



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas

# **Didžiųjų duomenų charakteristikų įverčių analizė**

Baigiamasis magistro projektas

---

**Jurgita Archipova**  
Projekto autorė

**Prof. dr. Robertas Alzbutas**  
Vadovas

**Doc. dr. Šviesa Leitonienė**  
Vadovė

---



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas

## **Didžiųjų duomenų charakteristikų įverčių analizė**

Baigiamasis magistro projektas  
Didžiųjų verslo duomenų analitika (621G12002)

---

**Jurgita Archipova**  
Projekto autorė

**Prof. dr. Robertas Alzbutas**  
Vadovas  
**Doc. dr. Šviesa Leitonienė**  
Vadovė

**Doc. dr. Evaldas Vaičiukynas**  
Recenzentas  
**Prof. dr. Lina Dagilienė**  
Recenzentė

---



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas  
Jurgita Archipova

## **Didžiųjų duomenų charakteristikų įverčių analizė**

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Jurgitos Archipovos, baigiamasis projektas tema „Didžiųjų duomenų charakteristikų įverčių analizė“ yra parašytas visiškai savarankiškai ir 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 mokėjusi.

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

---

(vardą ir pavardę įrašyti ranka)

---

(parašas)

## Turinys

<b>Įvadas</b> .....	<b>9</b>
<b>1. Literatūros apžvalga</b> .....	<b>11</b>
1.1. Didieji duomenys .....	11
1.2. Duomenų įvairovė.....	13
1.3. Duomenų sparta .....	15
1.4. Duomenų apimtis .....	17
1.5. Duomenų teisingumas.....	18
1.6. Duomenų kintamumas .....	19
<b>2. Tyrimų metodai</b> .....	<b>21</b>
2.1. Duomenų apimties vertinimas .....	21
2.2. Duomenų teisingumo vertinimas .....	23
2.3. Istorinio kintamumo vertinimas .....	25
2.4. Kintamumo matų paklaidos ir efektyvumas .....	29
2.5. Programinės priemonės.....	30
<b>3. Tiriamoji dalis</b> .....	<b>32</b>
3.1. Duomenys .....	32
3.2. Duomenų spartos vertinimas.....	35
3.3. Entropijos ir informacijos kiekio įvertinimas .....	36
3.4. Duomenų teisingumo vertinimas .....	37
3.5. Duomenų kintamumo vertinimas.....	38
<b>Išvados</b> .....	<b>49</b>
<b>Literatūros sąrašas</b> .....	<b>51</b>
<b>Priedai</b> .....	<b>53</b>

## Paveikslų sąrašas

1 pav. Didžiųjų duomenų charakteristikų skaičiaus raida .....	12
2 pav. Duomenų tipai .....	13
3 pav. Struktūrizuotų ir nestruktūrizuotų duomenų formos .....	13
4 pav. Nestruktūrizuotų ir struktūrizuotų duomenų procentas pasaulyje .....	14
5 pav. Kas įvyksta per minutę internete .....	16
6 pav. Dienos uždarymo akcijų kainų grafikas .....	33
7 pav. Kintamumas keturiais skirtingais laikotarpiais.....	40
8 pav. Kombinuotas kintamumo ir entropijos grafikas .....	41
9 pav. Kasmetinis kintamumas gautas naudojant dieninius duomenis (1986–2018 m.).....	42
10 pav. Kintamumas skirtingais laikotarpiais.....	44
11 pav. Dienos uždarymo akcijų kainų ir ekonominių ciklų grafikas .....	45
12 pav. Kintamumas skirtingų JAV ekonominių ciklų metu .....	46

## Lentelių sąrašas

1 lentelė. Aprašomosios duomenų statistikos.....	32
2 lentelė. Vidutinis įrašų skaičius per metus .....	36
3 lentelė. Entropijos ir informacijos kiekio reikšmės skirtingais laikotarpiais.....	37
4 lentelė. Kintamumo įvertinimas laikotarpiu 1986-03-13–1999-12-31.....	39
5 lentelė. Kintamumo įvertinimas laikotarpiu 1986-03-13–1987-12-31.....	43
6 lentelė. Kintamumo įvertinimas laikotarpiu 1986 m. kovas – 1990 birželis.....	46
7 lentelė. Duomenų tipų klasifikavimas .....	53
8 lentelė. Nuorodos.....	54
9 lentelė. Šventinės ir sutrumpintos dienos JAV.....	55
10 lentelė. Kintamumo įvertinimas laikotarpiu 2000-01-01–2006-12-31.....	56
11 lentelė. Kintamumo įvertinimas laikotarpiu 2007-01-01–2011-12-31.....	56
12 lentelė. Kintamumo įvertinimas laikotarpiu 2012-01-01–2018-03-13.....	57
13 lentelė. Kasmetinis kintamumo įvertinimas laikotarpiu 1986–2018 m.....	58
14 lentelė. Kintamumo įvertinimas laikotarpiu 1988-01-01–1996-12-31.....	59
15 lentelė. Kintamumo įvertinimas laikotarpiu 1997-01-01–2000-12-31.....	59
16 lentelė. Kintamumo įvertinimas laikotarpiu 2001-01-01–2005-12-31.....	60
17 lentelė. Kintamumo įvertinimas laikotarpiu 2006-01-01–2008-12-31.....	60
18 lentelė. Kintamumo įvertinimas laikotarpiu 2009-01-01–2017-12-31.....	61
19 lentelė. Kintamumo įvertinimas laikotarpiu 2018-01-01–2018-03-13.....	61
20 lentelė. Kintamumo įvertinimas laikotarpiu 1990 m. liepa – 1991 m. kovas.....	62
21 lentelė. Kintamumo įvertinimas laikotarpiu 1991 m. balandis – 2001 m. vasaris .....	62
22 lentelė. Kintamumo įvertinimas laikotarpiu 2001 m. kovas – 2001 m. lapkritis .....	63
23 lentelė. Kintamumo įvertinimas laikotarpiu 2001 m. gruodis – 2007 m. lapkritis.....	63
24 lentelė. Kintamumo įvertinimas laikotarpiu 2007 m. gruodis – 2009 m. birželis .....	64
25 lentelė. Kintamumo įvertinimas laikotarpiu 2009 m. liepa – 2018 m. kovas.....	64

Archipova, Jurgita. Didžiųjų duomenų charakteristikų įverčių analizė. Magistro baigiamasis projektas / vadovai prof. dr. Robertas Alzbutas ir doc. dr. Šviesa Leitonienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Mokslo kryptis ir sritis: matematika, fiziniai mokslai

Reikšminiai žodžiai: didieji duomenys, charakteristikos, kintamumas, analizė

Kaunas, 2018. 64 p.

### **Santrauka**

Šio baigiamojo darbo tikslas – išnagrinėti didžiųjų duomenų charakteristikų įvairovę, atskirus jų matavimus bei įverčius ir pademonstruoti, kaip jie gali būti taikomi didelės verslo įmonės vienos duomenų imties tyrimo atveju.

Literatūros apžvalgoje apibrėžiama didžiųjų duomenų sąvoka, jų svarba, aktualumas bei aprašomos šių duomenų charakteristikos, kurios bus vertinamos turimiems duomenims, t. y. apimtis, įvairovė, teisingumas, sparta ir kintamumas. Būtent pastaroji charakteristika šiame darbe ir yra akcentuojama. Metodinėje darbo dalyje apžvelgiami aprašytų charakteristikų skaičiavimo/vertinimo metodai. Taip pat apžvelgiama skaičiavimams tiriamojoje dalyje naudojama programinė įranga.

