŽIAGOS IR TYRIMŲ METODAI

## 4. PRADINIAI DUOMENYS IR JŲ APDOROJIMAS

Kaip jau buvo minėta anksčiau, šio tyrimo tikslas yra nustatyti informacinių pranešimų, pasirodančių interneto naujienų portaluose, poveikį akcijų kainoms. Prieš atliekant tyrimą yra svarbu nustatyti aktualų naujienų šaltinį bei pasirinkti aktualias akcijas, kurių kainų kitimo analizė bus atliekama. Duomenų pasirinkimas, duomenų gavimo procesas, pradinė duomenų struktūra bei duomenų paruošimas analizei pateikiami tolesniuose šio skyriaus poskyriuose.

### 4.1. Didžiųjų farmacijos kompanijų akcijų laiko eilutės

Keletą pastarųjų metų yra stebimas biotechnologijų šakų plėtros augimas, kuriam įtaką daro naujų technologijų įtraukimas kuriant vaistus bei tobulinant gydymo būdus ir technologiją. Šis augimas yra pastebimas pasauliniu mastu ir, kaip pastebi Anand'as [73], 2014 metais buvo akcijų (pavyzdžiui, Aurobindo Pharma), kurios investuotojų turtą padidino daugiau nei du kartus. Dėl šios priežasties aktualu analizuoti farmacijos kompanijų akcijų kainos kitimą remiantis šio sektoriaus pastarųjų metų augimu bei prognozuojamu potencialiu augimu.

Šiam tyrimui buvo pasirinktos 10 didžiųjų farmacijos kompanijų pasauliniu mastu akcijų kainos. Kompanijų dešimtukas buvo atrinktas remiantis pelnu bei įtaka rinkoje 2016 metais [74]. Tolesniam tyrimui buvo pasirinktos šios farmacijos kompanijos (sąrašas pateikiamas pelno bei rinkos įtakos mažėjimo tvarka):

1. Johnson & Johnson (pelnas 70 mlrd. dolerių);
2. Bayer (pelnas 51,4 mlrd. dolerių);
3. Novartis (pelnas 49,4 mlrd. dolerių);
4. Pfizer (pelnas 48,9 mlrd. dolerių);
5. Roche (pelnas 48,1 mlrd. dolerių);
6. Merck (pelnas 39,5 mlrd. dolerių);
7. Sanofi (pelnas 34,5 mlrd. dolerių);
8. Gilead (pelnas 32,6 mlrd. dolerių);
9. AstraZeneca (pelnas 24,7 mlrd. dolerių);
10. GlaxoSmithKline (pelnas 23,92 mlrd. dolerių).

Visos patektos kompanijos yra listinguojamos akcijų biržose (Nasdaq, NYSE ir kt.), todėl duomenys analizei buvo gauti taikant programos R paketą quantmod. Naudojantis šiuo paketu gautos visų kompanijų kiekvienos dienos uždarymo akcijos kainos. Tyrimui pasirinktas laikotarpis nuo 2014 m. sausio mėn. 1 d. iki 2016 m. gruodžio mėn. 31 d. Svarbu pastebėti, jog prekyba akcijomis vyksta tik darbo dienomis, todėl tariama, jog savaitę sudaro penkios dienos.

Atliekant literatūros analizę, buvo pastebėta, jog akcijų kainų kitimo analizei bei prognozei dažnai naudojamas ARIMA modelis. Šiam tyrimui taip pat pasirinktas ARIMA modelis, kadangi šiame modelyje galima įtraukti išorinius kintamuosius, siekiant nustatyti jų poveikį akcijos kainų laiko eilutei. Tačiau prieš atliekant šią analizę reikalinga nustatyti laiko eilutės būdingų bruožų (trendo, sezoniškumo ir kt.) egzistavimą ar nebuvimą. Laiko eilučių paruošimas bei būdingų bruožų nustatymas plačiau aptariami 5 skyriaus 5.1. poskyryje.

#### **4.2. Informacinių pranešimų atrinkimas ir gavyba**

Informacinių pranešimų gavimui pasirinktas Reuters naujienų portalas, kuris skelbia ir įvairias naujienas iš viso pasaulio, ir finansų rinkų duomenis. Analizuojant galimą naujienų poveikį akcijoms, aktualu vertinti ir įprastas naujienas, tačiau taip pat ir finansinius pranešimus, kurie gali turėti tiesioginę įtaką akcijų kainų pokyčiams. Vis dėl to, svarbu paminėti, jog Reuters portale taip pat pateikiamos įvairios sporto, pramoginės naujienos, kurios, jei analizuojama ne sporto ar pramogų sfera nėra aktualios tekstinės informacijos analizei. Dėl šios priežasties, atliekant tyrimą buvo siekiama iš karto sumažinti informacijos srautą atsisakant nesusijusių sričių naujienų.

Pirmiausia buvo išanalizuota 15 straipsnių, susijusių su sveikatos priežiūros sektoriumi, esančių Reuters naujienų portale. Ši analizė buvo reikalinga siekiant atrinkti raktinius žodžius, kurie dažniausiai minimi tokio pobūdžio naujienose, tam, kad būtų galima remiantis raktiniais žodžiais atrinkti straipsnius, kurių informacija gali turėti įtakos tiriamų kompanijų akcijų kainoms. Straipsnių, kuriais remiantis parinkti raktiniai žodžiai, sąrašas pateikiamas 1 priede. Remiantis analizuotais straipsniais, buvo atrinkti šie raktiniai žodžiai:

- Health care / healthcare (sveikatos priežiūra);
- Pharmaceuticals (farmacijos priemonės, vaistai);
- Food and Drug Administration (Maisto ir vaistų administracija);

Taip pat svarbu pastebėti, jog paieškai buvo parinktos įvairios galimos šių žodžių formos bei trumpiniai, pavyzdžiui, tekste apie farmacijos kompaniją dažnai pastebimas žodžių junginys „Pharmaceuticals company“, tačiau šis žodžių junginys kitame straipsnyje gali būti pakeistas vienu žodžiu – „pharma“, turinčiu tą pačią prasmę. Taip pat buvo atsižvelgta į galimus maisto ir vaistų administracijos trumpinius: FDA, F.D.A. Verta pastebėti, jog pirmųjų raktinių žodžių ryšys su didžiosiomis farmacijos kompanijomis aiškus: šios kompanijos priklauso sveikatos priežiūros (angl. health care) sektoriui ir yra farmacijos (angl. pharmaceuticals) kompanijos. Tačiau Maisto ir vaistų administracijos ryšys su farmacijos kompanijomis pastebimas todėl, jog ši agentūra yra atsakinga už įvairių vaistų patvirtinimą, jog jie saugūs ir gali būti platinami rinkoje. Dėl šios priežasties, kompanijoms sukūrus naujus vaistus, naujienos, jog maisto ir vaistų administracija



patvirtino masinį vaisto naudojimą ar jį atmetė yra labai aktualios, kadangi nauji vaistai farmacijos kompanijoms reiškia augančius pardavimo mastus bei pelną, o tai savo ruožtu veikia akcijų kainas.

Taip pat prie raktinių žodžių, atrenkant straipsnius informacinių naujienų analizei, buvo priskirti ir tiriamų didžiųjų farmacijos kompanijų pavadinimai, kadangi informacija apie tiriamą įmonę gali būti labai stipriai susijusi su akcijos kainos pokyčiais, jei naujienos nėra neutralios. Taigi, pagal sudarytą raktinių žodžių sąrašą buvo siekiama sumažinti aktualų informacijos srautą ir taip atrinkti aktualios kategorijos straipsnius. Straipsnių tekstinės informacijos iš interneto gavimas atliekamas naudojant Python programavimo kalbą bei pasitelkiant programinio paketo „Beautiful Soup“ technines galimybes.

Pastarojo programinio paketo galimybės leidžia patogiai peržiūrėti internetinių tinklapių struktūrą, kuri gali būti sudaryta naudojant HTML, Javascript ar PHP programavimo kalbas. Reuters naujienoms pateikti internete yra naudojama HTML kalba, todėl straipsnių gavybai naudotas Python kalba parašytas programinis kodas yra pritaikytas HTML kalbos struktūrai. Straipsnių gavimui taikytas kodas pateiktas 2 priede. Gautų straipsnių kiekis pagal bendrus raktinius žodžius bei pagal kompanijos paminėjimą straipsnyje, pateikiamas lentelėje (4.2.1. lent.).

**4.2.1. lentelė.** Straipsnių pasiskirstymas pagal temą tiriamuoju laikotarpiu

Straipsnio tema	Straipsnių skaičius per metus		
	2014	2015	2016
AstraZeneca	285	147	108
Bayer	150	48	55
Gilead	66	28	16
GlaxoSmithKline	252	95	70
Johnson & Johnson	56	48	60
Merck & Co	117	75	54
Novartis	193	109	71
Pfizer	247	177	222
Roche	164	54	60
Sanofi	147	83	66
Visos temos	10017	9773	12321

Svarbu pastebėti, jog nurodytas straipsnių skaičius nėra atitinkamai lygus dienų skaičiui, t. y. vieną dieną kompanija galėjo būti paminėta keliuose straipsniuose. Gavus straipsnius pagal raktinius žodžius tiriamajam laikotarpiui buvo pastebėta, jog raktiniai žodžiai, tinkami sveikatos priežiūros sektoriaus naujienoms, taip pat gali būti aktualūs ir kitų temų straipsnių kontekste. Pavyzdžiui, žodis „healthcare“ gali būti aktualūs apibūdinant sveikatos priežiūros sektorių ar jo įmones, tačiau šis žodis taip pat dažnas ir kitos tematikos straipsniuose, nes jis gali reikšti vieno ar daugiau asmenų individualų rūpinimąsi sveikata. Dėl šios priežasties buvo nuspręsta taikyti latentinį Dirichlet'o paskirstymo (angl. Latent Dirichlet Allocation) modelį, kuriuo sudaromos

temos, nusakančios straipsnio turinį pagal populiariausius temai priskirtus žodžius. Plačiau modelio specifinės savybės bei taikymas aptartas 5 skyriaus 5.2. poskyryje.

#### **4.3. Teksto sentimentų vertinimas informaciniuose pranešimuose**

Sentimentai gali būti skirstomi į dvi kategorijas: teigiami ir neigiami, arba dar gali būti išskiriama ir neutralių žodžių kategorija. Svarbu pastebėti, jog teksto sentimentus vertinant pagal sudarytus žodynus, yra du galimi žodynų tipai:

- Pagal kategoriją, kiekvienas žodyno žodis vertinamas vienodu įverčiu, pavyzdžiui, visi teigiamos kategorijos žodžiai vertinami 1;
- Kiekvienas žodis turi atskirą individualų vertinimą pagal savo pozityvumą ar negatyvumą, t. y. kiekvienam žodžiui suteikiamas svorinis įvertis, kuris nusako kiek stipriai žodžiu yra išreiškiamas teigiamas ar neigiamas sentimentas. Tokiu atveju vertinamas žodžių poliariškumas.

Šiame tyrime buvo sudarytas sentimentų žodynas iš keleto teigiamų ir neigiamų žodžių sąrašų, kurie buvo apjungti į vieną bendrą žodyną, pašalinant skirtinguose sąrašuose pasikartojančius žodžius. Svarbu pastebėti, jog žodyne pateikiama didžioji dalis galimų skirtingų žodžio formų, o tai leidžia vertinti gryno (angl. raw) teksto sentimentus be papildomo apdorojimo. Sudarytame žodyne yra 1312 teigiamų žodžių ir 2741 neigiamas žodis. Tyrimo metu kiekviename straipsnyje ieškoma sudaryto žodyno žodžių ir kiekvienam teigiamą sentimentą turinčiam žodžiui yra priskiriamas įvertis 1, neigiamų sentimentų žodžiams yra priskiriamas -1.

Taip pat tyrimui buvo parinkta ir dar viena alternatyva, vertinti straipsnio sentimentus naudojant jau sudarytą sentimentų žodyną. Šiam vertinimui buvo pasirinktas Python programinei kalbai pritaikytas paketas „TextBlob“. Šiuo paketu buvo vertinamas straipsnių poliariškumas, žodžiams suteikiant įvertį intervale  $[-1;1]$  atitinkamai, priklausomai nuo žodžio poliariškumo: jei žodis yra negatyvus, reikšmės yra artimesnės -1, jei pozityvus, įverčio reikšmė artimesnė 1. Straipsnio poliariškumas nustatomas kaip visų tekste esančių sentimentų įverčių vidurkis.

Straipsniai yra agreguojami dienos lygmenyje. Pirmiausia yra apskaičiuojamas kiekvienos dienos aktualaus straipsnio sentimentų įvertis, vėliau randama vidutinė sentimentų reikšmė kiekvienai dienai apskaičiuojant visų straipsnių sentimentų įverčių vidurkį. Tačiau šiuo atveju svarbu pastebėti, jog tyrime nebuvo atsisakyta informacinių pranešimų, kurie pasirodė internete savaitgalį, kadangi manoma, jog šią informaciją, jei ji nėra neutrali, investuotojai taip pat vertina, todėl šeštadienį ar sekmadienį pasirodžiusi informacija gali turėti poveikį akcijos kainoms prasidėjus naujai darbo savaitei, pavyzdžiui, pirmadienį ar taip pat ir antradienį. Todėl informaciniai pranešimai, pasirodę šeštadienį ar sekmadienį yra priskiriami artimiausio pirmadienio naujienoms. Tokiu būdu yra sudaroma informacinių pranešimų laiko eilutė. Taigi,

apžvelgus tyrimo duomenis bei jų ypatumus, aktualu apibrėžti modelius, kurie bus taikomi tyrimui. Taikytų modelių struktūra pateikiama 5 skyriuje.

## 5. TYRIMO METODAI

### 5.1. Akcijų laiko eilučių analizė

Kaip jau buvo minėta anksčiau, laiko eilutės gali turėti tam tikrus dėsningumus, pasikartojimus – trendą (angl. trend), sezoniškumą, cikliškumą. Tačiau svarbu paminėti, jog pirmiausia aktualu nustatyti ar tiriamą laiko eilutę yra stacionari. Stacionaria laiko eilutė yra laikoma tokia eilutė, kurios dispersija per laiką nekinta, kuri turi laike nekintančią autokoreliacijos struktūrą ir neturi periodinių svyravimų [75]. Todėl laiko eilutės stacionarumui įvertinti reikia nustatyti anksčiau paminėtus laiko eilutės būdingus dėsningumus.

Vienas pirmųjų žingsnių analizuojant laiko eilutes yra nustatyti trendą, kurio vertinimas gali būti atliekamas vertinant grafinį akcijos kainų kitimo laike vaizdą. Kaip jau minėta anksčiau, trendu gali būti įvertinama ilgalaikė akcijų kainų tendencija – nepaisant kainos padidėjimo ar sumažėjimo trumpuoju laikotarpiu, pavyzdžiui, kas dieną, vertinat bendrą tendenciją ilguoju laikotarpiu galima matyti kainų augimą ar sumažėjimą.

Sezoniškumas laiko eilutėje atsispindi kaip tam tikro šablono pasikartojimas praėjus tam tikram laiko tarpui, pavyzdžiui, ryškus akcijos kainos augimas ar kritimas pastebimas kiekvieno metų ketvirčio pradžioje. Sezoniškumas yra apibūdinamas formule [76]:

$$(1 - B^n)x_t = x_t - x_{t-n}; \quad (10)$$

čia  $x_t$  – laiko eilutės reikšmė  $t$  momentu,  $x_{t-n}$  laiko eilutės reikšmė su atsilikimu (angl. lag) žyminti sezoninę komponentę. Nustačius sezoniškumą, ji turi būti pašalinama iš kiekvieno laiko eilutės nario. Taip pat, esant trendui, laiko eilutė gali būti diferencijuojama ar logaritmuojama siekiant pašalinti jo įtaką. Atlikus reikiamus pakeitimus, laiko eilutė gali būti naudojama analizei ir prognozavimui. Šiame tyrime laiko eilučių stacionarumo vertinimas atliekamas naudojant R programinės įrangos paketus „tseries“, „xts“ ir „forecast“.

### 5.2. Latentinio Dirichlet'o pasiskirstymo modelio taikymas tekstinei informacijai

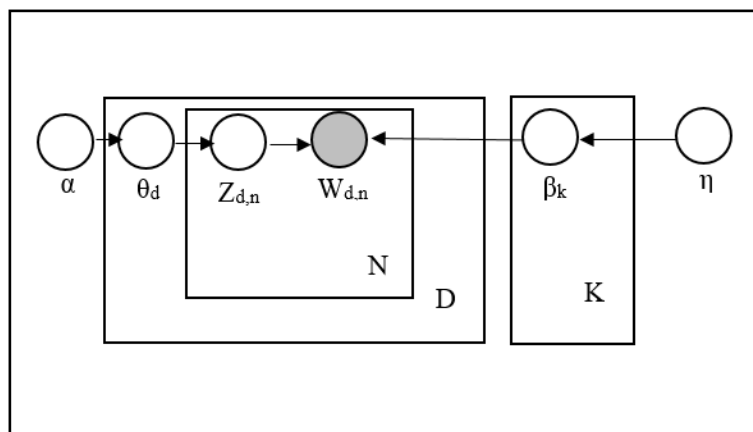
Nustačius, jog tarp tyrimui atrinktų straipsnių pagal raktinius žodžius esama straipsnių, kurie nėra susiję su sveikatos priežiūros sektoriumi, aktualu dar kartą sumažinti dimensiją, jog būtų galima vertinti aktualių straipsnių sentimentus bei informacijos įtaką akcijų kainai. Tokiam tikslui gerai tinka LDA modelis, kadangi juo galima nustatyti tekstinėje informacijoje vyraujančią temą pagal tame tekste esančius žodžius.

Prieš taikant LDA modelį, turimi duomenys (straipsniai) turi būti atitinkamai paruošti. Pirmiausia atliekant tekstinę analizę, visas straipsnio tekstas yra išskaidomas į atskirus žodžius.

Tokie atskiri žodžiai yra vadinami tokenais (angl. token). Taip pat iš teksto yra pašalinami visi skyrybos ženklai. Labai dažnai yra taikomas bereikšmių žodžių (angl. stop-words) valymas iš teksto. Šie žodžiai gali būti įvairūs jungtukai, artikkeliai<sup>2</sup> ar kiti tekstui naudingos informacijos nesuteikiantys žodžiai. Atlikus šiuos veiksmus, galima atlikti teksto lematizavimą (angl. lemmatization).

Teksto lematizavimas yra labai svarbi teksto analizės dalis. Lematizavimo metu galima nustatyti kokios kalbos daliai priklauso tekste esantis žodis. Šio tyrimo atveju kalbos dalies žymės buvo priskiriamos žodžiams, kurie buvo daiktavardžiai, būdvardžiai, veiksmažodžiai arrieveksmiai. Lematizavimo prasmė yra rasti to paties žodžio įvairias reikšmes ir jas visame tekste pakeisti į bendrinę reikšmę, pavyzdžiui, anglų kalbos veiksmažodis „be“ yra žodžio bendratis, o veiksmažodis „is“ yra esamojo laiko veiksmažodis. Jei nebūtų atliekamas lematizavimas, šie veiksmažodžiai būtų laikomi negiminingais, nors tai ir yra tas pats žodis, tačiau pateiktas kita forma.

Šiame tyrime straipsnio tekstas yra transformuojamas į žodžių vektorių, todėl lematizavimas leidžia rasti tuos pačius žodžius (kurių dažnio vertinimas aktualus vėliau) nepriklausomai nuo žodžio formos. Iš visų modifikuotų straipsnių žodžių yra sudaromas korpusas (angl. corpus) ir žodynas. Yra laikoma, jog korpusą sudaro  $M$  dokumentų, kurių kiekvieno ilgis yra  $N_i$ , t. y. korpuse pateikiami visi straipsniai kaip jiems priklausančių žodžių vektorius. Žodynas yra sudaromas iš visų straipsniuose esančių žodžių. Toliau atliekamas temų formavimas remiantis šia logine schema [77], pateikta paveiksle (5.2.1. pav.):



**5.2.1 pav.** LDA modelio loginė schema

Pateiktame 5.2.1. paveiksle,  $K$  žymi bendrą temų skaičių;  $\beta_k$  temos pasiskirstymą per visą žodyną, t. y. pažymimi vienai temai priklausančys žodžiai iš žodyno;  $D$  žymi dokumentų (šiam tiriamajame darbe straipsnių) skaičių;  $\theta_d$  nusako kiek skirtingų temų sudaro vieną dokumentą;  $N$

<sup>2</sup> Tyrime buvo naudojama informacija anglų kalba. Artikkeliais įvardinami žodžiai „a“, „an“, „the“ ir kt.

žymi žodžių skaičių viename dokumente;  $Z_{d,n}$  žymi kiekvieno žodžio priklausymą temoms (vienas žodis gali priklausyti ne vienai temai);  $W_{d,n}$  žymi tiriamą žodį;  $\alpha$  ir  $\eta$  yra Dirichlet'o parametrai.

Atlikus anksčiau aprašytus žingsnius, gautos 50 temų, sudarytos iš tiriamų straipsnių žodžių. Peržiūrėjus sudarytų temų sąrašą, atrinktos 4 temos, kuriose geriausiai atsiskyrė informacija apie tiriamas kompanijas. Geriausiai tikusios temos, kuri apėmė beveik visas tiriamas kompanijas, žodžiai bei jų tikimybės pateikiamos lentelėje (5.2.1. lent.). Likusių atrinktų temų žodžiai bei tikimybės pateikiami 3 priede.

**5.2.1. lentelė.** LDA modelio temos žodžiai bei jų pasiskirstymo straipsniuose tikimybės

Žodis	Žodžio tikimybė
drug	0,047736058
novartis	0,018766272
medicine	0,014632814
billion	0,014360143
gsk	0,014006422
new	0,013906346
sale	0,013562665
cancer	0,01206863
roche	0,010254964
drugmaker	0,010235737
treatment	0,007876196
year	0,007511401
product	0,007450657
analyst	0,007087882
patent	0,007049183
swiss	0,006992707
astrazeneca	0,006959797
merck	0,006870165
ben	0,006553036
sanofi	0,006371402

Taigi, remiantis šiomis temomis turimi straipsniai buvo filtruojami dar kartą – temų žodžiai buvo naudojami kaip raktiniai atrenkant straipsnius ir priskiriant juos kiekvienai kompanijai atskirai. Atsižvelgus į tai, jog straipsniuose gali būti minima ne viena kompanija, pirmiausia straipsnis buvo priskiriamas abiem kompanijoms kaip galimai aktuali informacija. Taip buvo atliktas straipsnių priskyrimas kiekvienai kompanijai.

Kaip jau buvo pastebėta, tiriant straipsnius yra galima situacija, jog viename straipsnyje gali būti paminėtos kelios kompanijos, todėl siekiant atsižvelgti į tai, jog straipsnis, kuriame kompanija tik paminėta, tačiau tai nėra aktuali informacija apie pačią kompaniją, turi būti pašalinamas iš kompanijai priskirtų straipsnių. Šiame tyrime toks straipsnių vertinimas buvo naudojamas taikant TF-IDF (angl. term frequency-inverse document frequency) statistiką. Šia statistika galima

įvertinti žodžio svarbą tekste pagal jo dažnį. Todėl kiekvienos kompanijos atrinktų straipsnių žodžiams buvo skaičiuojama ši statistika ir remiantis ja, jei, tarkime, straipsnyje tirama „AstraZeneca“ kompanija paminėta tik kartą, tačiau esama kompanijų, pavyzdžiui, „Pfizer“ ar „Gilead“, kurios paminėtos daugiau nei kartą, toks straipsnis pašalinamas iš „AstraZeneca“ straipsnių sąrašo. Suskaičiavus TF-IDF statistikos reikšmes straipsniams, pateikiamas pašalintų bei toliau tyrimui likusių straipsnių skaičius lentelėje (5.2.2. lent.).

**5.2.2. lentelė.** Straipsnių skaičius sentimentų laiko eilutės sudarymui

Kompanija	Pašalintų straipsnių kiekis	Likę straipsniai
AstraZeneca	19	521
Bayer	7	246
Gilead	11	99
GlaxoSmithKline	8	409
Johnson & Johnson	2	162
Merck & Co	9	237
Novartis	8	365
Pfizer	22	624
Roche	6	272
Sanofi	12	284

### 5.3. ARIMA modelio taikymas naujienų poveikio akcijų kainoms analizei

ARIMA modelis gali būti taikomas įprastos laiko eilutės modeliavimui, tačiau šiam modeliui taip pat gali būti papildomai įtraukiami išoriniai kintamieji. Šio tyrimo atveju išorinis kintamasis yra informacinių naujienų laiko eilutė. Įprasto ARIMA modelio formulė (5) pateikta anksčiau. ARIMA modelio su išoriniu kintamuoju formulė atrodo kiek kitaip – pridedamas papildomas kintamasis:

$$Y_t - \phi_1 Y_{t-1} = \mu - \theta_1 e_{t-1} + \beta(X_t - \phi_1 X_{t-1}); \quad (11)$$

čia reiškiny  $\beta(X_t - \phi_1 X_{t-1})$  žymi papildomai įtraukiamą išorinį kintamąjį.

Šiame tyrime taip pat buvo siekiama įvertinti ne tik informacinio pranešimo poveikį tos dienos akcijų kainai, kai pranešimas pasirodė, tačiau taip pat nustatyti, ar pasirodžius pranešimui jo poveikis pastebimas dar keletą dienų po pasirodymo. Dėl šios priežasties kaip išoriniai kintamieji taip pat buvo įtrauktos pavėlintos (angl. lag) straipsnių sentimentų reikšmės. Vertinamos iki 3 dienų pavėlintos reikšmės.

Sudarant ARIMA modelį, svarbu teisingai parinkti AR ir MA dalies eilę. Šiam tikslui buvo naudojama „auto.arima“ funkcija, kuri automatiškai nustato geriausias galimas AR ir MA reikšmes pagal Akaikės informacinį kriterijų (AIC). Šiuo kriterijumi yra įvertinamas modelio tinkamumas duomenims bei galimas informacijos praradimas sudarant modelį, t. y. AIC kriterijaus

reikšmė geresnė to modelio, kurio sudarymui informacijos praradimas yra mažesnis. Taip pat svarbu pastebėti, jog kuo mažesnė AIC reikšmė, tuo modelis geresnis. Tačiau šiuo kriterijumi nėra vertinama paties modelio kokybė. Toliau pateikiama AIC formulė:

$$AIC = 2k - 2\ln(\hat{L}); \quad (12)$$

čia  $k$  žymi kintamųjų skaičių, o  $\hat{L}$  – tikėtinumo funkciją, kuri nusakoma formule:

$$\hat{L} = P(x|\hat{\theta}, M); \quad (13)$$

čia  $M$  žymi modelį,  $x$  – duomenų rinkinį, o  $\hat{\theta}$  – parametrų, skirtų maksimizuoti tikėtinumo funkciją, reikšmes.

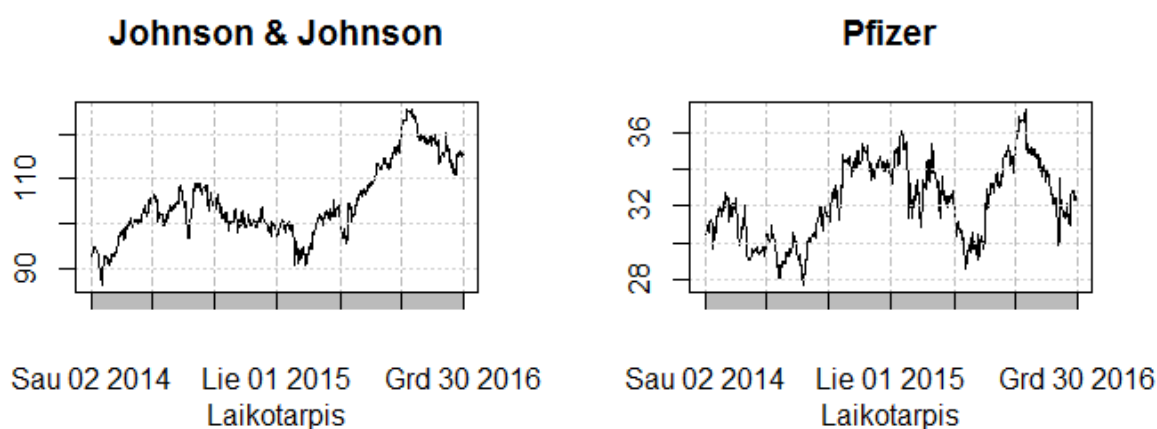
Taip pat atliekant tyrimą buvo siekiama įvertinti priežastingumo ryšį tarp akcijų kainos bei informacinių pranešimų. Šio ryšio nustatymui taikytas Granger'io priežastingumo testas. Šiuo testu norima nustatyti ar naujienos bei jų pavėlintos reikšmės (angl. lag) veikia akcijos kainą, ar esama atvirkštinio priežastingumo. Tokiam ryšiui nustatyti naudojamas R programos paketo „lmtest“ funkcija „grangertest“. Šia funkcija yra vertinami du galimi modeliai: vienas, su išoriniu kintamuoju, kitas be jo. Veiksnių Granger'io priežastingumui nustatyti yra tikrinama nulinė hipotezė, jog „veiksny X neturi priežastingumo ryšio su veiksnium Y“. Nulinė hipotezė priimama ar atmetama remiantis reikšmingumo įverčio  $p$  reikšme – jei  $p$  reikšmė mažesnė už pasirinktą reikšmingumo lygmenį, nulinė hipotezė atmetama.

Taigi, nustačius tyrimui taikytus modelius, aktualu apžvelgti gautus rezultatus analizuojant informacinių pranešimų poveikį tiriamų kompanijų akcijų kainoms. Pirmiausia pateikiami tyrimo rezultatai analizuojant laiko eilutės, sudarytos vertinant informacinių pranešimų sentimentus, įtakos akcijų kainoms tyrimo rezultatai. Toliau pateikiami rezultatai, gauti tiriant poliariškumu vertinamų naujienų poveikį akcijų kainoms.

## TYRIMŲ REZULTATAI IR JŲ APTARIMAS

### 6. TYRIMO REZULTATAI

Pirmiausia buvo atliktas akcijų kainų laiko eilučių stacionarumo vertinimas. Kiekvienos iš tiriamų įmonių akcijų kainų laiko eilutė buvo tikrinama dėl sezoniškumo bei cikliškumo, taip pat buvo vertinamas bendras trendas. Šios analizės rezultatai pateikiami toliau. Pirmiausia buvo analizuojamos grynos tiriamų kompanijų akcijų laiko eilutės. „Johnson & Johnson“ bei „Pfizer“ kompanijų laiko eilučių grafinis vaizdas pateikiamas 6.1. paveiksle. Likusių laiko eilučių grafinis vaizdas pateikiamas 4 priede.



**6.1. pav.** Grynųjų akcijų kainų laiko eilučių grafinis vaizdas

Taigi, remiantis šių laiko eilučių grafiniu vaizdu, galima pastebėti, jog kelios kompanijos turi pastebimą tendą: „Johnson & Johnson“ pastebimas akcijų kainos augimas, „GlaxoSmithKline“ atveju pastebimas akcijų kainos mažėjimo trendas, o „Bayer“ atveju kainų mažėjimo tendencija pastebima tik nuo 2015 metų. Taip pat svarbu pastebėti, jog „Roche“ bei „AstraZeneca“ atveju matomas staigus kainos kritimas yra susijęs su akcijų padalijimu (angl. stock split), o ne su informacinėmis naujienomis. Likusių kompanijų akcijų kainos pasižymi didele variacija, kuri rodo laiko eilutės nestacionarumą. Dėl šios priežasties siekiama nustatyti ar diferencijuojant laiko eilutę ji taptų stacionari.

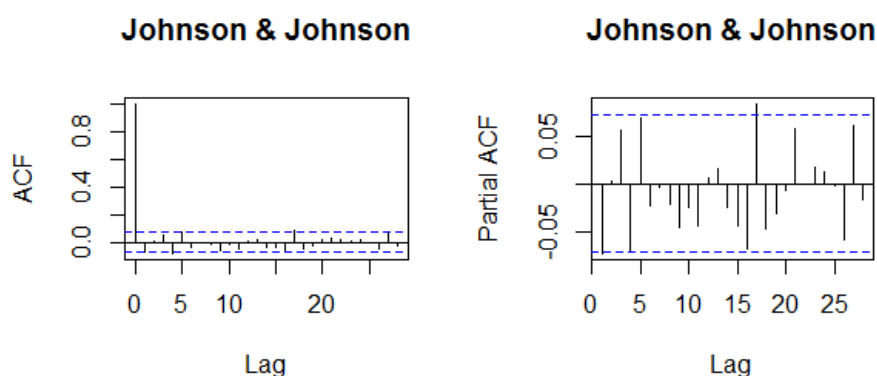
Visos tiriamos laiko eilutės buvo diferencijuojamos, kiekvienai iš eilučių buvo pritaikytas papildytas Dikio-Fulerio testas - ADF (angl. Augmented Dickey-Fuller test). Šio testo nulinė hipotezė teigia, jog laiko eilutė yra stacionari. Gautos testo reikšmės pateikiamos lentelėje (6.1.1. lent.).



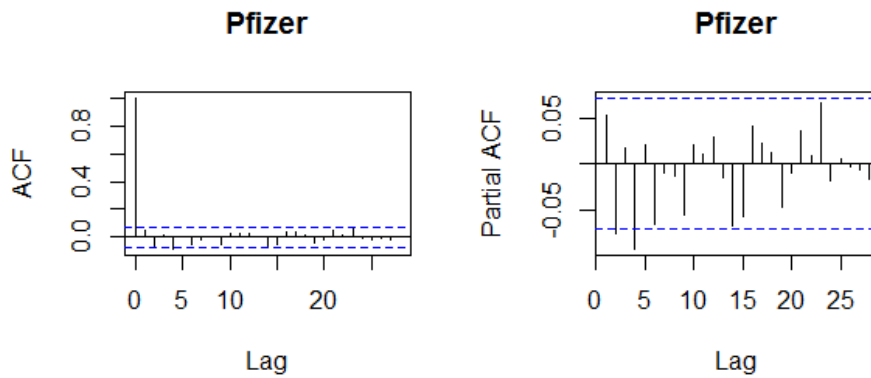
### 6.1.1. lentelė. Papildyto Dikio – Fulerio testo reikšmės akcijų kainų laiko eilutėms

Kompanija	ADF testo reikšmė	p reikšmė
AstraZeneca	-9,090	0,01
Bayer	-10,259	0,01
Gilead	-9,106	0,01
GlaxoSmithKline	-9,906	0,01
Johnson & Johnson	-9,316	0,01
Merck & Co	-9,453	0,01
Novartis	-9,322	0,01
Pfizer	-9,598	0,01
Roche	-9,140	0,01
Sanofi	-9,921	0,01

Svarbu pastebėti, jog lentelėje pateikta p reikšmė yra didesnė nei tikroji, kadangi R programoje pateikiama p reikšmė su priedu, jog tikroji reikšmė yra mažesnė nei atspausdinta. Remiantis rezultatais, pastebima, jog laiko eilutės stacionarios, kadangi p reikšmės yra mažesnės nei reikšmingumo lygmuo. Stacionarumo įvertinimui akcijų kainoms taip pat braižomi autokoreliacijos (ACF) ir dalinės autokoreliacijos (PACF) grafikai, kuriais remiantis, galima patikrinti laiko eilutės stacionarumą. Kompanijų „Johnson & Johnson“ bei „Pfizer“ grafikai po pirmojo diferencijavimo atitinkamai pateikiami 6.2. ir 6.3. paveiksluose. Kairėje pusėje pateikiamas autokoreliacijos, dešinėje – dalinės autokoreliacijos grafikas. Likusių kompanijų pilnos ir dalinės autokoreliacijos grafikai pateikiami 5 priede.