Tiriamojoje dalyje analizuojami „Microsoft“ korporacijos akcijų kainų duomenys. Atsižvelgiant į kainų svyravimus sudaromi tyrimo laikotarpiai. Toliau apskaičiuojamos/įvertinamos jau minėtos charakteristikos. Vienos iš jų vertinamos bendrai visam duomenų rinkiniui, kitos – skirtingais laikotarpiais. Be to, kintamumo charakteristika vertinama ne tik pagal akcijų kainų svyravimus sudarytais laikotarpiais, bet ir laikotarpiais, gautais atsižvelgiant į pačio kintamumo kasmetinių reikšmių svyravimus, bei laikotarpiais, sudarytais atsižvelgiant į JAV ekonominius ciklus. Galiausiai atliekama apskaičiuotų charakteristikų analizė bei aprašomas kintamumo charakteristikos taikymas versle.

Archipova, Jurgita. Analysis of estimates of big data characteristics: Master's Thesis in Business Big Data Analytics / supervisors prof. dr. Robertas Alzbutas and doc. dr. Šviesa Leitonienė. The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Research area and field: mathematics, physical sciences

Key words: big data, characteristics, volatility, analysis

Kaunas, 2018. 64 p.

### **Summary**

In this Thesis, the research aim is to analyse the variety of big data characteristics, their individual measurements and estimates, and to demonstrate how they can be used to analyse a data sample from a large business enterprise.

The Literature Review Section defines big data, explains their importance and relevance, and describes the data characteristics, that will be evaluated with respect of the available data, such as volume, variety, veracity, velocity, and volatility, with the main focus of this Thesis placed on the latter characteristic. The Method Section of the Thesis gives an overview of the described characteristic computing/estimating methods. Furthermore, it reviews software used for big data computing in the Research Section.

The Research Section analyses share price data from Microsoft Corporation. The research periods are defined with regard to the price fluctuations. Then the above characteristics are computed/estimated. Some of them are computed for the whole data set, while others are evaluated with respect of different periods. Moreover, volatility is measured not only in the periods, defined with respect to share price fluctuations, but also in periods, marked out on the basis of the fluctuation of the annual volatility values, and in periods, designed with a view to the US economic cycles. Finally, the computed characteristics are analysed and the use of the volatility characteristic in business is described.



## Ivadas

**Temos aktualumas.** Pasaulis tapo informacine visuomene, kuri labai priklauso nuo duomenų. Kadangi informacinės sistemos kiekvieną dieną, kiekvieną sekundę generuoja didžiulį įrašų kiekį, atrodo, kad pasaulis pasiekia duomenų perkrovos lygį. Akivaizdu, kad norint apdoroti tokius duomenų kiekius reikalingos didelės saugyklos bei skaičiavimo ištekliai. Saugyklų pajėgumų augimą riboja įrangos ir technologijų raida, o duomenų kiekio augimas yra neribojamas. Konkrečiau, šiandien daugelis organizacijų priėmė ir plačiai naudoja informacines sistemas, veikiančias technologinėse platformose, ir daugelio iš jų darbotvarkės tapo priklausomos nuo duomenų. Organizacijose duomenys tiesiogiai veikia verslo procesų logiką. Informacija tapo jų verslo pagrindu. Todėl verslas reikalauja duomenų [1].

Kažkada duomenims apdoroti ir laikyti užtekdavo paprastų duomenų bazių, bet kai beveik viskas, kas vyksta mūsų visuomenėje, persikėlė į virtualią erdvę, visam tam didžiuliam duomenų kiekiui suvaldyti prireikė įspūdingų programų. Taip ir atsirado terminas didieji duomenys (angl. *Big Data*).

Anglų kalboje šis terminas apibrėžiamas charakteristikomis, kurios prasideda raidėmis „V“. Šių charakteristikų yra keliasdešimt ir jų skaičius vis auga. Vis dėlto yra trys klasikinės charakteristikos: *Volume* – milžiniškos ir beribės duomenų apimtys, *Velocity* – didelė sparta, nes milžiniškiems duomenų kiekiams apdoroti reikia milžiniško greičio ir *Variety* – įvairovė, nes apdorojamų duomenų gali būti pačių įvairiausių tipų, nuo paprasčiausių tekstinių įrašų iki finansinių apskaitų. Tačiau šiame darbe bus nagrinėjamos ir dar dvi naujesnės, bet ne mažiau svarbios charakteristikos: *Veracity* – teisingumas, nes duomenų patikimumas ir tikrumas yra būtinas, norint gauti naudingą informaciją bei priimant sprendimus, ir *Volatility* – kintamumas, duomenys nėra pastovūs, jie keičiasi ir tą būtina įvertinti (šiuo darbe, būtent, ir yra koncentruojamasi į kintamumo charakteristiką ir jos analizę).

Kuo toliau, tuo labiau *Big Data* tampa neišvengiamai svarbiu dalyku visur: pradedant viešuoju sektoriumi ir baigiant verslais. Naudojantis *Big Data* galima sutaupyti laiko, pinigų, greičiau priimti naudingus sprendimus, patogiau ir efektyviau analizuoti informaciją bei daryti reikšmingus atradimus.

**Tyrimo objektas** – didžiųjų duomenų charakteristikų įverčiai, jų apskaičiavimo bei analizės metodai.

**Tyrimo tikslas** – išnagrinėti didžiųjų duomenų charakteristikų įvairovę, atskirus jų matavimus bei įverčius ir pademonstruoti, kaip jie gali būti taikomi didelės verslo įmonės vienos duomenų imties tyrimo atveju.

**Uždaviniai:**

- 1) remiantis moksline literatūra, apžvelgti didžiųjų duomenų charakteristiką, jų skaičiavimo bei vertinimo metodikas;
- 2) išanalizuoti „Microsoft“ korporacijos akcijų kainų dinamiką nuo 1986 m. kovo 13 d. iki 2018 m. kovo 13 d.;
- 3) esamiems akcijų kainų duomenims įvertinti charakteristiką, tokias kaip: apimtis, įvairovė, teisingumas, sparta bei kintamumas;
- 4) atlikti apskaičiuotų charakteristikų įverčių analizę, t. y., interpretuoti gautų įverčių reikšmes, aptarti kintamumo charakteristikos svarbą bei taikymą versle.

**Tyrimo metodai** – mokslinės literatūros, didžiųjų duomenų charakteristikų skaičiavimo algoritmų, metodikų, programinės įrangos bei turimų duomenų analizė.

## 1. Literatūros apžvalga

Šioje darbo dalyje aprašomi didieji duomenys, jų svarba, aktualumas bei keletas juos apibūdinančių charakteristikų, kurios yra įvertinamos darbo tiriamojoje dalyje.

### 1.1. Didieji duomenys

Didieji duomenys ir atitinkamas terminas šiuo metu yra labai aktualus ir populiarus, nes tai nauja tema, kuri neseniai pradėjo vystytis. Atsižvelgiant į įvairius šaltinius, kuriais įmonės gali rinkti informaciją, reikia naujų būdų saugoti ir pasiekti šiuos duomenis, kad būtų galima juos tinkamai panaudoti. Didieji duomenys apima nestruktūruotus duomenis, pusiau struktūruotus duomenis ir struktūrizuotus duomenis. Yra įdiegtos naujos infrastruktūros, kuriomis galima tvarkyti didelius įvairių rūšių duomenų kiekius, tačiau tradiciniai duomenų modeliai nėra pajėgūs apdoroti tokių duomenų. Įmonės gali naudoti didžiuosius duomenis, kad pritaikytų produktą ir surinktų visą informaciją, reikalingą maksimaliam pelnui gauti ir konkurenciniam pranašumui įgyti.

Didžiųjų duomenų didėjimo svarbą lemia kompiuterinių tinklų, tokių kaip internetas bei duomenų saugojimo plėtra, kurie leidžia užfiksuoti didelius duomenų kiekius. Grupinės kompiuterinės sistemos suteikia galimybę saugoti duomenis, taip pat juos analizuoti ir tvarkyti, o tai užtikrina duomenų rinkinio patikimumą ir, galiausiai, debesų skaičiavimo technologiją, leidžiančią padidinti saugojimo pajėgumus ir taip sutaupyti daug pinigų.

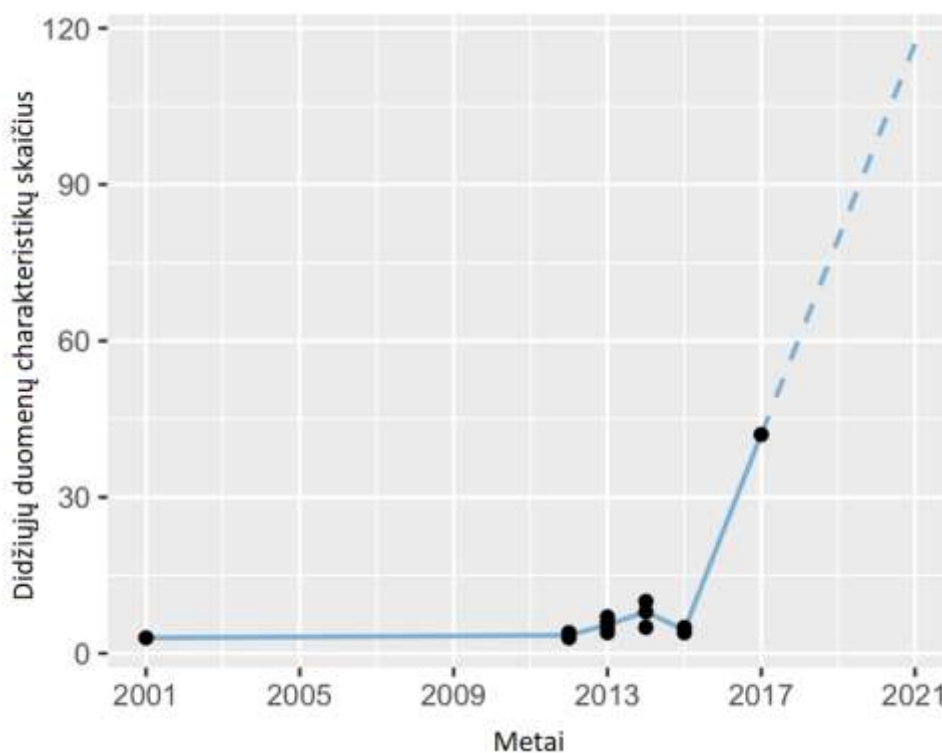
Didieji duomenys buvo sukurti dėl kelių veiksnių, tačiau pagrindiniai veiksniai susiję su atminties „kainos“ sumažėjimu: tapo lengva turėti didžiulę atmintį bei apdoroti duomenis. Kitas svarbus veiksnys – nesudėtingas kelių kompiuterių sujungimas.

Masinių technologijų lyderių taryba (2010, angl. *Mass Technology Leadership Council*) teigė, kad didieji duomenys yra duomenys, kurių dydis yra didesnis nei bet kurios duomenų bazės galimybė saugoti ir analizuoti. Kitas didžiųjų duomenų apibrėžimas yra tai, kad šių duomenų „didumas“ yra susijęs ne tik su jų apimtimi, bet taip pat ir su greičiu bei įvairove. Paprasčiau tariant, kai duomenys yra per dideli, kad tradicinės sistemos galėtų juos apdoroti, tai jie vadinami didžiais duomenimis (M. M. Gobble, 2013).

„Gartner Inc.“ 2012 m. oficialiai pristatė didžiųjų duomenų apibrėžimą: didieji duomenys – tai didelės apimties, didelės spartos ir (arba) įvairaus pobūdžio informacija, kurios apdorojimui reikalingos naujos apdorojimo formos, kad būtų sudarytos sąlygos priimti sprendimus, daryti išvalgas ir optimizuoti procesus [2].

Iš pradžių kai kurios pagrindinės kompanijos, tokios kaip „SAP“ ir „IBM“, galėjo turėti duomenų bazes, kurios, kaip manoma, sugebėdavo saugoti didelius duomenų kiekius, tačiau padidėjus ryšių priemonių ir įrankių skaičiui, tapo reikalingos naujos duomenų bazės, kad būtų galima saugoti šį didžiulį duomenų kiekį [3].

Apimtis/kiekis nėra vienintelė didžiųjų duomenų charakteristika, jų yra ir daugiau. Didžiųjų duomenų charakteristikos anglų kalboje yra apibūdinamos V raidėmis. Yra trys tarsi klasikinėmis laikomos charakteristikos: apimtis, sparta ir įvairovė. 1.1 paveiksle pateikta charakteristikų skaičiaus raida bėgant metams. Galime pastebėti, jog taip vadinamų V raidžių skaičius staigiai ima augti nuo 2015 m. Praėjusiais, t. y. 2017 m., V raidžių skaičius išaugo iki 42 ir, manoma, jog šis skaičius tik didės [4].



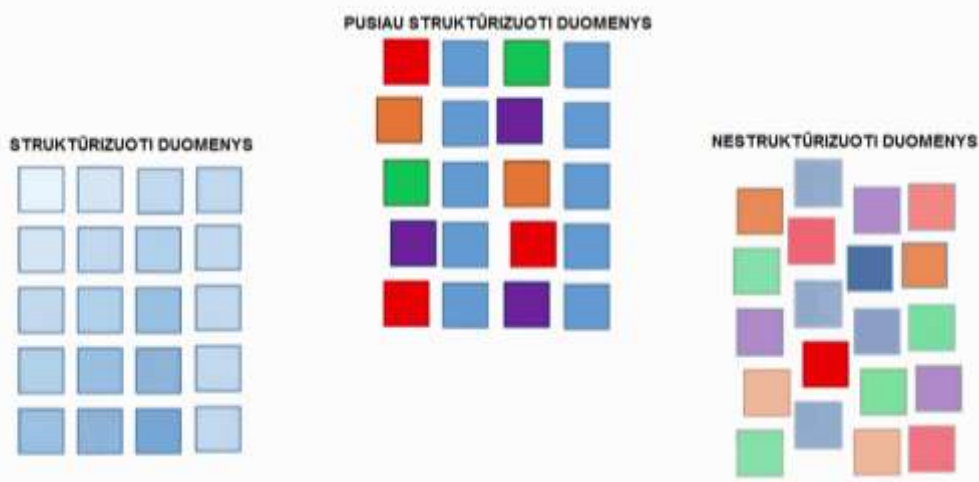
1 pav. Didžiųjų duomenų charakteristikų skaičiaus raida [4]

Iš tiesų, nėra vieningo didžiųjų duomenų apibrėžimo ar kažkokio šablono, pagal kurį būtų galima spręsti, jog duomenys yra didieji. Šiam apibrėžimui nėra ribų. Požiūris, jog tik duomenys, kurie pasižymi, jau minėtomis trimis klasikinėmis charakteristikomis, yra laikomi didžiaisiais – klaidingas. Sparta yra tarsi pagrindinė savybė, padedanti išsiskirti didiesiems duomenims iš kitų, o apimtis ir įvairovė nėra būtinos didžiųjų duomenų sąlygos. Didieji duomenys yra „greiti“ ir dažnai duomenys, pasižymintys didele apimtimi, įvairove, lankstumu ir panašiomis savybėmis, vis tiek yra laikomi „mažais“. Būtent jau minėtoji spartos charakteristika labiausiai prisidėjo prie didelio dėmesio ir investicijų į *Big Data* įmones. Yra labai mažai darbų, kuriose būtų siekiama išsamiai

išnagrinėti didžiųjų duomenų ontologiją, o ne tiesiog teigti, kad tai yra duomenys, turintys tam tikras savybes. Tiesą sakant, daugumoje tyrimų, kuriuose diskutuojama apie didžiuosius duomenis, šis terminas vartojamas kaip visiška amforinė frazė, numatanti, jog visi tokio tipo duomenys pasižymi tam tikrais bendrais bruožais, tačiau, iš tiesų, egzistuoja įvairios didžiųjų duomenų formos [5].