6.2. pav. „Johnson & Johnson“ autokoreliacijos ir dalinės autokoreliacijos grafinis vaizdas



**6.3. pav.** „Pfizer“ autokoreliacijos ir dalinės autokoreliacijos grafinis vaizdas

Remiantis gautais pilnos ir dalinės autokoreliacijos grafikai, matoma, jog laiko eilučių diferencijavimas išsprendžia nestacionarumo problemą: po atliktų korekcijų laiko eilutės yra stacionarios, kadangi nei pilnos, nei dalinės autokoreliacijos atvejais reikšmės beveik neviršija rėžių, žymimų punktyrine linija. Vis dėl to, modeliavimui buvo naudojamos grynosios laiko eilutės, kadangi taikomas integruotas ARIMA modelis gali nustatyti ar diferencijavimas reikalingas ir atitinkamai pritaikyti laiko eilutę. Informaciniai pranešimai tyrime buvo vertinami pagal sentimentus ir poliariškumą. Pirmiausia pateikiami rezultatai, gauti vertinant sentimentų laiko eilutės įtaką akcijų kainų laiko eilutei kiekvienos kompanijos atveju. Toliau pateikiami rezultatai, gauti vertinant poliariškumo laiko eilutės (sudarytos naudojant „TextBlob“ paketą) įtaką akcijų kainų laiko eilutei kiekvienos kompanijos atveju.

## **6.1. Informacinių pranešimų pagal sentimentus įtakos akcijų kainoms vertinimas**

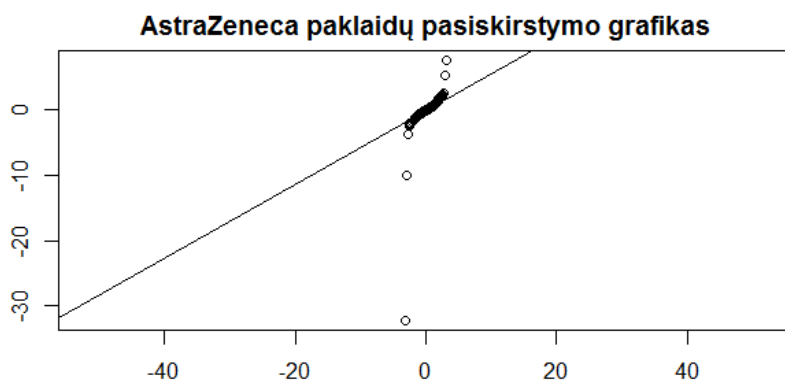
### **6.1.1. „AstraZeneca“ kompanija**

„AstraZeneca“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei įtraukiant sentimentų laiko eilutę ir taip pat pavėlintų reikšmių sentimentų eilutes. Į modelį yra įtraukiami 3 papildomi kintamieji, kurie žymi sentimentų laiko eilutės pavėlinimą (angl. lag): atitinkamai 1, 2 ir 3 dienų pavėlinimą. Taigi modeliavimas atliekamas su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.1.1. lent.) pateikiami gauti modeliavimo rezultatai.

#### 6.1.1.1. lentelė. Sentimentų modelio „AstraZeneca“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	0,022951	0,02811
Sentiment_lag1	-0,02665	0,0364
Sentiment_lag2	-0,01104	0,3862
Sentiment_lag3	-0,00967	0,35475

Remiantis lentelėje pateiktais rezultatais, matoma, jog sentimentų laiko eilutės įtaka akcijų kainų laiko eilutei, bei viena diena pavėlinta sentimentų eilutė yra statistiškai reikšmingi kintamieji. Tačiau likę pavėlintų reikšmių kintamieji nėra reikšmingi. Gauto modelio AIC reikšmė lygi 2726,3, RMSE reikšmė lygi 1,47174. Taigi galima teigti, jog modelio paklaida yra maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.1.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



6.1.1. pav. „AstraZeneca“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Kadangi modelis yra tinkamas, taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė bei jos pavėlintos reikšmės neveikia akcijų kainų laiko eilutės bei jos pavėlintų reikšmių, kadangi Fišerio statistikos p reikšmė yra lygi 0,1544, todėl priimama nulinė hipotezė, jog neegzistuoja Granger'io priežastingumas. Tačiau nustatyta, jog akcijos kaina bei jos pavėlintos reikšmės veikia informacinių pranešimų sentimentus bei jų pavėlintas reikšmes, kadangi su 99 proc. pasiklovimo lygmeniu gaunama p reikšmė yra 0,001029, todėl nulinė Granger'io priežastingumo hipotezė atmetama ir teigiama, jog egzistuoja akcijų kainos ir sentimentų laiko eilutės priežastingumo ryšys.

#### 6.1.2. „Bayer“ kompanija

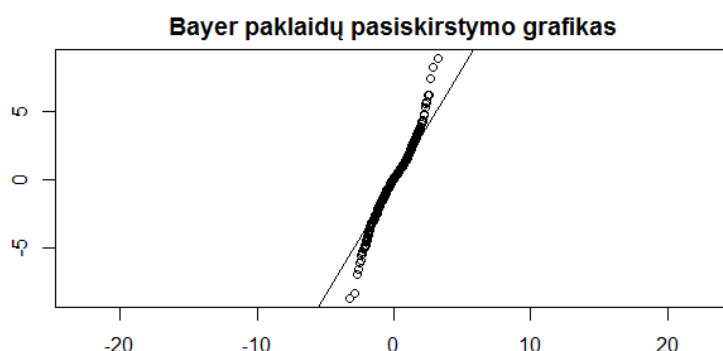
„Bayer“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (1, 1, 0). Pagal šiuos

parametrus sudarytas modelis su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.2.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.1.2.1. lentelė.** Sentimentų modelio „Bayer“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
ar1	-0,112943	0,001863
Sentiment	0,074705	0,000636
Sentiment_lag1	0,056661	0,024008
Sentiment_lag2	0,030413	0,225221
Sentiment_lag3	0,00918	0,672972

Remiantis lentelėje pateiktais rezultatais, matoma, jog statistiškai reikšminga autokoreliacija žyminti modelio dalis ar1, taip pat sentimentų laiko eilutė bei vienos dienos pavėlinta sentimentų laiko eilutė. Tačiau dviejų ir trijų dienų pavėlintų sentimentų laiko eilutės nėra reikšmingos. Gauto modelio AIC reikšmė lygi 3201,06, RMSE reikšmė lygi 2,015317. Taigi galima teigti, jog modelio paklaida yra maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.2.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.1.2.1. pav.** „Bayer“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį, tenkinama paklaidų normalumo prielaida. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė bei jos pavėlintos reikšmės neveikia akcijos kainų laiko eilutės bei jos pavėlintas reikšmių, kadangi Fišerio statistikos p reikšmė yra lygi 0,9086, todėl priimama nulinė hipotezė, jog Granger'io priežastingumas neegzistuoja. Taip pat nenustatyta, jog akcijų kainos veiktų sentimentų laiko eilutę, kadangi Fišerio statistikos p reikšmė lygi 0,9168. Taigi „Bayer“ kompanijos atveju galima teigti, jog informacinių pranešimų įtaka akcijų kainai pastebima esamu momentu, kai pasirodo pranešimas, ir po vienos dienos pasirodžius pranešimui. Tačiau nenustačius Granger'io priežastingumo, galima teigti, informacinių pranešimų, vertinamų pagal sentimentus, ir akcijos kainų pavėlintos laiko eilutės tarpusavyje neturi priežastinio ryšio.

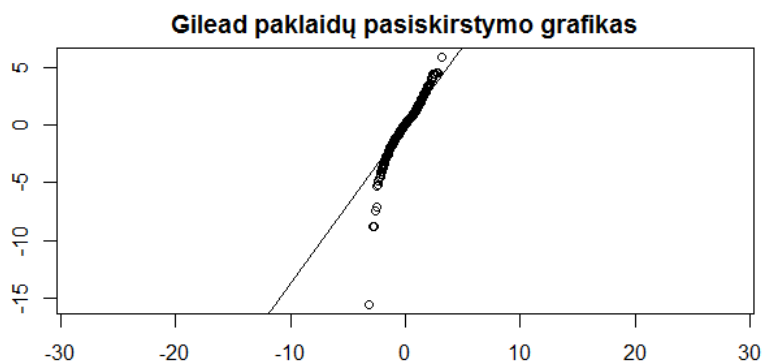
### 6.1.3. „Gilead“ kompanija

„Gilead“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.3.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.1.3.1. lentelė.** Sentimentų modelio „Gilead“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	-0,01154206	0,5911
Sentiment_lag1	-0,00044484	0,9861
Sentiment_lag2	0,00844222	0,7379
Sentiment_lag3	-0,00102094	0,9612

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 3007,53, RMSE reikšmė lygi 1,774354. Taigi galima teigti, jog modelio paklaida yra maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.3.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.1.3.1. pav.** „Gilead“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė neveikia akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,4567. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,7457. Taigi galima teigti, jog „Gilead“ kompanijos atveju nustatyta, jog informaciniai pranešimai neturi poveikio akcijų kainai.

### 6.1.4. „GlaxoSmithKline“ kompanija

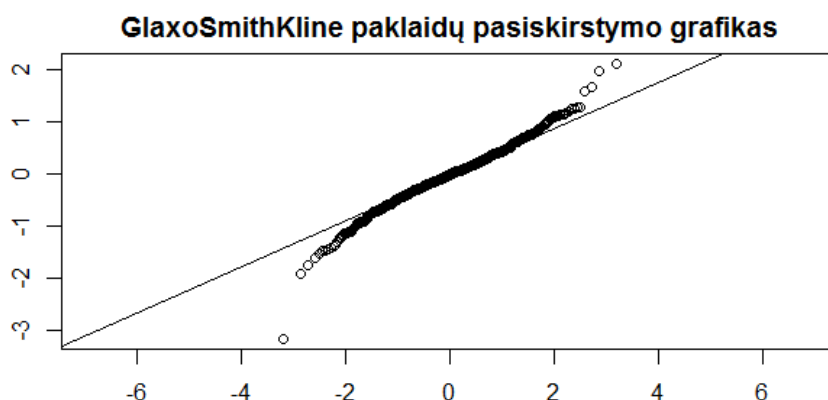
„GlaxoSmithKline“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos

parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.4.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.1.4.1. lentelė.** Sentimentų modelio „GlaxoSmithKline“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	0,00357074	0,2454
Sentiment_lag1	0,00421931	0,2568
Sentiment_lag2	-0,00028971	0,9379
Sentiment_lag3	-0,00277146	0,3673

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 1183,71, RMSE reikšmė lygi 0,5277134. Taigi galima teigti, jog modelio paklaida yra maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.4.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.1.4.1. pav.** „GlaxoSmithKline“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė neveikia akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,1465. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,5131. Taigi galima teigti, jog „GlaxoSmithKline“ kompanijos atveju nustatyta, jog informaciniai pranešimai neturi poveikio akcijų kainai.

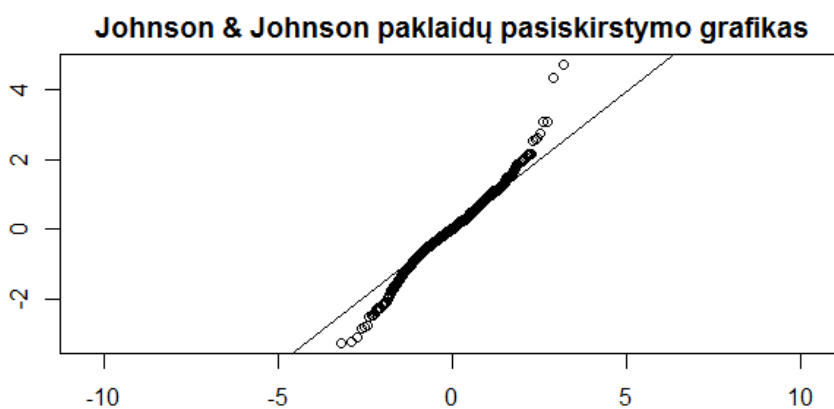
#### **6.1.5. „Johnson & Johnson“ kompanija**

„Johnson & Johnson“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.5.1. lent.) pateikiami gauti modeliavimo rezultatai.

#### 6.1.5.1. lentelė. Sentimentų modelio „Johnson & Johnson“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	-0,00030872	0,9697
Sentiment_lag1	0,00620744	0,5185
Sentiment_lag2	-0,00161051	0,8669
Sentiment_lag3	-0,00596174	0,4616

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 2082,06, RMSE reikšmė lygi 0,9589677. Sudaryto modelio paklaida yra labai maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.5.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



6.1.5.1. pav. „Johnson & Johnson“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė neveikia akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,3187. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,5083. Taigi galima teigti, jog „Johnson & Johnson“ kompanijos atveju nustatyta, jog informaciniai pranešimai neturi poveikio akcijų kainai.

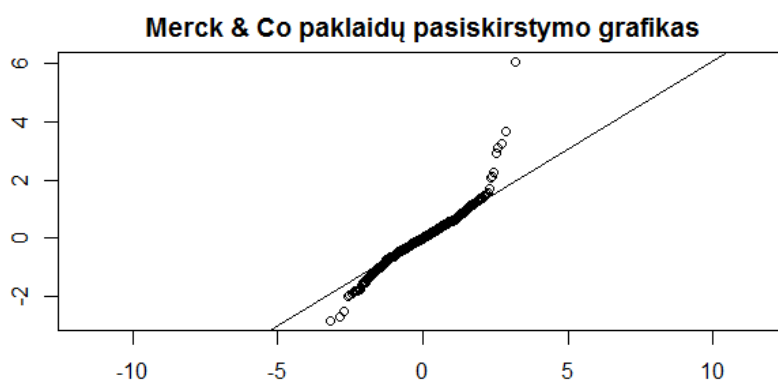
#### 6.1.6. „Merck & Co“ kompanija

„Merck & Co“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.6.1. lent.) pateikiami gauti modeliavimo rezultatai.

#### 6.1.6.1. lentelė. Sentimentų modelio „Merck & Co“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	0,00352	0,3829
Sentiment_lag1	-0,0016535	0,7383
Sentiment_lag2	0,0060232	0,2235
Sentiment_lag3	0,0031877	0,4294

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 1683,68, RMSE reikšmė lygi 0,7358122. Taigi galima teigti, jog modelio paklaida yra maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.6.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



6.1.6.1. pav. „Merck & Co“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė neveikia akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,7107 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,1701. Taigi galima teigti, jog „Merck & Co“ kompanijos atveju nustatyta, jog informaciniai pranešimai neturi poveikio akcijų kainai.

#### 6.1.7. „Novartis“ kompanija

„Novartis“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.7.1. lent.) pateikiami gauti modeliavimo rezultatai.



#### 6.1.7.1. lentelė. Sentimentų modelio „Novartis“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	0,0020898	0,7034
Sentiment_lag1	-0,0047046	0,4751
Sentiment_lag2	-0,0044674	0,4977
Sentiment_lag3	0,0019274	0,7255

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 2121,51, RMSE reikšmė lygi 0,9844547. Taigi galima teigti, jog modelio paklaida yra maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.7.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



6.1.7.1. pav. „Novartis“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp sentimentų laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,9617 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,5118. Taigi galima teigti, jog „Novartis“ kompanijos atveju nustatyta, kad informaciniai pranešimai neturi poveikio akcijų kainai.

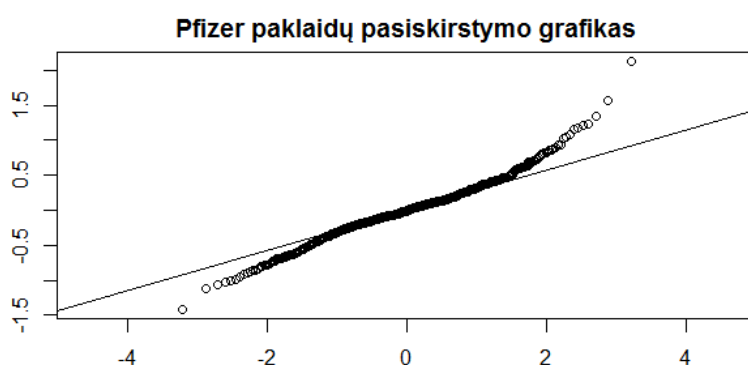
#### 6.1.8. „Pfizer“ kompanija

„Pfizer“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.8.1. lent.) pateikiami gauti modeliavimo rezultatai.

#### 6.1.8.1. lentelė. Sentimentų modelio „Pfizer“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	0,0038311	0,1676
Sentiment_lag1	0,0025175	0,4453
Sentiment_lag2	0,005383	0,1027
Sentiment_lag3	0,0028313	0,3078

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 665,65, RMSE reikšmė lygi 0,3739385. Taigi galima teigti, jog modelio paklaida yra maža. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.8.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



6.1.8.1. pav. „Pfizer“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį, patvirtinama paklaidų normalumo prielaida. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp sentimentų laiko eilutė ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,5283 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,3098. Taigi galima teigti, jog „Pfizer“ kompanijos atveju nustatyta, kad informaciniai pranešimai neturi poveikio akcijų kainai.