## 1.2. Duomenų įvairovė

Ši charakteristika nusako skirtingus duomenų formatus, kurie negali būti saugomi struktūrizuotose reliacinėse duomenų bazių sistemose [6]. Pagrindiniai duomenų tipai: struktūrizuoti, pusiau struktūrizuoti ir nestruktūrizuoti duomenys. Visi jie turi būti panaudoti, kad organizacija gautų tikrąją didžiųjų duomenų vertę. Įvairių duomenų šaltinių integravimas ir analizavimas leidžia organizacijoms padaryti tam tikras įžvalgas. Įmonės, kurios plečia duomenų šaltinių sritį, taip pat sukuria ir didesnę vertę [7].



2 pav. Duomenų tipai [8]

Struktūrizuoti duomenys apima tradicinius transakcijų duomenis, o nestruktūrizuoti apima tokius duomenis, kaip: dokumentai, elektroniniai laišakai, socialinės žiniasklaidos teksto pranešimai, vaizdo įrašai, nejudantys vaizdai, garso įrašai, grafikai, GPS signalai ir dar daugiau [6]. Pagrindinis iššūkis yra nestruktūrizuoti duomenys, o tradicinius duomenų tipus yra gana paprasta tvarkyti bei saugoti.



3 pav. Struktūrizuotų ir nestruktūrizuotų duomenų formos [9]

Struktūrizuoti duomenys puikiai tinka RDBVS (reliacinių duomenų bazių valdymo sistemų) atveju. Bet dauguma sugeneruojamų duomenų yra nestructūrizuoti. Skaitmeninis pasaulis atvėrė duris nestructūrizuotiems duomenims, todėl RDBVS nebebuvo perspektyvios. Tiesą sakant, vien „Facebook“ sukuria daugiau nei 30 petabaitų nestructūrizuotų duomenų (paveikslėlių, pranešimų ir pan.). Šiuo metu beveik 80 % duomenų yra nestructūrizuoti. Šio tipo duomenys negali būti suskirstyti į lenteles, tačiau *Big Data* technologijų pagalba dabar galima juos suprasti [3].



4 pav. Nestructūrizuotų ir struktūrizuotų duomenų procentas pasaulyje [10]

Savo ruožtu, juos lengva įrašyti, bet užtrunka daugiau laiko juos perskaityti ir sutvarkyti [7]. Būtent šio tipo duomenų kiekiai auga daug greičiau nei struktūrizuoti duomenys. Nustatyta, kad nestructūrizuoti duomenys kas tris mėnesius padvigubėja [6].

Dauguma didžiųjų duomenų projektų susiduria su daugiau iššūkių, susijusių su duomenų tipu, ir su mažiau iššūkių, susijusių su jų kiekiu. Duomenų tipai – vienas svarbiausių iššūkių, turinčių įtakos didelės duomenų sistemos našumui. Iššūkių atsiradimą gali lemti šios priežastys:

- Augimas, dėl kurio reikia įdiegti naujas sistemas.
- Naujų sistemų, kurios teikia duomenis, tačiau kelia problemų, nustatant jų svarbą didelėse duomenų sistemose, įdiegimas [7].

Daugelį metų įmonės tvarkė didelius duomenų kiekius ir žiūrėjo į tai, kaip į įprastą procesą, tačiau galimybė nestructūrizuotus duomenis įtraukti į analizę, yra vertinama kaip unikali ir nauja perspektyva [6].

### 1.3. Duomenų sparta

Dar viena svarbi didžiųjų duomenų charakteristika yra sparta, kuri apibūdinama, kaip greitis, kuriuo galima pasiekti saugomus duomenis, kad juos būtų galima naudoti. Duomenys turi būti prieinami kuo arčiau realaus laiko, kad būtų galima taikyti daugumą realaus pasaulio programų. Esminis šiuolaikinių technologijų apribojimas yra galimybė greitai ir racionaliai priėti prie didžiulių duomenų kiekių [3]. Kai kurie duomenys gali būti keičiami dinamiškai po jų sukūrimo, kai kurie iš jų gali būti beveik pasenę per kelias sekundes. Siekiant išgauti arba keisti informaciją, šių duomenų analizė turi būti atliekama iš karto [1].

Yra žinoma, kad socialinės žiniasklaidos srautai ir duomenys nėra tokie dideli, kaip mašinu sugeneruoti duomenys, tačiau jie pateikia vertingą informaciją. Pavyzdžiui, „Twitter“ naujienų srauto didelis greitis užtikrina maždaug 8 terabaitus duomenų per dieną. Duomenys, surinkti iš socialinių tinklų („Facebook“ ir „Twitter“), įskaitant nuomones ir kitą asmeninę informaciją, atnaujinami labai dideliu greičiu. Organizacijai tai gali būti labai naudinga, siekiant pritaikyti bei parduoti savo produktą. B. Sheppard (2011) teigė, kad per tinklalapius renkami duomenys teikia daugybę duomenų, kuriuos reikia filtruoti ir apdoroti realiuoju laiku, kad būtų galima gauti maksimalią informaciją, kurios reikia. G. Collett (2011) apibūdino greitį kaip duomenų apdorojimo laiką. Svarstoma, ar šimtai gigabaitų yra didieji duomenys, atsižvelgiant į greitį ir laiką, reikalingą jų apdorojimui [3].

Kai kurie mokslininkai mano, kad 90 % viso pasaulio duomenų buvo sukurta per paskutinius keletą metų. Didieji duomenys kelia didelį iššūkį socialinių tinklų svetainėms. Pavyzdžiui, „Facebook“ turi saugoti petabaitus duomenų, kuriuos sukuria aktyvūs vartotojai. Tokie srauto duomenys turi būti saugomi, o užklausos turi būti apdorojamos realiuoju laiku.



5 pav. Kas įvyksta per minutę internete [10]

Greitis turi tris svarbius aspektus: judantys duomenys (angl. *Data-In-Motion*), naudingumo trukmė (angl. *Lifetime of Data Utility*) ir jų analizė realiu laiku (angl. *Real Time Big Data Analytics*). Mokslininkai dažnai kalba apie kintančius ir nekintančius duomenis. Vienas iš rūpesčių yra greitai judančių duomenų srautų nuoseklumas ir išsamumas. Greitis taip pat apima savalaikiškumo ar uždelimo charakteristikas – ar duomenys yra užfiksuoti greitai, ar su laiko praleidimu, t. y., uždelsimu. Duomenų naudingumo tarnavimo laikas – antrasis greičio aspektas, nurodantis, kiek laiko duomenys bus vertingi. Ar jie visada bus vertingi, o galbūt jie greitai senėja ir praranda savo reikšmę ir svarbą. Duomenų analizė realiu laiku – trečiasis aspektas, reiškiantis greitį, kuriuo duomenys turi būti išsaugomi. Tai yra vienas iš svarbiausių saugojimo, analizės ir diegimo architektūros veiksnių, kuriuos įmonės turi atlikti šiandien. Kai apsilankoma sudėtingo turinio svetainėje, pvz., „Yahoo!“, tie skelbimai, matomi tinklalapyje, yra specialiai parinkti, atsižvelgiant į naujausią jūsų žiniatinklio vizitų istoriją. Tas pats procesas vyksta, kai prisijungiama prie „Amazon“ ar „Netflix“ ir yra matomi jums rekomenduojami pirkiniai. Fiksavimo, analizės ir diegimo architektūra turi atlikti analizę realiu laiku ir daryti tai kiekvieną minutę su tūkstančiais naujų lankytojų. Kas pasikeitė? Duomenys visada buvo, bet galimybė fiksuoti, analizuoti ir veikti (realiuoju laiku) iš tiesų yra visiškai naujas didžiųjų duomenų technologijos bruožas [6].



## 1.4. Duomenų apimtis

### Atminties kiekis

Apimtis yra fizinis duomenų kiekis, kuris sparčiai didėja. Atitinkamas duomenų augimas yra greitesnis nei turimų saugyklų pajėgumų augimas [1]. Kitaip sakant, apimtis nusako surinktų ir saugomų duomenų kiekį. Šie kiekiai svyruoja nuo terabaitų iki petabaitų duomenų [3]. Didžiosioms įmonėms nėra svetimi didieji duomenys. Jau 1980 m. didžiausia pasaulyje siuntų pristatymo kompanija „United Parcel Service Inc.“ (UPS) pradėjo kaupti ir stebėti siuntinių judėjimo duomenis, kurie dabar sudaro 16,3 milijono siuntinių per dieną ir yra saugoma daugiau nei 16 petabaitų duomenų. „Wal-Mart Stores Inc.“ įrašo daugiau nei 1 milijoną klientų operacijų per valandą bei generuoja daugiau nei 2,5 petabaitų duomenų [6].

Dėl spartaus duomenų kiekio augimo jų saugojimo išlaidos šiandien yra drastiškai sumažėjusios. Tai rodo pavyzdys, kad 2011 m. diską, kuris galėtų saugoti visą pasaulio muziką, buvo galima lengvai įsigyti už maždaug 600 dolerių. Šis pavyzdys yra naudojamas įrodyti, kad didžiųjų duomenų problema yra ne jų kiekis, o priemonės, skirtos duomenų saugojimui, analizavimui ir prieigai. B. Sheppard (2011) teigė, kad duomenų apimtis skiriasi priklausomai nuo organizacijos dydžio: didelė organizacija gali turėti keletą petabaitų, o mažos ir vidutinio dydžio organizacijos gali turėti kelis terabaitus duomenų. Kai tvarkyti surinktus ir saugomus duomenis tampa problema, duomenys, atsižvelgiant į organizacijos dydį, gali būti vadinami didžiaisiais [3].

Egzistuoja tarsi bendras susitarimas, kad jei duomenys telpa gigabaituose, tai tokie duomenys greičiausiai nėra didieji, bet jeigu jie užima terabaitus, petabaitus ar netgi daugiau, tai tokius duomenis drąsiai galime vadinti didžiaisiais. Kiekis yra pagrindinė priežastis, dėl kurios tradicinės reliacinės duomenų bazių valdymo sistemos (RDBVS, duomenų saugyklos) nesugeba tvarkyti didžiųjų duomenų. Taip yra dėl patikimumo trūkumo, ilgai trunkančių užklausų ir jų nesugebėjimo tvarkyti naujus nestruktūrotų arba pusiau struktūrotų duomenų šaltinius, tokius kaip tekstas ir panašiai.

Pagal kiekį duomenys yra skirstomi į mažos (angl. *Low Volume*) ir didelės apimties (angl. *High Volume*) duomenis. Būtent didelis duomenų kiekis yra viena iš pagrindinių didžiųjų duomenų charakteristikų, todėl šie duomenys priklauso *High Volume* grupei. [2]. Kaip ir buvo minėta anksčiau, duomenys vadinami didžiaisiais, kuomet pasiekia petabaitus, zetabaitus ar net didesnę kiekį. Tačiau kiekis nėra susijęs tik su masiniu duomenų baitų kiekiu. Jis taip pat gali būti kiekybiškai įvertintas pagal įrašų, sandorių, lentelių ar failų kiekį duomenų rinkinyje [2]. Duomenų tipų klasifikavimo lentelė, kurioje pateiktas ir skirtingiems tipams būdingas duomenų kiekis, pateikta 1 priede.

Didėjant duomenų kiekiui darosi vis sudėtingiau atrinkti svarbią informaciją ir daryti efektyvius, greitus bei teisingus sprendimus. Dėl šios priežasties labai svarbu išmatuoti informacijos kiekį, o tai galime padaryti pasitelkę Claude Shannon informacijos teoriją.

## 1.5. Duomenų teisingumas

### Duomenų kokybė

Kiekvieną dieną sukuriama didžiuliai duomenų kiekiai. Šie duomenys pateikiami iš skaitmeninių nuotraukų, vaizdo įrašų, įrašų iš socialinės žiniasklaidos svetainių, pažangių jutiklių, pirkimo operacijų įrašų ir mobiliųjų telefonų GPS signalų, ir t. t. Trumpai tariant, didieji duomenys yra nepatogūs duomenys [7]. Be anksčiau išvardintų keturių „V“, dar vienas svarbus ir seniai pripažintas aspektas, kuris susijęs su duomenų tikslumu ir aiškumu, yra teisingumas. Šis elementas buvo įvestas „IBM“ kaip duomenų patikimumo klausimas. Tai skamba racionaliai, nes kuo didesnis duomenų kiekis, tuo daugiau dėmesio reikia kreipti į duomenų teisingumą. Ši didžiųjų duomenų charakteristika atspindi duomenų neapibrėžtumą [11].

Tik neseniai buvo pripažinta informacijos kokybės svarba, ir raginama didžiuosius duomenis apibūdinti ne tik anksčiau nustatytais matmenimis (apimtimi, įvairove, greičiu ir kintamumu), bet taip pat dar vienu aspektu: teisingumu. Teisingumas siejamas su didiesiems duomenims būdingu neapibrėžtumu, kuris, kaip tikimasi, sparčiai didės per artimiausius metus. Bet nežiūrint į neapibrėžtumą, duomenyse vis dar yra vertingos informacijos. Norint ją išgauti iš duomenų, turi būti patikrintas duomenų teisingumas. Neapibrėžtumo valdymas skaitiniams duomenims gali būti atliekamas kombinuojant kelis mažiau patikimus šaltinius, siekiant sukurti tikslesnius ir naudingesnius duomenis arba naudojant „pažangią matematiką“, apimančią neapibrėžtumą, pavyzdžiui, tvirtus optimizavimo metodus ir „Fuzzy“ logiką. Teksto duomenų neapibrėžtumo valdymas sudėtingesnis, nes tekstiniai duomenys, ypač socialinės žiniasklaidos priemonėse, yra labai neaiškūs tiek išraiška, tiek turiniu. Turinio neapibrėžtumą siūloma valdyti, kiekybiškai įvertinant turinio objektyvumo (angl. *Objectivity*), tiesos (angl. *Truthfulness*) ir patikimumo (angl. *Credibility*) (OTC) lygį.

Didieji duomenys gali pasižymėti skirtingomis savybėmis, kurios gali turėti įtakos jų kokybei. Priklausomai nuo kilmės, duomenų apdorojimo technologijų ir metodų, naudojamų duomenų rinkimui ir moksliniams atradimams, didieji duomenys gali turėti dviprasmybių ir netikslumų, kuriuos reikia nustatyti ir apskaičiuoti, siekiant sumažinti išvesties klaidas ir padidinti gautų įžvalgų tikslumą. Didelis duomenų teisingumas dabar pripažįstamas kaip būtina savybė, papildanti kitas anksčiau nustatytas charakteristikas (apimtį, įvairovę ir kintamumą).