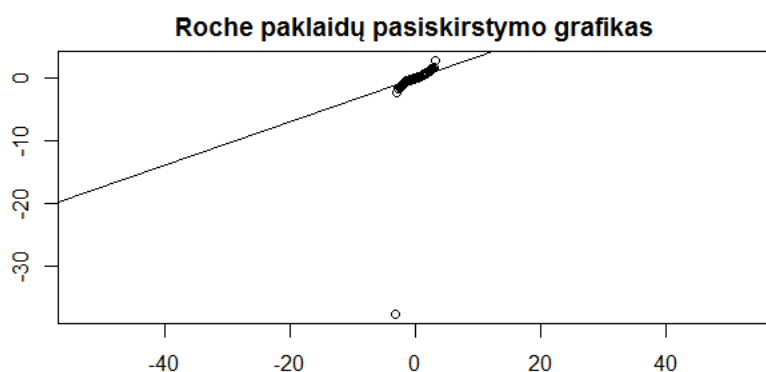
#### 6.1.9. „Roche“ kompanija

„Roche“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.9.1. lent.) pateikiami gauti modeliavimo rezultatai.

#### 6.1.9.1. lentelė. Sentimentų modelio „Roche“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	0,0022539	0,7986
Sentiment_lag1	0,0027735	0,7936
Sentiment_lag2	0,0024352	0,8182
Sentiment_lag3	0,0000654	0,9941

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 2692,83, RMSE reikšmė lygi 1,439354. Taigi galima teigti, jog modelio paklaida nėra didelė. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.9.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



6.1.9.1. pav. „Roche“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė neveikia akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,594 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,9862. Taigi galima teigti, jog „Roche“ kompanijos atveju nustatyta, jog informaciniai pranešimai neturi poveikio akcijų kainai.

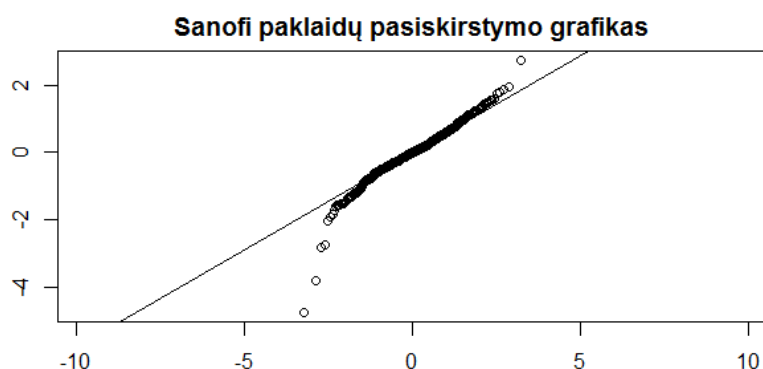
#### 6.1.10. „Sanofi“ kompanija

„Sanofi“ informacinių pranešimų, sudarytų pagal sentimentus, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.1.10.1. lent.) pateikiami gauti modeliavimo rezultatai.

#### 6.1.10.1. lentelė. Sentimentų modelio „Sanofi“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Sentiment	-0,0024671	0,555
Sentiment_lag1	-0,0026858	0,5979
Sentiment_lag2	-0,0044056	0,3871
Sentiment_lag3	-0,0021061	0,6144

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų. Gauto modelio AIC reikšmė lygi 1568,97, RMSE reikšmė lygi 0,681779. Taigi galima teigti, jog modelio paklaida nėra didelė. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.1.10.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



6.1.10.1. pav. „Sanofi“ sentimentų modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog sentimentų laiko eilutė neveikia akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,9692 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų sentimentų, šio testo Fišerio statistika lygi 0,867. Taigi galima teigti, jog „Sanofi“ kompanijos atveju nustatyta, jog informaciniai pranešimai neturi poveikio akcijų kainai.

Taigi, atlikus informacinių pranešimų sentimentų laiko eilutės įtakos akcijų kainoms tyrimą taikant ARIMA modelius, nustatyta, jog informaciniai pranešimai turi įtaką tik „AstraZeneca“ bei „Bayer“ kompanijų atvejais. Taip pat nustatyta, jog šių kompanijų informaciniai pranešimai turi poveikį ne tik tos dienos akcijai, kada pranešimas pasirodė spaudoje, tačiau ir kitą dieną, t. y. vienos dienos vėluojančios sentimentų reikšmės „AstraZeneca“ ir „Bayer“ atveju yra statistiškai reikšmingos. Visų kitų kompanijų informacinių pranešimų reikšmingo poveikio akcijų kainai nebuvo nustatyta.

## 6.2. Informacinių pranešimų pagal poliariškumą įtakos akcijų kainoms vertinimas

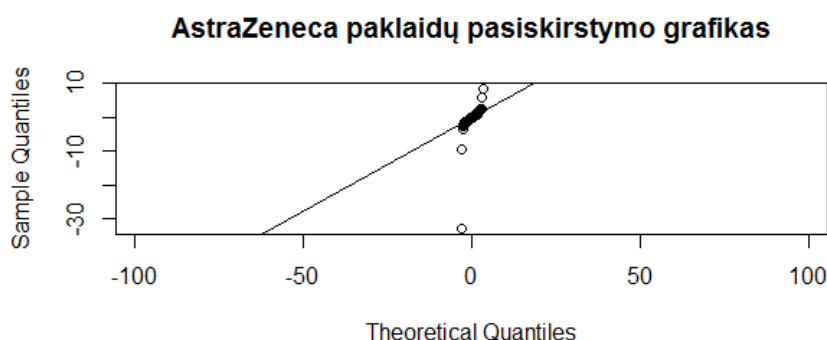
### 6.2.1. „AstraZeneca“ kompanija

„AstraZeneca“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais: poliariškumo laiko eilutės kintamasis bei 1, 2 ir 3 vėluojančių laiko momentų (angl. lag) poliariškumo laiko eilutės. Toliau lentelėje (6.2.1.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.1.1. lentelė.** Poliariškumo modelio „AstraZeneca“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	-1,83933	0,07742
Polarity_lag1	-1,87487	0,14187
Polarity_lag2	-2,08605	0,10219
Polarity_lag3	-0,70522	0,49837

Remiantis lentelėje pateiktais rezultatais, matoma, jog tik su 90 proc. pasikliautumu informacinių pranešimų poliariškumo laiko eilutė yra statistiškai reikšminga. Šiuo atveju gauti rezultatai artimi šios kompanijos modeliui su sentimentų laiko eilute, tačiau šiuo atveju viena diena vėluojanti poliariškumo laiko eilutė nėra statistiškai reikšminga, priešingai nei sentimentų modelio atveju. Gauto modelio AIC reikšmė lygi 2744,6, RMSE reikšmė lygi 1,489764. Poliariškumo ir sentimentų modelių „AstraZeneca“ kompanijai AIC bei RMSE reikšmės beveik nesiskiria, todėl galima teigti, jog abu modeliai gerai tinka duomenims. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.1.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.2.1.1. pav.** „AstraZeneca“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger‘io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta,

jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,7502 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo, šio testo Fišerio statistika lygi 0,5412. Taigi, galima teigti, jog informaciniai pranešimai, vertinami poliariškumu, iš dalies turi poveikį akcijų kainai, kadangi gauta, jog poliariškumo laiko eilutė yra statistiškai reikšmingas kintamasis esant 90 proc. pasiklovimo intervalui.

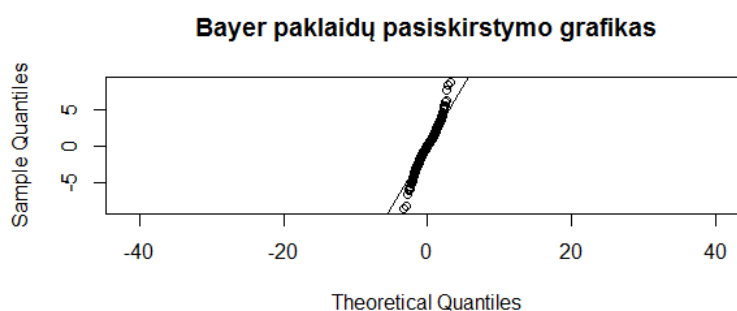
### 6.2.2. „Bayer“ kompanija

„Bayer“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (1, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.2.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.2.1. lentelė.** Poliariškumo modelio „Bayer“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
ar1	-0,118318	0,001094
Polarity	0,575944	0,7399
Polarity_lag1	2,24795	0,272052
Polarity_lag2	3,162029	0,122317
Polarity_lag3	1,444192	0,405048

Remiantis lentelėje pateiktais rezultatais, matoma, jog statistiškai reikšminga tik modelio autokoreliaciją žyminti dalis ar1. Kitų statistiškai reikšmingų kintamųjų nėra. Šiuo atveju gauti rezultatai yra priešingi šios kompanijos modeliui su sentimentų laiko eilute, kadangi poliariškumo modelio atveju yra reikšminga tik modelio autokoreliaciją žyminti dalis ar1. Gauto modelio AIC reikšmė lygi 3210,59, RMSE reikšmė lygi 2,028128. Poliariškumo ir sentimentų modelių „Bayer“ kompanijai AIC bei RMSE reikšmės beveik nesiskiria, todėl galima teigti, jog abu modeliai gerai tinka duomenims. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.2.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.2.2.1. pav.** „Bayer“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,8931 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,3557.

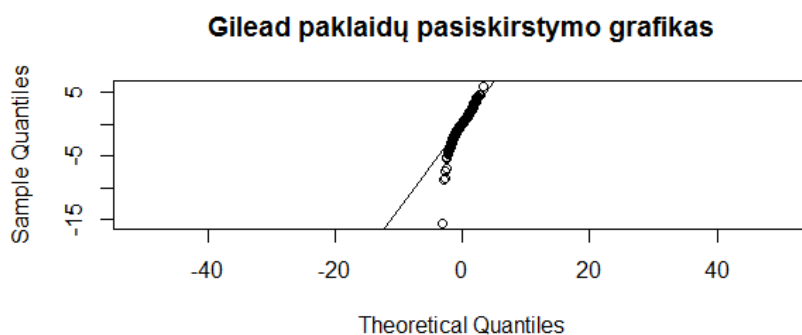
### 6.2.3. „Gilead“ kompanija

„Gilead“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.3.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.3.1. lentelė.** Poliariškumo modelio „Gilead“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	-0,37545	0,8531
Polarity_lag1	-1,504	0,5357
Polarity_lag2	-1,92496	0,425
Polarity_lag3	-0,36627	0,8545

Remiantis lentelėje pateiktais rezultatais, matoma, jog statistiškai reikšmingų kintamųjų nėra. Šios kompanijos sentimentų bei poliariškumo modeliai iš esmės nesiskiria, kadangi abiejų modelių rezultatai rodo, jog informacinių pranešimų įtaka akcijų kainai nėra statistiškai reikšminga. Gauto modelio AIC reikšmė lygi 3007,44, RMSE reikšmė lygi 1,774254. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.3.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.2.3.1. pav.** „Gilead“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta,

jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,2939 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,837.

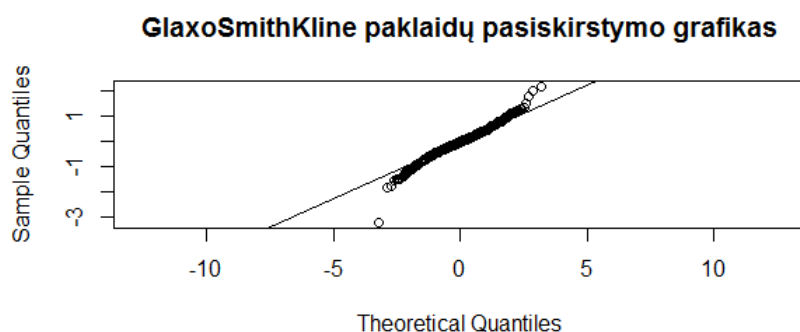
#### 6.2.4. „GlaxoSmithKline“ kompanija

„GlaxoSmithKline“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.4.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.4.1. lentelė.** Poliariškumo modelio „GlaxoSmithKline“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	0,24664	0,4927
Polarity_lag1	0,57424	0,1889
Polarity_lag2	0,66591	0,1276
Polarity_lag3	0,17494	0,6266

Remiantis lentelėje pateiktais rezultatais, matoma, jog statistiškai reikšmingų kintamųjų nėra. Šios kompanijos sentimentų bei poliariškumo modeliai iš esmės nesiskiria, kadangi abiejų modelių rezultatai rodo, jog informacinių pranešimų įtaka akcijų kainai nėra statistiškai reikšminga. Gauto modelio AIC reikšmė lygi 1185,15, RMSE reikšmė lygi 0,5282184. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.4.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.2.4.1. pav.** „GlaxoSmithKline“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat tikrinamas Granger'io priežastingumas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,2634 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo nebuvimą.



Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,4915. Taigi galima teigti, jog informaciniai pranešimai neturi įtakos akcijų kainai „GlaxoSmithKline“ kompanijos atveju.

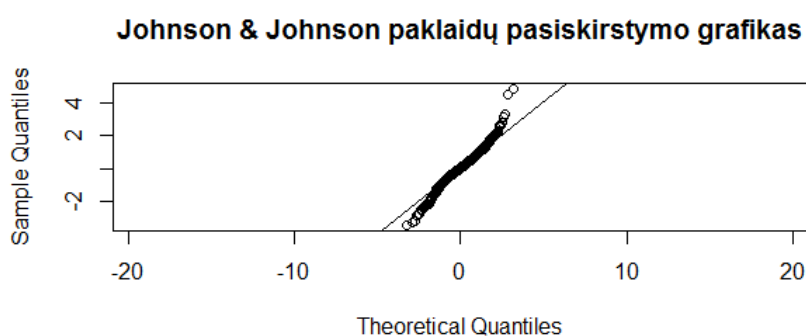
### 6.2.5. „Johnson & Johnson“ kompanija

„Johnson & Johnson“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.5.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.5.1. lentelė.** Poliariškumo modelio „Johnson & Johnson“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	0,32088	0,6967
Polarity_lag1	-0,27046	0,7814
Polarity_lag2	-0,2189	0,8222
Polarity_lag3	-0,20856	0,7996

Remiantis lentelėje pateiktais rezultatais, matoma, jog statistiškai reikšmingų kintamųjų nėra. Šios kompanijos sentimentų bei poliariškumo modeliai iš esmės nesiskiria, kadangi abiejų modelių rezultatai rodo, jog informacinių pranešimų įtaka akcijų kainai nėra statistiškai reikšminga. Gauto modelio AIC reikšmė lygi 2083,33, RMSE reikšmė lygi 0,959776. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.5.1. pav.), siekiant nustatyti ar paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį.



**6.2.5.1. pav.** „Johnson & Johnson“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat yra atliekamas Granger‘io priežastingumo testas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,1407 ir yra priimama nulinė hipotezė apie Granger‘io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo –

šio testo Fišerio statistika lygi 0,9273. Taigi galima teigti, jog informaciniai pranešimai neturi įtakos akcijų kainai „Johnson & Johnson“ kompanijos atveju.

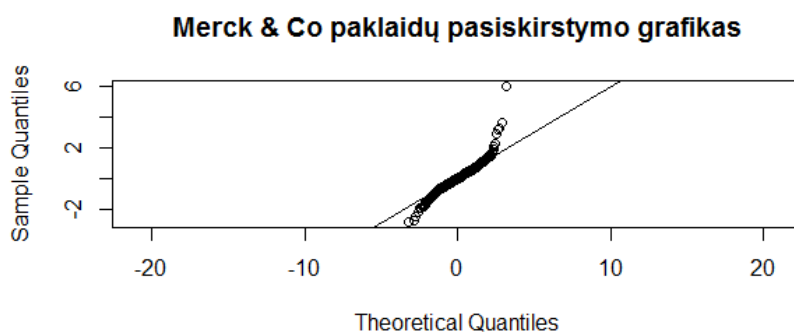
### 6.2.6. „Merck & Co“ kompanija

„Merck & Co“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.6.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.6.1. lentelė.** Poliariškumo modelio „Merck & Co“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	-0,091626	0,8692
Polarity_lag1	0,402392	0,5531
Polarity_lag2	0,637681	0,3473
Polarity_lag3	-0,360226	0,5173

Remiantis lentelėje pateiktais rezultatais, matoma, jog statistiškai reikšmingų kintamųjų nėra. Šios kompanijos sentimentų bei poliariškumo modeliai iš esmės nesiskiria, kadangi abiejų modelių rezultatai rodo, jog informacinių pranešimų įtaka akcijų kainai nėra statistiškai reikšminga. Gauto modelio AIC reikšmė lygi 1685,69, RMSE reikšmė lygi 0,7367955. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.6.1. pav.), siekiant nustatyti modelio paklaidų pasiskirstymą.



**6.2.6.1. pav.** „Merck & Co“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat yra atliekamas Granger‘io priežastingumo testas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,6224 ir yra priimama nulinė hipotezė apie Granger‘io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,3582. Taigi galima teigti, jog informaciniai pranešimai neturi įtakos akcijų kainai „Merck & Co“ kompanijos atveju.

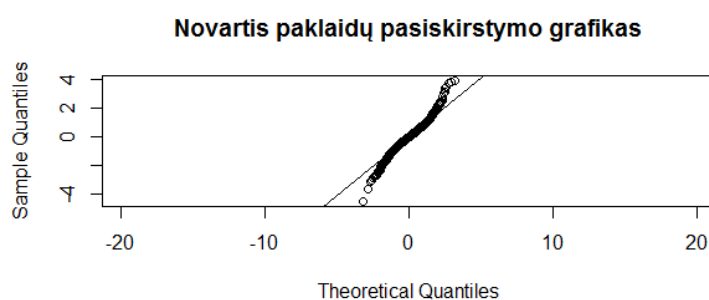
### 6.2.7. „Novartis“ kompanija

„Novartis“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.7.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.7.1. lentelė.** Poliariškumo modelio „Novartis“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	0,21406	0,73757
Polarity_lag1	-0,41924	0,58279
Polarity_lag2	-1,26694	0,09691
Polarity_lag3	-0,43468	0,49625

Remiantis lentelėje pateiktais rezultatais, matoma, jog poliariškumo dviejų dienų vėluojanti laiko eilutė yra statistiškai reikšminga esant 90 proc. pasiklovimo lygmeniui. Tai reiškia, jog pasirodžiusio straipsnio poveikis akcijos kainai pastebimas praėjus dviem dienoms nuo straipsnio pasirodymo. Tačiau kiti kintamieji yra statistiškai nereikšmingi. Todėl verta pastebėti, jog „Novartis“ kompanijos atveju, informacinių pranešimų poveikis skiriasi priklausomai nuo to, kaip yra vertinami straipsniai – įprastais sentimentais ar poliariškumu. Gauta poliariškumo modelio AIC reikšmė lygi 2120, RMSE reikšmė lygi 0,9834643. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.7.1. pav.).



**6.2.7.1. pav.** „Novartis“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat yra atliekamas Granger‘io priežastingumo testas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,7808 ir yra priimama nulinė hipotezė apie Granger‘io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,2067. Taigi galima teigti, jog 2 dienų vėluojanti poliariškumo laiko eilutė yra statistikai reikšminga, tačiau priežastingumas tarp informacinių pranešimų ir akcijų kainos nenustatytas.

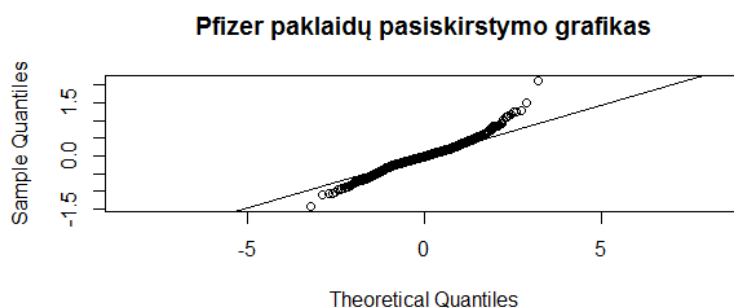
### 6.2.8. „Pfizer“ kompanija

„Pfizer“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.8.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.8.1. lentelė.** Poliariškumo modelio „Pfizer“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	0,620013	0,01247
Polarity_lag1	0,505925	0,10173
Polarity_lag2	0,137692	0,65604
Polarity_lag3	0,041506	0,86716

Remiantis lentelėje pateiktais rezultatais, matoma, jog poliariškumo laiko eilutė yra statistiškai reikšminga esant 95 proc. pasikliautinumo lygiui. Tačiau likę kintamieji nėra statistiškai reikšmingi. Gauto modelio AIC reikšmė lygi 662,57, RMSE reikšmė lygi 0,3731735. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.8.1. pav.), siekiant nustatyti modelio paklaidų pasiskirstymą.



**6.2.8.1. pav.** „Pfizer“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat yra atliekamas Granger‘io priežastingumo testas. Tačiau remiantis gautais šio testo rezultatais, nustatyta, jog priežastinio ryšio tarp poliariškumo laiko eilutės ir akcijų kainos nėra, kadangi Fišerio statistikos reikšmė lygi 0,6112 ir yra priimama nulinė hipotezė apie Granger‘io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,6115. Todėl galima daryti išvadą, jog esama informacinių naujienų, vertinamų poliariškumu, poveikio „Pfizer“ kompanijos akcijų kainai, tačiau vertinant poliariškumo kartu su pavēlintomis reikšmėmis priežastingumo ryšį su akcijų kaina, nenustatyta, jog toks ryšys egzistuoja.

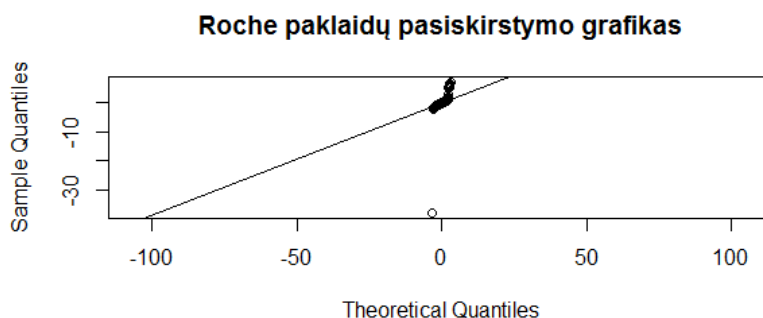
### 6.2.9. „Roche“ kompanija

„Roche“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (5, 2, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.9.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.9.1. lentelė.** Poliariškumo modelio „Roche“ kompanijai rezultatai

Kintamasis	Įvertis (Estimate)	P reikšmė
ar1	-0,829103	$2,2 * 10^{-16}$
ar2	-0,620447	$2,2 * 10^{-16}$
ar3	-0,498334	$2,2 * 10^{-16}$
ar4	-0,356285	$2,86 * 10^{-12}$
ar5	-0,191003	$9,576 * 10^{-5}$
Polarity	0,29952	0,7589
Polarity_lag1	0,180259	0,8785
Polarity_lag2	0,337565	0,7744
Polarity_lag3	0,047455	0,961

Remiantis lentelėje pateiktais rezultatais, matoma, jog statistiškai reikšmingi yra tik autoregresiją modelyje atspindintys kintamieji. Taigi galima daryti išvadą, jog poliariškumu vertinami straipsniai neturi poveikio „Roche“ kompanijos akcijų kainoms. Gauto poliariškumo modelio AIC reikšmė lygi 2813,37, RMSE reikšmė lygi 1,550867. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.9.1. pav.).



**6.2.9.1. pav.** „Roche“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat yra atliekamas Granger'io priežastingumo testas. Remiantis gautais šio testo rezultatais, nustatyta, jog nėra priežastingumo tarp poliariškumo laiko eilutės ir akcijų kainos, kadangi Fišerio statistikos reikšmė lygi 0,6691 ir yra priimama nulinė hipotezė apie Granger'io priežastingumo

nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,9732.

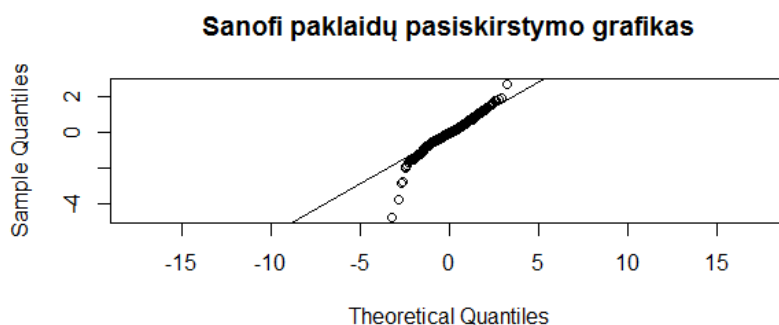
#### 6.2.10. „Sanofi“ kompanija

„Sanofi“ informacinių pranešimų, sudarytų pagal poliariškumą, ir akcijos kainai sudarytas ARIMA modelis, kurio parametrai pagal „auto.arima“ funkciją yra (0, 1, 0). Pagal šiuos parametrus sudarytas modelis laiko eilutei su 4 išoriniais kintamaisiais. Toliau lentelėje (6.2.10.1. lent.) pateikiami gauti modeliavimo rezultatai.

**6.2.10.1. lentelė. Poliariškumo modelio „Sanofi“ kompanijai rezultatai**

Kintamasis	Įvertis (Estimate)	P reikšmė
Polarity	-0,424027	0,4077
Polarity_lag1	-0,308738	0,6029
Polarity_lag2	-0,066534	0,9109
Polarity_lag3	0,118201	0,8183

Remiantis lentelėje pateiktais rezultatais, matoma, jog nėra statistiškai reikšmingų kintamųjų, todėl „Sanofi“ kompanijos atveju taip pat pastebima, jog informaciniai pranešimai, vertinami poliariškumu, neturi įtakos kompanijos akcijų kainai. Sudaryto ARIMA modelio AIC reikšmė lygi 1568,9, RMSE reikšmė lygi 0,6817459. Taip pat pateikiamas modelio paklaidų pasiskirstymo grafikas (6.2.10.1. pav.), siekiant nustatyti modelio paklaidų pasiskirstymą.



**6.2.10.1. pav.** „Sanofi“ poliariškumo modelio paklaidų pasiskirstymo grafikas

Taigi, pagal pateiktą paveikslą matoma, jog paklaidos pasiskirstę pagal normalųjį skirstinį. Taip pat yra atliekamas Granger‘io priežastingumo testas. Remiantis gautais šio testo rezultatais, nustatyta, jog priežastinio ryšio tarp poliariškumo laiko eilutės ir akcijų kainos nėra, kadangi Fišerio statistikos reikšmė lygi 0,3732 ir yra priimama nulinė hipotezė apie Granger‘io priežastingumo nebuvimą. Taip pat nustatyta, jog akcijų kaina neveikia informacinių pranešimų poliariškumo – šio testo Fišerio statistika lygi 0,7042. Taigi ir šiuo atveju priimama nulinė

Granger'io testo hipotezė. Todėl galima daryti išvadą, jog informacinių naujienų, vertinamų poliariškumu, poveikio „Sanofi“ kompanijos akcijų kainai nėra.

Taigi, atlikus informacinių pranešimų poveikio akcijų kainoms tyrimą, nustatyta, jog vertinant informacinių pranešimų turinio pozityvumą pagal sentimentus, pastebėta, jog „AstraZeneca“ ir „Bayer“ kompanijoms naujienų poveikis yra statistiškai reikšmingas. Taip pat nustatyta, jog šių kompanijų akcijų kainoms egzistuoja sentimentų laiko eilutės 1 diena pavėlintų reikšmių poveikis, t. y. praėjus vienai dienai nuo straipsnio pasirodymo, vis dar pastebimas poveikis akcijų kainai. Tačiau informacinių pranešimų pagal sentimentus poveikio likusių kompanijų akcijų kainoms nebuvo nustatyta. Aktualu pastebėti, jog Granger'io priežastingumo testo rezultatai beveik visais atvejais indikuoja, jog priežastinis ryšys tarp informacinių pranešimų bei akcijų kainų neegzistuoja.

Tyrimo, kai informaciniai pranešimai buvo vertinami pagal žodžių poliariškumą, atveju tik „AstraZeneca“ kompanijos atveju su 90 proc. pasiklivimo lygmeniu buvo nustatytas statistinis poliariškumo laiko eilutės reikšmingumas. Visoms kitoms kompanijoms, informacinių pranešimų poveikio vertinimui sudarytos laiko eilutės iš poliariškumo įverčių nėra statistiškai reikšmingi kintamieji. Todėl galima teigti, jog naujienų poveikio akcijų kainoms nėra.

Taip pat verta pastebėti, jog beveik visų kompanijų (išskyrus „Roche“ kompaniją vertinant informacinius pranešimus pagal poliariškumą) ARIMA modelių parametrai, pasiūlyti pagal „auto.arima“ funkciją yra (0, 1, 0). Todėl galima daryti išvadą, jog akcijų kainos kitimas priklauso atsitiktiniam klaidžiojimui, t. y. remiantis praeities stebiniais negalima nusakyti ateities.

## IŠVADOS

Remiantis atlikta literatūros analize, nustatyta, jog akcijų kainos analizei dažniausiai taikoma techninė bei laiko eilučių analizė. Tačiau nemaža dalis autorių tirdami akcijų kainas nustatė, jog techninė analizė nėra efektyvi. Laiko eilučių analizės principų taikymas analizuojant akcijų kainas yra sėkmingas, bet verta pastebėti, jog tokiu būdu analizuojama tik tiesinė akcijų kainos priklausomybė nuo praeities kainų. Todėl vis dažniau akcijų kainos tyrimui yra taikomi sistemų mokymosi ir duomenų tyrybos metodai, kuriais galima nustatyti netiesinę akcijų kainos priklausomybę. Taip pat atliekant literatūros analizę pastebėta, jog galimas tiesinę ir netiesinę priklausomybę vertinančių modelių derinimas kuriant hibridinius modelius.

Pastebėta, jog pastarąjį dešimtmetį aktuali akcijų kainos analizės kryptis yra tekstinės informacijos analizė, kurią galima atlikti taikant sistemų mokymosi ir duomenų tyrybos metodus. Tekstinės informacijos turinys gali būti vertinamas sentimentais – teigiamais arba neigiamais. Taip pat galima nustatyti teksto poliariškumą, kuriuo įvertinamas pozityvumo ar negatyvumo lygmuo pasirenkamoje skalėje. Atliekant informacinių pranešimų poveikio didžiųjų farmacijos kompanijų akcijoms tyrimą, informaciniai pranešimai buvo vertinami kaip teigiami ar neigiami signalai, kurie potencialiai gali veikti akcijos kainą, t. y. teigiama naujiena vertinama kaip pozityvus signalas, kuris gali turėti įtakos akcijų kainos augimui, neigiama naujiena gali veikti akcijų kainos kritimą. Tyrime informacinių pranešimų pozityvumas ir negatyvumas buvo vertinami pagal sentimentus – priskiriant reikšmes 1 ir -1 atitinkamai jei pranešime yra teigiamas ar neigiamas žodis. Teigiami ir neigiami žodžiai buvo vertinami pagal sudarytus pozityvių ir negatyvių žodžių sąrašus. Informacinių pranešimų pozityvumas taip pat buvo vertinamas pagal sudarytą Python programinei kalbai pritaikytą „TextBlob“ teksto analizės paketą. Šiuo atveju buvo vertinamas straipsnio poliariškumas pozityvius ir negatyvius žodžius vertinant įverčiais, esančiais intervale [-1; 1].

Remiantis gautais tyrimo rezultatais nustatyta, jog „AstraZeneca“ bei „Bayer“ kompanijoms informaciniai signalų, vertinamų pagal sentimentus, poveikis yra statistiškai reikšmingas. Taip pat nustatyta, jog šių kompanijų atveju informacinių signalų poveikis akcijų kainai pastebimas ir praėjus vienai dienai nuo naujienos pasirodymo spaudoje. Tačiau vertinant informacinius signalus poliariškumu, buvo nustatyta, jog statistiškai reikšmingo informacinių pranešimų bei pavėlintų jų reikšmių poveikio akcijų kainoms nėra.

Vis dėl to, aktualu pastebėti, jog gauti rezultatai „AstraZeneca“ bei „Bayer“ atveju priklausomai nuo informacinių pranešimų vertinimo metodo skiriasi. „AstraZeneca“ atveju informacinių pranešimų laiko eilutės (pagal sentimentus ir pagal poliariškumą) poveikis pastebimas abiem atvejais, tačiau skiriasi pasiklovimo lygmuo, kuriam esant kintamieji yra statistiškai reikšmingi. Tačiau „Bayer“ kompanijos atveju gauti rezultatai priklausomai nuo



informacinių pranešimų vertinimo yra priešingi: sentimentų laiko eilutė bei vienos dienos pavėlinta sentimentų eilutė yra statistiškai reikšminga, tačiau poliariškumo laiko eilutė nėra reikšminga. Todėl šiuo atveju galima teigti, jog informacinių pranešimų poveikis priklauso nuo naujienų vertinimo metodo, t. y. priklausomai nuo to, kaip jautriai yra vertinamas pranešimo pozityvumas ir negatyvumas galima spręsti apie galimą informacinių pranešimų poveikį akcijos kainai.

Taip pat aktualu pastebėti, jog ARIMA modelių parametrai, pasiūlyti pagal „auto.arima“ funkciją daugeliui kompanijų yra  $(0, 1, 0)$ . Pirmuoju parametru nurodoma reikalinga autoregresijos dalis, antruoju nurodoma ar reikalingas laiko eilutės integravimas, trečiuoju nurodoma slenkančių vidurkių ARIMA modelio dalis. Todėl remiantis automatinės funkcijos siūlomų parametru reikšmėmis, galima pastebėti, jog reikalingas integravimas, tačiau modeliuose nepastebima nei autokoreliacijos, nei slenkančių vidurkių. Literatūroje tokios ARIMA modelio reikšmės turi atskirą kategoriją – tokie modelio parametrai išskiriami kaip atsitiktinio klaidžiojimo modelis. Tai reiškia, jog remiantis praeities stebiniais negalima nusakyti ateities. Dėl šios priežasties galima teigti, jog tiriamų kompanijų akcijų kainos kitimas priklauso atsitiktiniam stochastiniam procesui, taigi šių kompanijų akcijų kainos prognozavimas naudojant tiesinę priklausomybę analizuojančius metodus nėra galimas.