Charakteristikos, tokios kaip kiekis, greitis ir įvairovė gali būti pakankamai tiksliai kiekybiškai išmatuojamos. Tačiau teisingumas yra sudėtingesnis aspektas, kuris neturi suderintų matavimo būdų, ypač tekstinių didžiųjų duomenų rinkinių atveju. Aptinkant neapibrėžtumą tekstiniuose duomenyse, jis gali būti sumažintas, o tai gali pagerinti duomenų teisingumą. Taip yra todėl, kad neapibrėžtumas sukelia ne tik dviprasmybes, bet ir galimus faktinius neatitikimus. Taigi, norint nustatyti ir išmatuoti neapibrėžtumą, reikia apibūdinti pagrindinius neapibrėžtumo šaltinius: subjektyvumą (angl. *Subjectivity*), apgaulę (angl. *Deception*) ir netikėtumą (angl. *Implausibility*) (SDI). SDI didina tekstinių duomenų neapibrėžtumą, todėl tokiu būdu mažėja teisingumas. Taigi, siūloma apibrėžti tris pagrindinius teorinius teisingumo matmenis: objektyvumą, tikrumą ir patikimumą. Kiekvienas iš šių matmenų apibūdina įvairias didžiųjų duomenų problemas ir taip gali sumažinti duomenų kokybę kartu su jų verte [12].

## 1.6. Duomenų kintamumas

### Duomenų nepastovumo samprata

Pasaulis tampa vis labiau nepastovus. Ši įžvalga ateina ne tik iš mokslo, bet ir verslo pasaulio. Yra pakankamai pavyzdžių: atsargų arba žaliavų rinkų augimas/nuosmukis ir vis trumpesnis produkto gyvavimo ciklas. Kad galėtumėte reaguoti į padidėjusį nepastovumą, pirmiausia reikia jį įvertinti kiekybiškai. Reikia atsakyti į tris klausimus: ką reikėtų išmatuoti, kaip tai padaryti, kai keičiasi svyravimų aukštis ir koks yra svyravimų laipsnis. Iš jų pirmasis yra svarbiausias. Norint atsakyti į šį klausimą, reikia suprasti priežastis ir priežastinius nepastovumo veiksnius bendrovei. Tačiau kintamumas daro tokį patį poveikį visoms įmonėms; jis gali turėti daug formų ir priežasčių. Yra labai svarbu atlikti nepastovumo analizę, kuri būtų pritaikyta atsižvelgiant į įmonę. Tačiau kintamumo matavimas nėra galutinis tikslas: pagrindinis ketinimas yra aptikti nepastovumo rodiklius, kurie anksti įspėtų apie pokyčius. Tai turėtų pagerinti ir pagreitinti sprendimų priėmimą.

Nepastovumas ir su juo susijęs neapibrėžtumas yra reguliariai registruojamas reiškinys. Kintamumas laikomas viena iš dešimties mokslininkų, specialistų ir vadovų kontrolės sričių. Terminas „nepastovumas“ vartojamas daugelyje sričių. Apskritai, nepastovumas yra trumpalaikis laiko eilučių svyravimas aplink jo vidutinę vertę arba tendą. Kintamumas ekonomikoje dažnai naudojamas terminas, bet jis nėra aiškiai apibrėžtas.

Dažnai nepastovumas apibrėžiamas kaip nenumatytas išorinių ir vidinių ekonominių parametru svyravimo diapazonas ir dažnumas įmonėms, kurių sekos modeliai nėra arba yra sunkiai nuspėjami. Šis labai platus apibrėžimas apima daugybę įvykių: žaliavų kainų nepastovumą, paklausos pokyčius ir konkurencines pozicijas, pvz., produkto gyvavimo ciklus ir technologinius pokyčius, taip pat stiprius ekonominius svyravimus, kurie stebimi nuo 2008 m. finansų krizės.

Kintamumas, ypač įmonės lygmeniu, dažnai apibūdinamas kaip subjektyvus, kokybiškai suvokiamas pagreitintų pokyčių reiškinys be operatyvizuoto nepastovumo sampratos. Tam, kad būtų galima pateikti diferencijuotą ataskaitą apie nepastovumo raidą ir įtaką kontrolei, būtina išnagrinėti įvairių įmonių veiksnių svyravimus ir nustatyti nepastovumo rodiklius. Lemiamas veiksnys – konkrečios įmonės konkrečių rodiklių analizė, o ne standartinių matavimų raida [13].

Kintamumas reiškia duomenų verčių pokyčius laikui bėgant. Gali pablogėti įranga, taip pat gali būti įtraukiamos paklaidos ar poslinkiai, tokie kaip matavimai iš palydovų. Laikui bėgant sukauptų duomenų analizė tampa kritine, norint suprasti duomenų generavimo pokyčius, siekiant ištaisyti dreifą ar nepastovumą ir suteikti galimybę istoriškai analizuoti visus duomenis. Ši sąvoka nėra nauja didžiųjų duomenų srityje, tačiau ji vis dar išlieka svarbi [14].

Didelis duomenų nepastovumas nurodo, kiek laiko duomenys yra svarbūs ir kiek laiko jie turėtų būti saugomi. Šiame realaus laiko duomenų pasaulyje reikia nustatyti, koku momentu duomenys yra netinkami dabartinei analizei. Kintamumo/nepastovumo sąvoka pabrėžiama, kaip greitai keičiasi duomenys, o jei į tai neatsižvelgiama, tuomet bet kokie analizės rezultatai gali būti neteisingi. Galima suprasti, jog svarbu įvertinti numatomą duomenų kintamumą, o tai galima padaryti jį lyginant su istoriniu duomenų kintamumu [15].

## 2. Tyrimų metodai

Šioje baigiamojo projekto dalyje aprašomi didžiųjų duomenų charakteristikų vertinimo metodai, pateikiamos reikalingos formulės, analizuojama jų struktūra ir nurodomos kintamųjų reikšmės. Taip pat aprašoma tyrimui naudota programinė įranga.

### 2.1. Duomenų apimties vertinimas

#### Informacijos kiekis

Šiame darbe duomenų apimties arba, kitaip duomenų kiekio, charakteristika yra siejama su informacijos kiekiu. Gerai žinoma, jog informacija perduodama signalais. Tačiau kiek jos perduodama? Norint rasti atsakymą į šį klausimą, reikia apibūdinti pačią informacijos kiekio sąvoką. Šiuolaikinės informacijos teorijos pradininkas Claude Shannon pasiūlė informacijos kiekį apibūdinti pašalinto neapibrėžtumo kiekiu. Jei prieš kažkokį sąlyginį eksperimentą pradinis neapibrėžtumas yra  $H_1$ , o po eksperimento liekamas neapibrėžtumas yra  $H_2$ , tai eksperimento metu gautas informacijos kiekis yra

$$I = H_1 - H_2. \quad (2.1)$$

Neapibrėžtumas, kurio tarptautinis sinonimas yra entropija, kaip sąvoka būdinga tik netiksliai apibūdinamai situacijai. Pavyzdžiui, jeigu koks nors įvykis gali įvykti, tačiau gali ir neįvykti, tai abu išsakyti teiginiai vertinami tikimybėmis. Tikimybę, kad įvykis įvyks pažymėjus  $p$ , o tikimybę, kad jis neįvyks –  $1 - p$ , vidutinė entropija užrašoma tokiu būdu:

$$H = -p \log p - (1 - p) \log(1 - p). \quad (2.2)$$

Kai kalbama apie atsitiktinį dydį, kurio galimos reikšmės charakterizuojamos tikimybių pasiskirstymo dėsnio  $p_1, p_2, \dots, p_n$ , vidutinė entropija yra lygi:

$$H = - \sum_{i=1}^n p_i \log p_i. \quad (2.3)$$

Nesunku įsitikinti, jog didžiausia entropija yra gaunama, kai visos tikimybės vienodos, t. y.,  $p_i = 1/n$ . Tuomet entropija lygi  $H_{max} = \log n$ .