## LITERATŪROS SĄRAŠAS

1. KELLER, C. and M. SIEGRIST. Investing in stocks: The influence of financial risk attitude and values-related money and stock market attitudes. *Journal of Economic Psychology* [interaktyvus]. Amsterdam: Elsevier, 2006, (27)2, 285-303 [žiūrėta 2017-03-04]. Prieiga per: Science Direct. [http://ac.els-cdn.com/S016748700500084X/1-s2.0-S016748700500084X-main.pdf?\\_tid=9001ae68-25bc-11e7-85a5-00000aab0f26&acdnat=1492687925\\_904b9286fbfeb62ec12c650b71a79a5a](http://ac.els-cdn.com/S016748700500084X/1-s2.0-S016748700500084X-main.pdf?_tid=9001ae68-25bc-11e7-85a5-00000aab0f26&acdnat=1492687925_904b9286fbfeb62ec12c650b71a79a5a)
2. KUMAR, Rajesh. Risk and Return. Valuation Theories and Concepts [interaktyvus] 2016. Prieiga per: doi: <https://doi.org/10.1016/B978-0-12-802303-7.00002-4>
3. FAMA, Eugene. The behavior of Stock-Market Prices. *Journal of Business* [interaktyvus]. 1965, (38)1, 34-105 [žiūrėta 2017-03-05]. ISSN Prieiga per: JSTOR. <http://www.e-m-h.org/Fama65.pdf>
4. ELTON, Edwin, Martin GRUBER, Stephen BROW and William GOETZMANN. Modern Portfolio Theory and Investment Analysis. [http://www.academia.edu/20386504/Modern Portfolio Theory and Investment Analysis Edwin J. Elton Martin J. Gruber Stephen J. Brow](http://www.academia.edu/20386504/Modern_Portfolio_Theory_and_Investment_Analysis_Edwin_J._Elton_Martin_J._Gruber_Stephen_J._Brow)
5. MAJOUJI, R. The financial markets context [interaktyvus] 2017. [žiūrėta 2017-03-11]. Prieiga per: <http://www.open.edu/openlearn/money-management/money/accounting-and-finance/the-financial-markets-context/content-section-3>
6. MALKIEL, B. Efficient Market Hypothesis. In: *Finance*. Eatwell J., M. Milgate, P. Newman: The Macmillan Press Limited, 1994, pp 127-134 ISBN 978-0-333-49535-3
7. TIMMERMAN, A. and C. GRANGER. Efficient market hypothesis and forecasting. *International Journal of Forecasting* [interaktyvus]. Amsterdam: Elsevier, 2004, (20)1, 15-27 [žiūrėta 2017-03-06] ISSN Prieiga per: Science Direct. <http://www.sciencedirect.com/science/article/pii/S0169207003000128>
8. MALKIEL Burton. The Efficient Market Hypothesis and its critics. *Journal of Economic Perspectives* [interaktyvus] 2003. (17)1, 59-82 [žiūrėta 2017-04-02]. Prieiga per: doi: <http://doi.org/10.1257/089533003321164958>
9. HUANG, Chenn-Jung, Dian-Xiu YANG and Yi-Ta CHUANG. Application of wrapper approach and composite classifier to the stock trend prediction. *Expert Systems with Applications* [interaktyvus]. Amsterdam: Elsevier, 2008, (34)4, 2870-2878 [žiūrėta 2017-04-03] Prieiga per: doi: <https://doi.org/10.1016/j.eswa.2007.05.035>
10. ADAM Anokye and George TWENEBOAH. Macroeconomic Factors and Stock Market Movement: Evidence from Ghana. *SSRN Electronic Journal* [interaktyvus] 2008. Prieiga per: doi: <http://dx.doi.org/10.2139/ssrn.1289842>
11. MUZAMMAL I., S. MUZAMMIL and A. HUSSAIN. Macroeconomic Factors do influencing Stock Price: A Case Study on Karachi Stock Exchange. *Journal of Economics and Sustainable Development* [interaktyvus]. 2014, (5)7 [žiūrėta 2017-03-07] ISSN 2222-2855 (Online). Prieiga per: IISTE. <http://iiste.org/Journals/index.php/JEDS/article/viewFile/12581/12899>
12. JARENO F. And N. NEGRUT. US Stock Market and Macroeconomic Factors. *The Journal of Applied Business Research* [interaktyvus]. 2016, (32)1 [žiūrėta 2017-03-07]. Prieiga per: <https://previa.uclm.es/profesorado/fjareno/DOCS/9541-36008-2-PB.pdf>
13. CAPORALE, Guglielmo, Fabio SPAGNOLO and Nicola SPAGNOLO. Macro news and stock returns in the Euro area: A VAR-GARCH-in-mean analysis. *International Review of Financial Analysis* [interaktyvus]. 2016, (45), 180-188 [žiūrėta 2017-04-03]. Prieiga per: doi: <https://doi.org/10.1016/j.irfa.2016.03.016>
14. NARAYAN, Paresh and Deepa BANNIGIDADMATH. Does Financial News Predict Stock Returns? New Evidence from Islamic and Non-Islamic Stocks. *Pacific-Basin Finance Journal* [interaktyvus]. 2017, (42), 24-45 [žiūrėta 2017-04-04]. Prieiga per: doi: <https://doi.org/10.1016/j.pacfin.2015.12.009>
15. WALKER, Clive. The direction of media influence: Real-estate news and the stock market. *Journal of Behavioral and Experimental Finance* [interaktyvus]. 2016, (10), 20-31 [žiūrėta 2017-04-06]. Prieiga per: doi: <https://doi.org/10.1016/j.jbef.2016.02.001>

16. SHYNKEVICH, Yauheniya ir kt. Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning. *Decision Support Systems* [interaktyvus]. 2016, (85), 74-83 [žiūrėta 2017-04-07]. Prieiga per: doi: <https://doi.org/10.1016/j.dss.2016.03.001>
17. LI, Xiaodong ir kt. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems* [interaktyvus]. 2014, (69), 14-23 [žiūrėta 2017-04-07]. Prieiga per: doi: <https://doi.org/10.1016/j.knosys.2014.04.022>
18. SUN, Andrew, Michael LACHANSKI and Frank FABOZZI. Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. *International Review of Financial Analysis* [interaktyvus]. 2016, (48), 272-281 [žiūrėta 2017-04-07]. Prieiga per: doi: <https://doi.org/10.1016/j.irfa.2016.10.009>
19. OLIVEIRA, Nuno, Paulo CORTEZ and Nelson AREAL. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems wit Applications* [interaktyvus]. 2017, (73), 125-144 [žiūrėta 2017-04-08]. Prieiga per: doi: <https://doi.org/10.1016/j.eswa.2016.12.036>
20. Structured and Unstructured Data: What is it? [interaktyvus]. [žiūrėta 2017-04-10] Prieiga per: <http://shepasoftware.com/blog/structured-and-unstructured-data-what-is-it/>
21. MALCOLM Baker and Jeffrey WUGLER. Investor Sentiment in the Stock Market. *Journal of Economic Perspectives* [interaktyvus]. 2007, (21)2, 129-151 [žiūrėta 2017-04-10]. Prieiga per: doi: <http://doi.org/10.3386/w13189>
22. BROWN, Gregory and Michael CLIFF. Investor sentiment and the near-term stock market. *Journal of Empirical Finance* [interaktyvus]. 2004, (11)1, 1-27 [žiūrėta 2017-04-07]. Prieiga per: doi: <https://doi.org/10.1016/j.jempfin.2002.12.001>
23. CHANG, Amy, Shaokun YU, Alan REINSTEIN and Natalie CHURYK. An overview of investor sentiment in stock market. [interaktyvus] 2004 [žiūrėta 2017-04-11]. Prieiga per: <http://www.wiu.edu/cbt/jcibi/documents/NAASFeb2016/SpecialNAASIssueFeb2016-InvestorSentiment.pdf>
24. Dictionary of Financial Terms [interaktyvus]. 2008 [žiūrėta 2017-03-09] Prieiga per: <http://financial-dictionary.thefreedictionary.com/bear+market>
25. BEAL Vangie. Unstructured data. [interaktyvus] [žiūrėta 2017-03-04] Prieiga per: [http://www.webopedia.com/TERM/U/unstructured\\_data.html](http://www.webopedia.com/TERM/U/unstructured_data.html)
26. TANWAR, Mona, Reena DUGGAL and Sunil KHATRI. Unravelling unstructured data: A wealth of information in big data [interaktyvus] 2015 [žiūrėta 2017-04-12]. Prieiga per: doi: <http://doi.org/10.1109/ICRITO.2015.7359270>
27. KHAIDEM L., S. SAHA and S. DEY. Predicting the direction of stock market prices using random forest. [interaktyvus] 2016, [žiūrėta 2017-04-16]. Prieiga per: <https://arxiv.org/pdf/1605.00003.pdf>
28. BALLINGS Michel, Dirk Van den POEL, Nathalie HESPEELS and Ruben GRYP. Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications* [interaktyvus]. 2015, (42)20, 7046-7056 [žiūrėta 2017-03-06] Prieiga per: doi: <https://doi.org/10.1016/j.eswa.2015.05.013>
29. HELLSTRÖM T. and K. HOMSTRÖM. Predictable Patterns in Stock Returns. [interaktyvus]. 1998 [žiūrėta 2017-03-07] Ima-TOM-1997-09 Prieiga per: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=F0B895A2C3EEEDDA4766C66392E7C7EC?doi=10.1.1.50.3351&rep=rep1&type=pdf>
30. EDWARDS R., J. MAGEE and W.H.C. BASSETTI. Technical Analysis of Stock Trends. Ninth Edition. Boca Raton: CRC Press, 2007. ISBN 13:978-1-4200-1314-6. [https://books.google.lt/books?hl=lt&lr=&id=wklriRw9a1oC&oi=fnd&pg=PP1&dq=technical+stock+analysis&ots=MVZM0MyQ61&sig=KFtKLuQZNcEWnphtAIFfut6p2a0&redir\\_esc=y#v=onepage&q=technical%20stock%20analysis&f=false](https://books.google.lt/books?hl=lt&lr=&id=wklriRw9a1oC&oi=fnd&pg=PP1&dq=technical+stock+analysis&ots=MVZM0MyQ61&sig=KFtKLuQZNcEWnphtAIFfut6p2a0&redir_esc=y#v=onepage&q=technical%20stock%20analysis&f=false)
31. HUDSON, Robert, Michael DEMPSEY and Kevin KEASEY. A note on the weak form efficiency of capital markets: The application of simple technical trading rules to UK stock prices - 1935 to 1994. *Journal of Banking*

- & *Finance* [interaktyvus]. 1996, (20)6, 1121-1132, [žiūrėta 2017-04-12]. Prieiga per: doi: [https://doi.org/10.1016/0378-4266\(95\)00043-7](https://doi.org/10.1016/0378-4266(95)00043-7)
32. JENSEN Michael and George BENINGTON G. Random Walks and Technical Theories: Some Additional Evidence. *Journal of Finance* [interaktyvus]. 2002, (25)2 [žiūrėta 2016-03-09] Prieiga per: doi: <http://dx.doi.org/10.2139/ssrn.244160>
  33. CHONG, Terence And Wing-Kam NG. Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. *Applied Economic Letters* [interaktyvus]. 2008, (15)14, 1111-1114 [žiūrėta 2017-04-15]. Prieiga per: doi: <http://dx.doi.org/10.1080/13504850600993598>
  34. BROCK William, J. LAKONISHOK and B. LeBARON. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance* [interaktyvus]. 1992, (47)5, 1731-1764 [žiūrėta 2017-03-12] Prieiga per: <http://www.technicalanalysis.org.uk/moving-averages/BrLL92.pdf>
  35. MILLS T. Technical Analysis and the London Stock Exchange: Testing Trading Rules Using the FT30 [interaktyvus]. 1997 [žiūrėta 2017-03-12]. Prieiga per: <http://web.ist.utl.pt/adriano.simoese/tese/referencias/Papers%20-%20Adriano/Technical%20Analysis.pdf>
  36. FRANKEL J. and K. FROOT. Chartists, Fundamentalists and Trading in the Foreign Exchange Market. *The American Economic Review* [interaktyvus]. 1990, (80)2, 181-185 [žiūrėta 2017-03-13] Prieiga per: [http://web.pdx.edu/~ito/Frankel\\_Froot.pdf](http://web.pdx.edu/~ito/Frankel_Froot.pdf)
  37. SHILLER R. Stock Prices and Social Dynamics. Brookings Papers on Economic Activity [interaktyvus]. 1984, (2) [žiūrėta 2017-03-13]. Prieiga per: [https://www.brookings.edu/wp-content/uploads/1984/06/1984b\\_bpea\\_shiller\\_fischer\\_friedman.pdf](https://www.brookings.edu/wp-content/uploads/1984/06/1984b_bpea_shiller_fischer_friedman.pdf)
  38. STABINGIENĖ L. Ekonometrika [interaktyvus]. 2014 [žiūrėta 2017-03-14]. Prieiga per: <http://www.ilab.lt/stabingiene/index.html>
  39. TAY F., L. SHEN, L. CAO. *Ordinary Shares, Exotic Methods– Financial Forecasting Using Data Mining Techniques* [interaktyvus]. Singapore: World Scientific Publishing Co., 2003 [žiūrėta 2017-03-14] ISBN 981-238-075-2 Prieiga per: <https://books.google.lt/books?id=1MXUCgAAQBAJ&lpg=PR7&dq=tay%20ordinary%20shares%20exotic%20methods&lr&hl=lt&pg=PP1#v=onepage&q&f=false>
  40. RUSU V. and C. RUSU. Forecasting methods and stock market analysis [interaktyvus]. 2003 [žiūrėta 2017-03-15]. Prieiga per: [http://creative-mathematics.ubm.ro/issues/creative\\_2003\\_12\\_103\\_110.pdf](http://creative-mathematics.ubm.ro/issues/creative_2003_12_103_110.pdf)
  41. HYNDMAN J. and G. ATHANASOPOULOS. Forecasting: Principles and Practice [interaktyvus]. 2017 [žiūrėta 2017-03-16]. Prieiga per: <https://www.otexts.org/fpp2/>
  42. ZHONG Xiao And David ENKE. Forecasting daily stock market return using dimensionality reduction. *Expert Systems With Applications* [interaktyvus]. 2017, (67), 126-139 [žiūrėta 2017-03-17]. Prieiga per: doi: <http://dx.doi.org/10.1016/j.eswa.2016.09.027>
  43. PAI Ping-Feng and Chih-Sheng LIN. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* [interaktyvus]. 2005, (33)6, 497-505 [žiūrėta 2017-03-17]. Prieiga per: doi: <http://doi.org/10.1016/j.omega.2004.07.024>
  44. WANG Ju-Jie, Jian-Zhou WANG, Zhe-George ZHANG and Shu-Po GUO. Stock index forecasting based on a hybrid model. *Omega* [interaktyvus]. 2012, (40)6, 758-766 [žiūrėta 2017-03-19]. Prieiga per: doi: <http://doi.org/10.1016/j.omega.2011.07.008>
  45. LU Chi-Jie, Tian-Shyug LEE and Chih-Chou CHIU. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems* [interaktyvus] 2009, (47)2, 115-125 [žiūrėta 2017-03-21]. Prieiga per: doi: <http://doi.org/10.1016/j.dss.2009.02.001>
  46. HAN J., M. KAMBER and J. PEI. *Data Mining: Concepts and Techniques* [interaktyvus]. Amsterdam: Elsevier, 2012. [žiūrėta 2017-03-20] ISBN 978-0-12-381479-1 Prieiga per: <https://books.google.lt/books?id=pQws07tdpjoC&lpg=PP1&dq=data%20mining%20importance&lr&hl=lt&pg=PR6#v=onepage&q=data%20mining%20importance&f=false>

47. WITTEN I., E. FRANK, M. HALL and J. PAL. *Data Mining– Practical Machine Learning Tools and Techniques* [interaktyvus]. 2017 [žiūrėta 2017-03-25] ISBN 978-0-12-804291-5 Prieiga per: <https://books.google.lt/books?id=1SylCgAAQBAJ&lpg=PP1&dq=machine%20learning%20data%20mining%20tasks&lr&hl=lt&pg=PR4#v=onepage&q=machine%20learning%20data%20mining%20tasks&f=false>
48. KANTARDZIC M. *Data Mining – Concepts, Models, Methods and Algorithms* [interaktyvus]. 2011 [žiūrėta 2017-03-21] ISBN 978-0-470-89045-5 Prieiga per: <https://books.google.lt/books?id=ZZ7l6v0CvRMC&lpg=PA1&dq=data%20mining%20importance&lr&hl=lt&pg=PR4#v=onepage&q=data%20mining%20importance&f=false>
49. VARLAMIS Iraklis et al. Application of data mining techniques and data analysis methods to measure cancer morbidity and mortality data in a regional cancer registry: The case of the island of Crete, Greece. *Computer Methods and Programs in Biomedicine* [interaktyvus]. 2017, (145), 73-83 [žiūrėta 2017-03-28]. Prieiga per: doi: <http://doi.org/10.1016/j.cmpb.2017.04.011>
50. KAVAKIOTIS, Ioannis et al. Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal* [interaktyvus]. 2017, (15), 104-116 [žiūrėta 2017-05-02]. Prieiga per: doi: <https://doi.org/10.1016/j.csbj.2016.12.005>
51. ABURROUS Maher et al. Intelligent phishing detection system for e-banking using fuzzy data mining. *Expert Systems with Applications* [interaktyvus]. 2010, (37)12, 7913–7921 [žiūrėta 2017-03-24]. Prieiga per: doi: <http://doi.org/10.1016/j.eswa.2010.04.044>
52. MORO Sergio, Paulo CORTEZ and Paulo RITA. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications* [interaktyvus]. 2014, (42)3, 1314-1324 [žiūrėta 2017-03-22]. Prieiga per: doi: <http://doi.org/10.1016/j.eswa.2014.09.024>
53. CHARTE, Francisco et al. Comparative analysis of data mining and response surface methodology predictive models for enzymatic hydrolysis of pretreated olive tree biomass. *Computers & Chemical Engineering* [interaktyvus]. 2017(101)9, 23-30 [žiūrėta 2017-03-02]. Prieiga per: doi: <https://doi.org/10.1016/j.compchemeng.2017.02.008>
54. ADAMCZYK, Krzysztof et al. Recognition of culling reasons in Polish dairy cows using data mining methods. *Computers and Electronics in Agriculture* [interaktyvus]. 2016, (127), 26-37 [žiūrėta 2017-05-05]. Prieiga per: doi: <https://doi.org/10.1016/j.compag.2016.05.011>
55. ZHOU, Lina, Shimei PAN, Jianwu WANG and Athanasios VASILAKOS. Machine learning on big data: Opportunities and challenges. *Neurocomputing* [interaktyvus]. 2017, (237), 350-361 [žiūrėta 2017-04-19]. Prieiga per: doi: <https://doi.org/10.1016/j.neucom.2017.01.026>
56. JAYAWARDENA, Nirodha, Neda TODOROVA, Bin LI and Jen-Je SU. Forecasting stock volatility using after-hour information: Evidence from the Australian Stock Exchange. *Economic Modelling* [interaktyvus]. 2016, (52)B, 592-608 [žiūrėta 2017-03-03]. Prieiga per: doi: <https://doi.org/10.1016/j.econmod.2015.10.004>
57. SHEN, Wei, Xiaopen GUO, Chao WU and Desheng WU. Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. *Knowledge-Based Systems* [interaktyvus]. 2011, (24)3, 378-385 [žiūrėta 2017-03-05]. Prieiga per: doi: <https://doi.org/10.1016/j.knosys.2010.11.001>
58. JORDAN, Steven, Andrew VIVIAN and Mark WOHAR. Forecasting market returns: bagging or combining? *International Journal of Forecasting* [interaktyvus]. 2017, (33)1, 102-120 [žiūrėta 2017-04-22]. Prieiga per: doi: <https://doi.org/10.1016/j.ijforecast.2016.07.003>
59. PATEL, Jigar, Sahil SHAH, Priyank THAKKAR and K. KOTTECHA. Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications* [interaktyvus]. 2015, (42)4, 2162-2172 [žiūrėta 2017-04-11]. Prieiga per: doi: <https://doi.org/10.1016/j.eswa.2014.10.031>
60. LIAO, Shu-Hsien, Hsu-hui HO and Hui-wen LIN. Mining stock category association and cluster on Taiwan stock market. *Expert Systems with Applications* [interaktyvus]. 2008, (35)1-2, 19-29 [žiūrėta 2017-04-12]. Prieiga per: doi: <https://doi.org/10.1016/j.eswa.2007.06.001>
61. GEVA, Tomer and Jacob ZAHAVI. Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. *Decision Support Systems* [interaktyvus]. 2014, (57), 212-223 [žiūrėta 2017-04-03]. Prieiga per: doi: <https://doi.org/10.1016/j.dss.2013.09.013>



62. AGGARWAL, C. and C. ZHAI. Mining Text Data. New York: Springer Science, 2012. ISBN 978-1-4614-3222-7.
63. FELDMAN, R. et al. Knowledge Management: A Text Mining Approach. [interaktyvus] 1998, [žiūrėta 2017-04-02]. Prieiga per: [https://www.researchgate.net/publication/220795989\\_Knowledge\\_Management\\_A\\_Text\\_Mining\\_Approach](https://www.researchgate.net/publication/220795989_Knowledge_Management_A_Text_Mining_Approach)
64. NASSIRTOUSSI, Arman, Saeed AGHABOZORGI, Teh WAH and David NGO. Text mining for market prediction: A systematic review. *Expert Systems with Applications* [interaktyvus]. 2014, (41)16, 7653-7670 [žiūrėta 2017-03-27]. Prieiga per: doi: <https://doi.org/10.1016/j.eswa.2014.06.009>
65. HAGENAU, Michael, Michael LIEBMANN and Dirk NEUMANN. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems* [interaktyvus]. 2013, (55)3, 685-697 [žiūrėta 2017-03-22]. Prieiga per: doi: <https://doi.org/10.1016/j.dss.2013.02.006>
66. LI, Qing et al. The effect of news and public mood on stock movements. *Information Sciences* [interaktyvus]. 2014, (278)10, 826-840, [žiūrėta 2017-04-02]. Prieiga per: doi: <https://doi.org/10.1016/j.ins.2014.03.096>
67. LOSIEWICZ, Paul, Douglas OARD and Ronald KOSTOFF. Textual Data Mining to Support Science and Technology Management. *Journal of Intelligent Information Systems* [interaktyvus]. 2000, (15)2, 99-119, [žiūrėta 2017-04-19]. Prieiga per: doi: <http://doi.org/doi:10.1023/A:1008777222412>
68. STRAPPARAVA, C., A. VALITUTTI and O. STOCK. The affective weight of lexicon [interaktyvus]. 2006 [žiūrėta 2017-04-12]. Prieiga per: [http://www.lrec-conf.org/proceedings/lrec2006/pdf/186\\_pdf.pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/186_pdf.pdf)
69. NGUYEN, Thien, Kiyooki SHIRAI and Julien VELCIN. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* [interaktyvus]. 2015, (42)24, 9603-9611, [žiūrėta 2017-04-16]. Prieiga per: doi: <https://doi.org/10.1016/j.eswa.2015.07.052>
70. SETH, S. Safest Industries to Invest In [interaktyvus]. 2015 [žiūrėta 2017-04-18]. Prieiga per: <http://www.investopedia.com/articles/investing/052815/safest-industries-invest.asp>
71. BORZYKOWSKI, B. Health-care stocks—a great long-term market bet [interaktyvus]. 2015 [žiūrėta 2017-04-28]. Prieiga per: <http://www.cnbc.com/2015/06/02/health-care-stocks-a-great-long-term-market-bet.html>
72. ZUCCHI, K. Investing In The Healthcare Sector [interaktyvus]. [žiūrėta 2017-05-02]. Prieiga per: <http://www.investopedia.com/articles/stocks/08/investing-in-healthcare.asp>
73. ANAND, K. The big 'bull' market is in pharma sector, say experts; top bets [interaktyvus]. 2014 [žiūrėta 2017-04-23]. Prieiga per: <http://economictimes.indiatimes.com/markets/stocks/news/the-big-bull-market-is-in-pharma-sector-say-experts-top-bets/articleshow/45097682.cms>
74. DATTA, A. Top 10 pharmaceutical companies in 2016 [interaktyvus]. 2016 [žiūrėta 2017-04-19]. Prieiga per: <http://blog.proclinical.com/top-10-pharmaceutical-companies-2016>
75. Engineering statistics handbook [interaktyvus]. 2012 [žiūrėta 2017-04-28]. Prieiga per: <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm>
76. STAT 510 Applied Time Series Analysis [interaktyvus]. 2017 [žiūrėta 2017-04-12]. Prieiga per: <https://onlinecourses.science.psu.edu/stat510/node/67>
77. Segmentation of Twitter Timelines via Topic Modeling [interaktyvus]. 2015 [žiūrėta 2017-04-25]. Prieiga per: [http://alexperrier.github.io/jekyll/update/2015/09/16/segmentation\\_twitter\\_timelines\\_lda\\_vs\\_lsa.html](http://alexperrier.github.io/jekyll/update/2015/09/16/segmentation_twitter_timelines_lda_vs_lsa.html)

## **PRIEDAI**

## 1 PRIEDAS

### Straipsnių, teksto analizės raktiniams žodžiams atrinkti, sąrašas

- BRIEF-Omega Healthcare investors Q3 FFO per share \$0.80. Prieiga per: <http://www.reuters.com/article/idUSASC09EIF>
- BRIEF-Aetna and Banner Health announces a new JV health plan in Arizona. [žiūrėta 2017-03-03]. Prieiga per: <http://www.reuters.com/article/idUSASC09EG7>
- BRIEF-Novartis Bioventures reports 11.8 pct passive stake in Ra Pharmaceuticals - SEC filing. [žiūrėta 2017-03-03]. Prieiga per: <http://www.reuters.com/article/idUSFWN1D112K>
- BRIEF-Johnson & Johnson reports 6.4 pct passive stake in Tracon Pharmaceuticals. [žiūrėta 2017-03-03]. Prieiga per: <http://www.reuters.com/article/idUSFWN1D10ZX>
- BRIEF-Acadia Pharmaceuticals initiates phase II study of pimavanserin in Alzheimer's disease agitation. [žiūrėta 2017-03-03]. Prieiga per: <http://www.reuters.com/article/idUSFWN1D10VD>
- BRIEF-Catalyst Pharmaceuticals gets FDA special protocol assessment for phase 3 trial of lems treatment. [žiūrėta 2017-03-04]. Prieiga per: <http://www.reuters.com/article/idUSASC09ECY>
- BRIEF-Easton Pharma and Partner BMV Medica announce closing of first distribution agreement with Windsor Pharmaceuticals. [žiūrėta 2017-03-03]. Prieiga per: <http://www.reuters.com/article/idUSASC09ECD>
- BRIEF-Scintilla Pharma, a unit of Sorrento Therapeutics, to buy Semnur Pharma. [žiūrėta 2017-03-04]. Prieiga per: <http://www.reuters.com/article/idUSFWN1AW0NX>
- BRIEF-Ritter Pharmaceuticals completes enrollment for phase 2b/3 trial of RP-G28. [žiūrėta 2017-03-06]. Prieiga per: <http://www.reuters.com/article/idUSFWN1AW0D6>
- European stocks lifted to 7-week highs by pharma shares. [žiūrėta 2017-03-06]. Prieiga per: <http://www.reuters.com/article/europe-stocks-idUSL8N1AW107>
- BRIEF-Leica Biosystems and Merrimack Pharmaceuticals to collaborate. [žiūrėta 2017-03-09]. Prieiga per: <http://www.reuters.com/article/idUSFWN17H0SX>
- BRIEF-Omega Healthcare sets quarterly dividend of \$0.58/share. [žiūrėta 2017-03-08]. Prieiga per: <http://www.reuters.com/article/idUSASC08L8L>
- FDA expands use of Pfizer's best-selling pneumonia vaccine. [žiūrėta 2017-03-01]. Prieiga per: <http://www.reuters.com/article/us-pfizer-fda-idUSKCN0ZS2O4>
- Intercept Pharma's liver drug wins conditional EU approval. [žiūrėta 2017-03-02]. Prieiga per: <http://www.reuters.com/article/us-intercept-ema-idUSKBN1431E6>
- Novo aims to be ahead of pharma pack with Iran investment. [žiūrėta 2017-03-03]. <http://www.reuters.com/article/us-iran-pharmaceuticals-idUSKCN0UZ266>



## 2 PRIEDAS

### Programinis kodas, skirtas straipsnių gavimui iš Reuters naujienų portalo

```
import requests
from bs4 import BeautifulSoup
import re
import csv
import urllib
import itertools
import time

date=[]
with open('datos2.csv', 'r') as f:
    reader = csv.reader(f)
    datos = list(reader)
for i in datos:
    date.append(str(i).replace('[', '').replace('"', '').replace(']', ''))

words = []
with open('filter_words.csv', 'r') as f:
    reader = csv.reader(f)
    filter_words = list(reader)
for w in filter_words:
    words.append(str(w).replace('[', '').replace('"', '').replace(']', ''))

urls=[]
headlines=[]

for x in date:
    page = requests.get("http://www.reuters.com/resources/archive/us/{}.html".format(x))
    soup = BeautifulSoup(page.content, 'html.parser')
    for item in soup.find_all('div', class_='headlineMed'):
        if not item.find('video'):
            data = item.find('a')
            headlines.append(data.get_text())
            urls.append(data.attrs['href'])
            for d in range(0, len(urls)):
                with open("{}_{}.txt".format(x, d), "w") as f:
                    f.write(str(urls))
            print('Links for {} have been pulled'.format(x))

#get all links and headers for investigated dates, text of articles
info=[]
urls=[]
headlines=[]

for x in date:
    page = requests.get("http://www.reuters.com/resources/archive/us/{}.html".format(x))
    soup = BeautifulSoup(page.content, 'html.parser')
    for item in soup.find_all('div', class_='headlineMed'):
        if not item.find('video'):
            data = item.find('a')
            headlines.append(data.get_text())
            urls.append(data.attrs['href'])
            for d in range(0, len(urls)):
                info.append(x)
            print('Links for {} have been pulled'.format(x))

#Pulling text from articles
titles = []
def textual_data(urls):
    final = 0
    l = 0
    for u in urls:
        final += 1
        l += 1
        if l <= len(urls):
```

```

print(1)
try:
    page = requests.get("{}".format(u))
    soup = BeautifulSoup(page.content, 'html.parser')
    head = soup.find_all('title')[0]
    header = head.get_text()
    header = header[:-11].strip()
    if header in titles:
        continue
    else:
        titles.append(header)
        article_date = soup.find("meta",
{"property": "og:article:published_time"})['content'][0:10]
        for item in soup.find_all(id="article-text"):
            text_data = []
            for p in item.find_all('p'):
                for span in p.find_all('span'):
                    span.decompose()
                text_data.append(str(p).replace('<p>', '').replace("</p>", ""))
            text_words = " ".join(p.lower() for p in text_data).split()
            if set(w.lower() for w in words).intersection(text_words):
                try:
                    with open("{}_{}.txt".format(article_date, header), "w")
as f:
                    f.write(str(text_words))
                    print('{} article is being written -
{}'.format(article_date, header))
                except:
                    continue
            else:
                continue
        except:
            continue
        time.sleep(1)
    else:
        break

textual_data(urls)

```

### 3 PRIEDAS

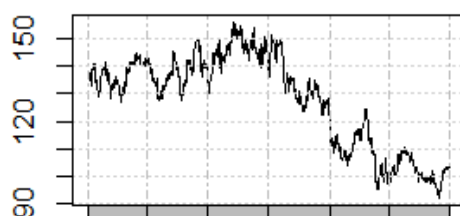
#### Tinkamiausių LDA modelių sudarytų temų žodžiai ir jų tikimybės

1 Tema		2 Tema		3 Tema	
Žodis	Žodžio tikimybė	Žodis	Žodžio tikimybė	Žodis	Žodžio tikimybė
tax	0,06231	drug	0,06067	bayer	0,09360
generic	0,05209	price	0,02872	animal	0,03168
pfizer	0,04706	share	0,01670	chemical	0,02860
mylan	0,03328	cost	0,01217	monsanto	0,01803
teva	0,03327	treatment	0,01107	crop	0,01642
allergan	0,02818	gilead	0,01106	seed	0,01518
deal	0,02646	pharmaceutical	0,01041	ag	0,01321
billion	0,02476	editing	0,00993	product	0,01252
pharmaceutical	0,02452	analyst	0,00991	farm	0,01170
drug	0,01479	hepatitis	0,00951	german	0,01150
drugmaker	0,01476	patient	0,00843	group	0,00942
industry	0,01223	pharmacy	0,00766	merck	0,00939
product	0,01194	new	0,00707	unit	0,00912
plc	0,01184	bengaluru	0,00695	farmer	0,00839
actavis	0,01102	market	0,00689	agriculture	0,00726
buy	0,01072	high	0,00610	market	0,00722
base	0,00947	year	0,00597	business	0,00708
business	0,00937	sovaldi	0,00562	materialscience	0,00702
israel	0,00825	benefit	0,00527	dekker	0,00700
acquisition	0,00814	stock	0,00522	pet	0,00663

## 4 PRIEDAS

### Grynujų akcijų kainų laiko eilutės

**Bayer**



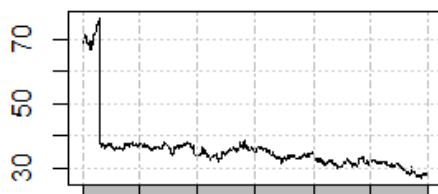
Sau 02 2014    Lie 01 2015    Grd 30 2016  
Laikotarpis

**Novartis**



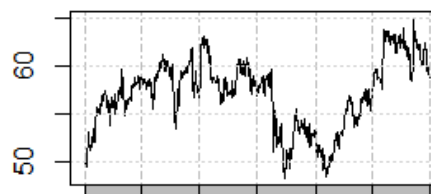
Sau 02 2014    Lie 01 2015    Grd 30 2016  
Laikotarpis

**Roche**



Sau 02 2014    Lie 01 2015    Grd 30 2016  
Laikotarpis

**Merck & Co**



Sau 02 2014    Lie 01 2015    Grd 30 2016  
Laikotarpis

**Sanofi**



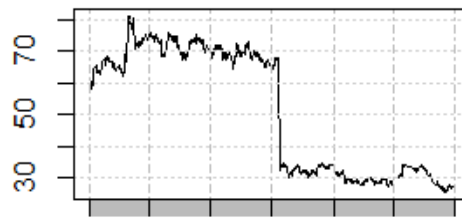
Sau 02 2014    Lie 01 2015    Grd 30 2016  
Laikotarpis

**Gilead**



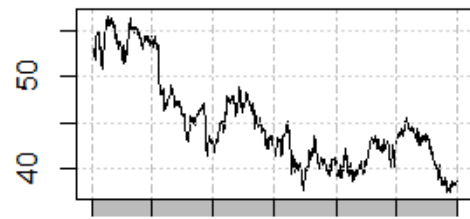
Sau 02 2014    Lie 01 2015    Grd 30 2016  
Laikotarpis