Perduodant informaciją be klaidų arba kai eksperimento metu neapibrėžtumas visiškai pašalinamas ( $H_2 = 0$ ), tada pagal (2.1) formulę informacijos kiekis išreiškiamas taip pat kaip entropija:

$$I = - \sum_{i=1}^n p_i \log p_i. \quad (2.4)$$

Informacijos kiekio, apibūdinto (2.1), (2.4) formulėmis, vienetas priklauso nuo pasirinkto logaritmo pagrindo. Kai logaritmo pagrindas yra 2, tai informacijos vienetas yra bitas, o jeigu logaritmo pagrindas 8, tai informacijos vienetas yra baitas. Aštuoni bitai atitinka vieną informacijos baitą.

Tarkime, kad iš informacijos šaltinio gaunami pranešimai perteikiami simboliškai  $c_1, c_2, \dots, c_n$  (kaip rašytinis tekstas – raidėmis). Jeigu simbolių pasikartojimo dažniai arba, kitaip sakant, jų tikimybės yra  $p_1, p_2, \dots, p_n$ , tai kiekvienas simbolis vidutiniškai perneša (2.4) formule išreikštą informacijos kiekį. Didžiausią informacijos kiekį kiekvienas simbolis perneša, kai visų jų tikimybės yra vienodos –  $p_i = 1/n$ . Tuomet vienas simbolis perneša  $I_{max} = \log n$  informacijos.

Informacijos kiekis, tenkantis pranešimo simboliams, yra siejamas su šaltinio kodavimu. Kai tikimybės  $p_1, p_2, \dots, p_n$  viena nuo kitos labai skiriasi, galima teigti, kad toks šaltinis užkoduotas neracionaliai. Kodavimo racionalumui įvertinti naudojamas rodiklis, vadinamas perteklumu ir išreiškiamas tokiu būdu:

$$r = 1 - H/H_{max} = 1 - \log(1/n) : \sum_{i=1}^n p_i \log p_i. \quad (2.5)$$

Kai šaltinio perteklumas yra didelis, tai jį galima sumažinti perkoduojant pranešimą taip, kad visų simbolių pasikartojimo dažnis būtų bent apytikriai vienodas. Pavyzdžiui, įrašant informaciją į kompiuterių kaupiklius. Dažnai pradiniai duomenys, tekstai ar dokumentai specialiai suspaudžiami specialiomis archyvavimo programomis.

Tarkime, kad šaltinis per laiko vienetą sukuria  $v$  simbolių, o kiekvienam simboliui tenkantis informacijos kiekis išreiškiamas (2.4) formule. Tai reiškia, kad šaltinis per laiko vienetą sukuria informacijos kiekį, lygų

$$I' = \sum_{i=1}^n p_i \log p_i, \quad (2.6)$$

kuris vadinamas informacijos šaltinio našumu.

Šios charakteristikos taikomos ir tolydinių procesų  $s(t)$  informacinėms savybėms nusakyti. Tolydinio signalo šaltinio entropija išreiškiamą formule:

$$h = - \int \log[w(x)] w(x) dx, \quad (2.7)$$

kur  $w(x)$  – signalo  $s(t)$  reikšmių pasiskirstymo tikimybių tankio funkcija. Informacijos, gautos perduodant vieną signalo  $s(t)$  atskaitą, kiekis išreiškiamas formule, analogiška (2.1):

$$I = h_1 - h_2. \quad (2.8)$$

Tolydinius signalus, skirtingai nuo diskrečių signalų, perduoti neiškraipytus neįmanoma, nes triukšmai, kad ir labai maži, visada yra, o ir signalų atkūrimo įtaisų paklaidos gali būti interpretuojamos kaip triukšmai.

Tarkime, kad signalas  $s(t)$  ir trukdis  $n(t)$  yra normalieji procesai. Jų tikimybių tankio funkcijos yra Gauso funkcija, o dispersijos atitinkamai  $\sigma_s^2 = S$  ir  $\sigma_n^2 = N$ . Apskaičiavus pagal (2.7) formulę gaunama:

$$h_1 = \log \sqrt{2\pi e S}, \quad h_2 = \log \sqrt{2\pi e N}. \quad (2.9)$$

Toliau pagal (2.8) formulę apskaičiuojamas informacijos kiekis:

$$I = \log \sqrt{S/N}. \quad (2.10)$$

Matome, jog tolydinio signalo pernešamas informacijos kiekis priklauso tik nuo signalo ir trukdžio dispersijų santykio. Ši išvada svarbi, nes ji teoriškai pagrindžia signalo ir triukšmo santykio reikšmę [16].

## 2.2. Duomenų teisingumo vertinimas

Literatūroje yra keletas duomenų teisingumo modelių, iš kurių vienas yra „didžiųjų duomenų teisingumo“ modelis: objektyvus, tikras ir patikimas (OTC). Jis naudojamas apskaičiuoti teisingumo matą. Toliau pateikiama informacija OTC modelį.

### Objektyvumas

OTC yra trimatis modelis, kurį pasiūlė T. Lukoianova. Pirmas teisingumo matas OTC modelyje yra duomenų objektyvumas / subjektyvumas. Objektyvumas iš esmės susijęs su faktais, tiesa ir realybe. Objektyvumas – tai žinios, kurios yra įrodytos, o subjektyvumas – tai žinios, kurios gali būti klaidingos. Yra lingvistiniai įrankiai, tokie kaip automatinis sentimentų ir nuomonės klasifikavimas tekste. Be to, objektyvumas yra tik tiesos versija, todėl teisingumas taip pat susijęs su kitu matu, kuris yra tiesa / apgaulė.

## **Tiesa**

Antrasis aspektas – tiesa, tekstiniuose duomenyse gali būti nustatyta patikrinant, ar tekste yra klaidingų nuomonių ir klaidingų išvadų. Tai galima patikrinti naudojant apgaulės testą. Jei testas sėkmingas, tai reiškia, kad tekstas yra teisingas, tačiau jei testas nesėkmingas, vartotojas turi ieškoti alternatyvų, kad būtų galima toliau tikrinti faktus. Be to, didėjant duomenų kiekiui, didėja ir apgaulingo/klaidingo teksto tikimybė. Dėl šios priežasties gali būti neteisingai atlikta analizė ir to pasekoje, gauti neteisingi rezultatai. Yra daug apgaulės aptikimo programinės įrangos, kuri apskaičiuoja statistikas, naudojamas teisingam ar apgaulingam tekstui klasifikuoti arba nustatant netikras produktų/paslaugų apžvalgas.

## **Patikimumas**

Paskutinis OTC modelio aspektas yra patikimumas.. Patikimumas susijęs su dviem savybėmis: įtikimumu ir patirtimi. Įtikimumas susijęs su tokiais savybėmis kaip tikrumas ar nešališkumas. Patirtis susijusi su patikimomis ir kompetentingomis savybėmis. Aukščiau paminėtas dvi patikimumo savybes galima pasiekti pasitelkiant patikimą šaltinį. Teksto patikimumas gali būti apskaičiuojamas naudojant bendrą informaciją tarp žodžių, t. y., analizuojant dažnai pasitaikančius daiktavardžius ir veiksmažodžius.