### AstraZeneca



Sau 02 2014   Lie 01 2015   Grd 30 2016  
Laikotarpis

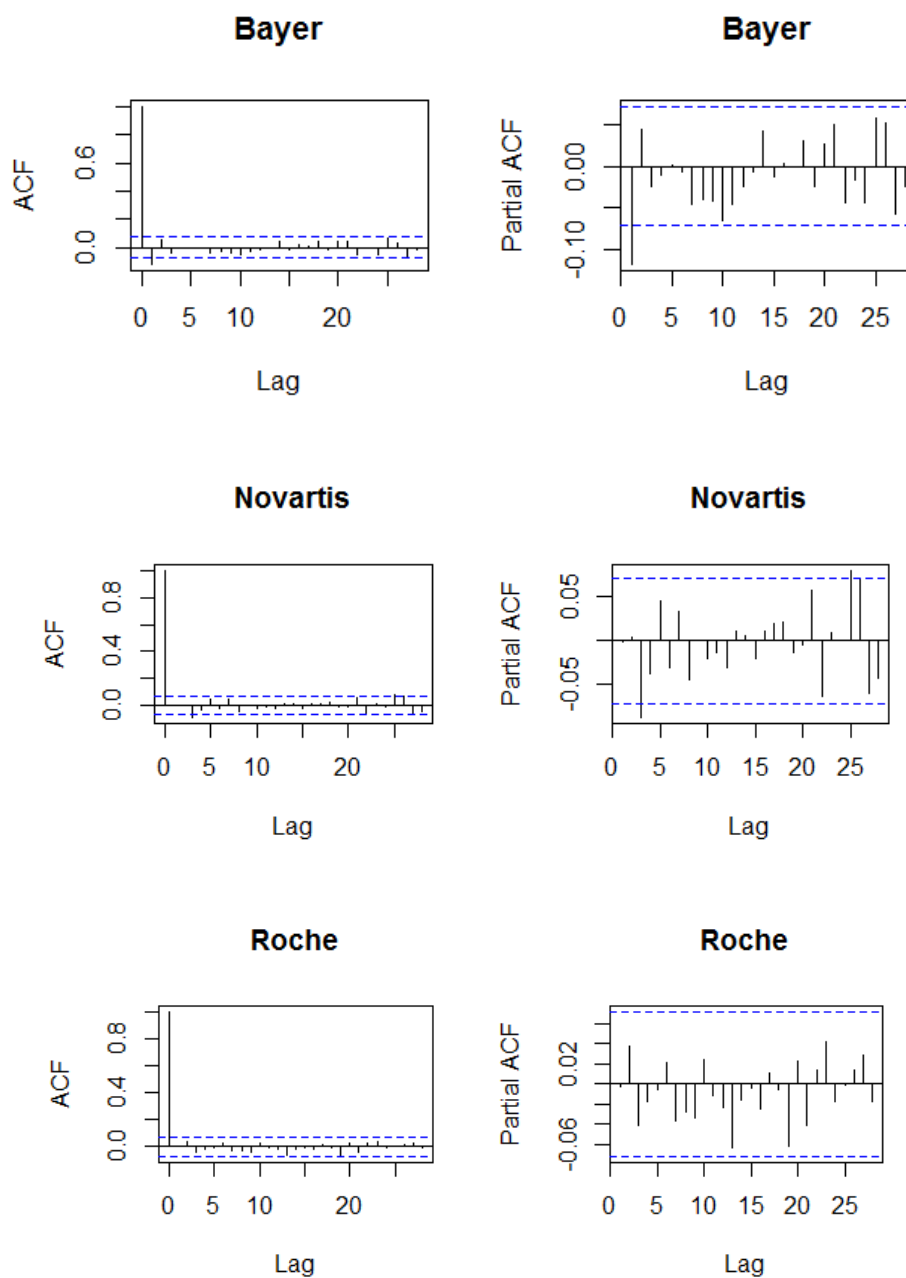
### GlaxoSmithKline



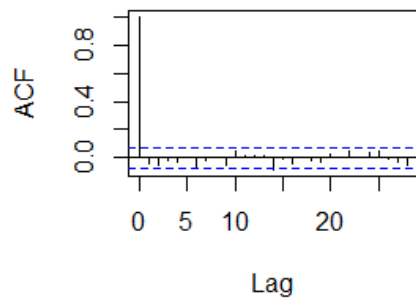
Sau 02 2014   Lie 01 2015   Grd 30 2016  
Laikotarpis

## 5 PRIEDAS

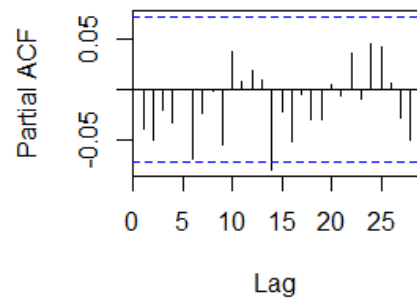
### Didžiųjų farmacijos kompanijų akcijų kainos laiko eilučių pilnosios ir dalinės autokoreliacijos grafikai



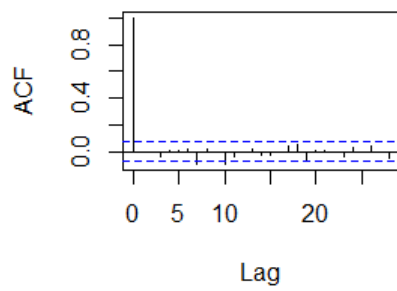
**Merck & Co**



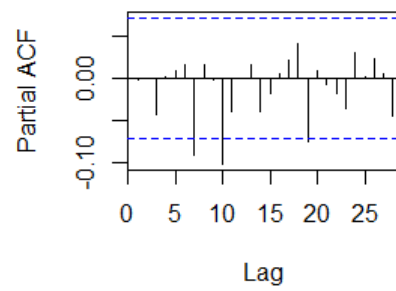
**Merck & Co**



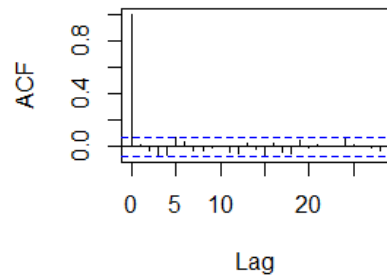
**Sanofi**



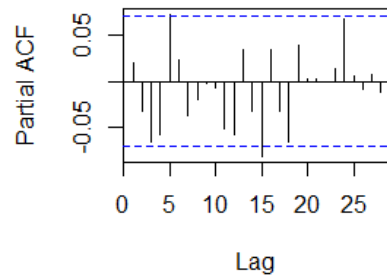
**Sanofi**



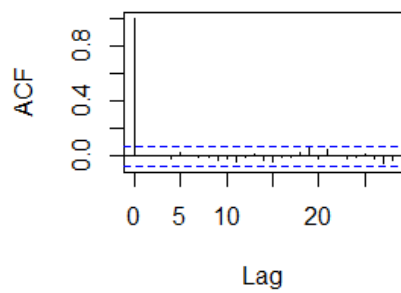
**Gilead**



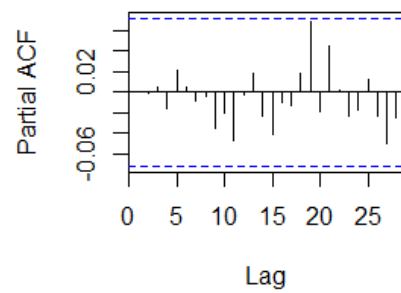
**Gilead**

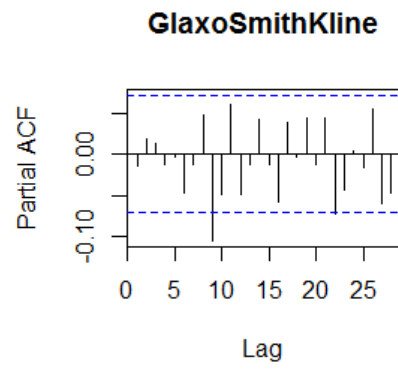
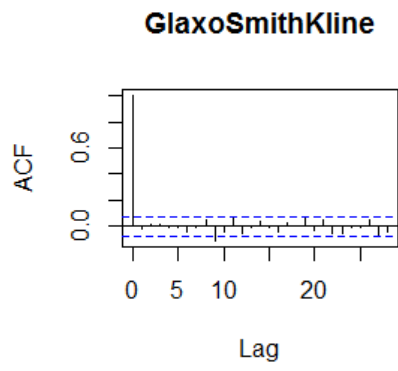


**AstraZeneca**



**AstraZeneca**







## 6 PRIEDAS

### Programinis kodas

#### Akcijų kainų duomenų gavimas

```
require(quantmod)
library(xlsx)

stockData <- new.env()
startDate = as.Date("2014-01-01")
endDate = as.Date("2016-12-31")

tickers <- c("JNJ", "BAYRY", "NVS", "PFE", "RHHBY", "MRK", "SNY", "GILD", "AZN", "GSK")
getSymbols(tickers, env = stockData, src = "yahoo", from = startDate, to = endDate)

JNJ <- Cl(stockData$JNJ)
BAYRY <- Cl(stockData$BAYRY)
NVS <- Cl(stockData$NVS)
PFE <- Cl(stockData$PFE)
RHHBY <- Cl(stockData$RHHBY)
MRK <- Cl(stockData$MRK)
SNY <- Cl(stockData$SNY)
GILD <- Cl(stockData$GILD)
AZN <- Cl(stockData$AZN)
GSK <- Cl(stockData$GSK)

#Akcijų uždarymo kainos įrašomos paeiliui į failus
write.xlsx(JNJ, "C:/Users/Karolina/Desktop/JNJ.xlsx")
```

#### Latentinio Dirichlet'o pasiskirstymo modelis

```
import time
import os
import re
import numpy as np
import nltk
import gensim
import string
import csv
import glob
import ast
import collections
import itertools
from gensim import utils

fnames = [f for f in glob.glob("*.txt") if len(f.split("__")) > 1]

def cleaner(doc):
    from nltk.tokenize import RegexpTokenizer
    tokenizer = RegexpTokenizer(r'\w+')
    for f in fnames:
        d = f[0:10]
        s = f[11:-4]
        with open('{}_{}.txt'.format(d, s), "w") as text:
            with open('{}'.format(f)) as t:
                u = []
                y = t.read().replace('[', ' ').replace('"', '').replace(" ", "").split(',')
                for i in y:
                    if not re.search(r'\bhref', i):
                        if not re.search(r'\<a', i):
                            if not re.search(r'\d+', i):
                                if not re.search(r'\/\x7f+', i):
                                    u.append(i)
                u = str(tokenizer.tokenize(str(u)))
            text.write(u)
```

```

def load_doc(fname):
    with open(fname) as f:
        return ast.literal_eval(f.read())

raw_docs = [load_doc(f) for f in fnames]

[cleaner(f) for f in fnames]

from nltk.tokenize import RegexpTokenizer
tokenizer = RegexpTokenizer(r'\w+')

def nltk_stopwords():
    return set(nltk.corpus.stopwords.words('english'))

def prep_corpus(docs, additional_stopwords=set(), no_below=5, no_above=0.5):
    print('Building dictionary...')
    dictionary = gensim.corpora.Dictionary(docs)
    stopwords = nltk_stopwords().union(additional_stopwords)
    stopword_ids = map(dictionary.token2id.get, stopwords)
    dictionary.filter_tokens(stopword_ids)
    dictionary.compactify()
    dictionary.filter_extremes(no_below=no_below, no_above=no_above, keep_n=None)
    # dictionary.compactify()

    print('Building corpus...')
    corpus = [dictionary.doc2bow(doc) for doc in docs]

    return dictionary, corpus

from nltk.stem.wordnet import WordNetLemmatizer
from nltk.corpus import wordnet
from collections import defaultdict

lmtzr = WordNetLemmatizer()

pos_tag_map = defaultdict(
    lambda: wordnet.NOUN,
    {'NN': wordnet.NOUN, 'JJ': wordnet.ADJ, 'VB': wordnet.VERB, 'RB': wordnet.ADV})

%%time
tagged_docs = [nltk.pos_tag(d) for d in raw_docs]

lemmatized_docs = [[lmtzr.lemmatize(token, pos_tag_map[tag[:2]]) for token, tag in d] for d in
tagged_docs]

dictionary, corpus = prep_corpus(lemmatized_docs)

%%time
lda = gensim.models.ldamodel.LdaModel(corpus=corpus, id2word=dictionary, num_topics=50,
passes=20)

vis_data = gensimvis.prepare(lda, corpus, dictionary)
pyLDavis.display(vis_data)

```

## TF-IDF statistikos skaičiavimas

Programinis kodas, prieinamas adresu <https://gist.github.com/vineetrok/1391957>, pritaikytas turimiems duomenims

```

import glob
import math
import time
%%time

line=''
s=set()
file_list=glob.glob(r'/home/vagrant/synced_dir/magistras/Companies/Bayer/*.txt')
for file in file_list:

```

```

        tfile=open(file,"r")
        line=tfile.read()
        tfile.close()
        s=s.union(set(line.split(' ')))
s=sorted(s)

i=0
ct=0
tf_line=''
doc_counts=[]
for term in s:
    doc_counts.append(0)

    for fdoc in file_list:

        doc=open(fdoc)
        line=doc.read()
        doc.close()
        ct=line.count(str(term))
        tf_line+=str(ct)+','
        if (ct>0):
            doc_counts[i]+=1
    i+=1
    tf_line=tf_line.strip()+'\n'

idf=[]
weights=[]
total_docs=len(file_list)

i=0

for doc_count in doc_counts:
    idf.append(math.log(total_docs/doc_count))
    weights.append(idf[i]*doc_count)
    i+=1

final_line='TERM'+','
i=1
for f in file_list:
    final_line+='D'+str(i)+' '+','
    i+=1
final_line+=','+'IDF'+','+'TF-IDF\n'

tf_arr=tf_line.split('\n')

i=0
for term in s:
    final_line+=term+', '+tf_arr[i]+' '+str(round(idf[i],2))+','+' '+
'+str(round(weights[i],2))+','+' '\n'
    i+=1

fdoc="/home/vagrant/synced_dir/magistras/Companies/Bayer/tftable_Bayer.csv"
outfile=open(fdoc,"w")
outfile.write(final_line)
outfile.close()

print('DONE')

```

## Sentimentų vektoriaus sudarymas

```

#CREATING SENTIMENT VECTOR FROM ARTICLES
setwd("C:/Users/Karolina/Desktop/Comp/Pfizer")

txt_files = list.files(pattern = "*.txt");
files <- strsplit(as.character(lapply(txt_files, readLines)), " ")

length(txt_files)

```

```

news_date <- list()
l=1
for (l in 1:length(txt_files)){
  news_date[[l]] <- substr(txt_files[l],1,10)
}

pos <- read.csv("C:/Users/Karolina/Desktop/Positive.txt", header=FALSE, sep="\t")
pos<-as.vector(t(pos))
neg <- read.csv("C:/Users/Karolina/Desktop/Negative.txt", header=FALSE, sep="\t")
neg<-as.vector(t(neg))

i=1
a <- list()
b <- list()
c <- list()
d <- list()
e <- list()
for(i in 1:length(files)){
  a[i] <- list(files[[i]][files[[i]] %in% pos])
  b[i] <- list(files[[i]][files[[i]] %in% neg])
  d[[i]] <- append(noquote(paste(news_date[i])), (length(a[[i]])))
  e[[i]] <- append(noquote(paste(news_date[i])), (-length(b[[i]])))
  c[[i]] <- append(noquote(paste(news_date[i])), (length(a[[i]])+(-length(b[[i]]))))
}

total <- data.frame(t(as.data.frame(c)))
positive <- data.frame(t(as.data.frame(d)))
negative <- data.frame(t(as.data.frame(e)))

library(xlsx)
write.xlsx(total,"C:/Users/Karolina/Desktop/Pfizer.xlsx")

```

## Poliariškumo vektoriaus sudarymas

```

import os
import re
from textblob import TextBlob
import pandas as pd
import csv

files = [f for f in os.listdir('.') if re.match(r'([0-9]{4}\-[0-9]{2}\-[0-9]{2})(.*)\.txt$',
f)]
datos = []
for t in files:
    with open('{}'.format(t), 'r') as f:
        d = t[0:10]
        datos.append(d)
        reader = csv.reader(f)
        your_list = str(list(reader))
        y = list("".join(your_list.replace("
","").replace("[","").replace("]", "").replace("'", ""))).split(",")
        y = str(y).replace("[","").replace("]", "").replace("'", "").replace(",","")
        m = []
        m.append(y)
        for i in m:
            blob = TextBlob(i)
            for s in blob.sentences:
                print(d, s.sentiment.polarity)

article_polaris = pd.DataFrame({'Date':datos,
                                'Polarity':polar})

article_polaris.to_csv('{}'.format(company))

print('Finished')

```

## Autokoreliacijos tikrinimas akcijų kainų laiko eilutėms

```
library(quantmod)
library(Quandl)
library(timeSeries)
library(PerformanceAnalytics)
library(FinTS)
library(forecast)
library(tseries)
library(rugarch)

JNJ <- read.xlsx("C:/Users/Karolina/Desktop/Pharma stocks/stocks/JNJ.xlsx", header=TRUE,
sheetName="Sheet1")
JNJ <- JNJ[2]
JNJ <- as.vector(t(JNJ))
opar=par(mfrow=c(1,2))
plot(JNJ, xlab="Laikotarpis", ylab="", main="Johnson & Johnson")
par(opar)
JNJ <- diff(JNJ)

#Full and partial autocorrelation
opar=par(mfrow=c(1,2))
acf(JNJ, main="Johnson & Johnson"); pacf(JNJ, main="Johnson & Johnson")
par(opar)

#Augmented Dickey Fuller test
JNJ_adf <- adf.test(JNJ, alternative = "stationary"); JNJ_adf
```

## ARIMA modelio taikymas

```
library(xlsx)
library(stats)
library(lmtest)
library(tseries)
library(dplR)
library(forecast)
stock_link = "C:/Users/Karolina/Desktop/Pharma stocks/stocks/SNY.xlsx"
#sent_link = "C:/Users/Karolina/Desktop/Comp/Sanofisent.xlsx"
#BLOB sentiments
sent_link = "C:/Users/Karolina/Desktop/Comp/Blob_sent/Blob_sanofi.xlsx"

stock <- read.xlsx(stock_link, header=TRUE, sheetName="Sheet1")
sent <- read.xlsx(sent_link, header=TRUE, sheetName="Sanofi")

stock <- stock[2]
Sentiment_lag1 <- sent[3] ; Sentiment_lag1 <- as.vector(t(Sentiment_lag1)) ;
Sentiment_lag2 <- sent[4] ; Sentiment_lag2 <- as.vector(t(Sentiment_lag2)) ;
Sentiment_lag3 <- sent[5] ; Sentiment_lag3 <- as.vector(t(Sentiment_lag3)) ;
Sentiment <- sent[2] ; Sentiment <- as.vector(t(Sentiment)) ;
stock <- as.vector(t(stock))

vars <- cbind(Sentiment, Sentiment_lag1,Sentiment_lag2,Sentiment_lag3)

auto.arima(stock, xreg=vars)

model <- arima(stock, order=c(0,1,0), xreg=vars, transform.pars=TRUE, method='CSS-ML')
summary(model)
coeftest(model)

par(mar = rep(2, 4))
qqnorm(model$residuals, main="Sanofi paklaidų pasiskirstymo grafikas", asp = 1)
qqline(model$residuals, asp=1)

grangertest(Sentiment ~ stock, order=3)
grangertest(stock ~ Sentiment, order=3)
```