## **OTC modelis**

Objektyvumas ir tiesa gali būti apskaičiuojami naudojant teksto apdorojimo biblioteką „Text Blob“ („Python“). „Text Blob“ biblioteka naudojama natūralios kalbos apdorojimo užduotims atlikti, pavyzdžiui, sentimentų analizei, klasifikacijai, kalbos vertimui, žodžių žymėjimui ir daugeliui kitų funkcijų. „Text Blob“ pagalba apskaičiuojamas pateikto teksto subjektyvumas ir poliškumas. Subjektyvumas yra neigiama objektyvumo pusė OTC modelyje. Apskaičiuotas subjektyvumas yra intervale  $[0,0 ; 1,0]$ , kur 0,0 reiškia, kad tekstas yra labai objektyvus, o 1,0 reiškia, kad tekstas yra labai subjektyvus.

Poliškumas susijęs su teigiamomis, neigiamomis ir neutraliomis stebėtomis emocijomis tekste. Šios emocijos gali būti naudojamos atskiriant teisingą ir apgaulingą tekstus. Įrodymai, kad melagiai naudoja daugiau neigiamų emocijų nei sakantys tiesą, gali būti naudojami kaip priemonė, leidžianti rasti poliškumą tekste, kuris yra antroji tiesos/apgaulės dimensija. Poliškumas, apskaičiuotas „Text Blob“ bibliotekoje, intervale  $[-1,0 ; 1,0]$ , kur -1,0 yra labai neigiama emocija, 1,0 reiškia labai teigiamą emociją ir 0,0 reiškia neutralią emociją.

„Text Blob“ biblioteka priklauso nuo „NLTK“ ir „Pattern“ bibliotekų (žr. 2 priedo lentelės 3 eilutę). Pagal nutylėjimą „Text Blob“ naudoja „Pattern Analyzer“ analizatorių (žr. 2 priedo lentelės



1 eilutę). „Pattern Analyzer“ – tai jausmų analizatorius, kuris naudoja „Pattern“ biblioteką (žr. 2 priedo lentelės 4 eilutę). „Pattern“ yra „Python“ biblioteka, atliekanti duomenų gavybos operacijas, tokias kaip natūralios kalbos apdorojimas, mašininis mokymasis, tinklo analizė ir vizualizacija. Šią biblioteką sukūrė Tom De Smelt ir Walter Daelemans (žr. 2 priedo lentelės 2 eilutę).

Patikimumas apskaičiuojamas naudojant bendrą informaciją tarp dviejų žodžių. Pirma, žodžiai tokenizuojami<sup>1</sup> naudojant „NLTK“ žodžių tokenizavimo funkciją, tuomet skaičiuojama, kiek iš viso žodžių yra tekste ( $n$ ). Pasirinktos imties dydis yra  $n \cdot n$ . Toliau yra suskaičiuojama, kiek kartų tekste pasikartoja kiekvienas žodis. Tegul šie skaičiai bus  $n_1$  ir  $n_2$ .

$$\text{Bendra informacija} = \frac{\text{pirmo ir antro žodžio tikimybė}}{\text{pirmo žodžio tikimybė} \cdot \text{antro žodžio tikimybė}}. \quad (2.11)$$

Taigi,

$$\text{Bendra informacija} = \frac{n_1 \cdot \frac{2}{n^2} + n_2 \cdot \frac{2}{n^2}}{\frac{2n - 1}{n^2}}. \quad (2.12)$$

Galiausiai, visi trys OTC modelio matmenys yra sujungti į vieną teisingumo indeksą, normalizuojant matmenis į intervalą [0 ; 1], kur 1 yra didžiausias objektyvumas, teisingumas, patikimumas ir 0 yra minimalus objektyvumas, teisingumas ir patikimumas. Šis sudėtinis indeksas suteikia galimybę sistemingai vertinti didelius duomenų kokybės skirtumus duomenų rinkiniuose su tekstiniais duomenimis. Indeksas gali būti naudingas nustatant tas didelio duomenų rinkinio dalis, kurios yra mažesnės kokybės. Šias dalis galima išskirti, jei reikia pagerinti viso duomenų rinkinio kokybę [17].

### 2.3. Istorinio kintamumo vertinimas

Paprastai numanomas kintamumas yra palyginamas su istoriniu kintamumu. Nors numanomas kintamumas tik vienas, tačiau yra daugybė skirtingų istorinio kintamumo įvertinimo priemonių.

Numanomas kintamumas turi fiksuotą vertę. Tačiau nėra vieningo istorinio kintamumo skaičiavimo. Istorinių dienų skaičius keičia istorinio kintamumo skaičiavimo rezultatus. Tačiau neturėtų būti skirtumo tarp vidutinio dienos ar savaitės istorinio kintamumo. Yra nagrinėjami įvairaus istorinio nepastovumo skaičiavimo matai, tarp jų ir close-to-close bei standartinis

---

<sup>1</sup> Tokenizavimas – tai procesas, kurio metu teksto įvestis suskaidoma į žodžius ir terminus, kitaip vadinamais tokenais.

nuokrypis, taip pat pažangūs kintamumo matai, tokie kaip Parkinson, Garman-Klass (įskaitant Yang-Zhang pratęsimą), Rogers ir Satchell bei Yang-Zhang. Istoriniam nepastovumui reikia dviejų parametrų:

- laiko trukmės  $N$  (pvz., dienų / savaitės / mėnesių skaičiaus);
- matavimo dažnumo  $F$  (pvz., kasdien / kas savaitę).

- ***Istorinio kintamumo laiko trukmė***

Istorinio kintamumo dienų skaičiaus pasirinkimas nėra trivialus. Kai kurie investuotojai mano, kad geriausias istorinių kintamumo dienų skaičius yra toks pat, kaip ir numanomas palūkanų kintamumas, pvz., 1 mėnesio kintamumas turėtų būti lyginamas su 21 prekybos dienos istoriniu nepastovumu (ir 3 mėnesių kintamumas turėtų būti lyginamas su 63 dienų istoriniu kintamumu ir t. t.). Nors vienodos trukmės istorinis kintamumas yra naudingas, norint pasiekti realią minimalią ir maksimalią vertę per ilgesnį laiką, tačiau tai ne visada geriausias laikotarpis, nustatant teisingą ilgalaikio numanomo kintamumo lygį.

- ***Istorinio kintamumo dažnis***

Nors istorinis kintamumas gali būti matuojamas kas mėnesį, kas ketvirtį ar kasmet, paprastai jis vertinamas kasdien arba per savaitę. Dažniausiai kasdieniniam kintamumui yra teikiama pirmenybė lyginant su savaitiniu. Tačiau jei ilgalaikis nepastovumas yra nagrinėjamas tarp dviejų skirtingų rinkų, savaitės svyravimai gali būti geriausia priemonė sumažinti skirtingų valstybinių švenčių (ir prekybos valandų) įtaką. Jei akcijų kainos yra nepriklausomos, tuomet kasdienis ir savaitinis istorinis nepastovumas vidutiniškai turėtų būti vienodas. Jei akcijų kainos grąža nėra nepriklausoma, tai gali skirtis. Autokoreliacija yra koreliacija tarp dviejų skirtingų grąžų, todėl nepriklausoma grąža turi 0 % autokoreliaciją.

**Close-to-close matas**

Tai yra paprasčiausias ir labiausiai paplitęs matas, naudojantis uždarymo kainą (angl. *Closing Price*). Kintamumas apibrėžiamas kaip logaritminės grąžos metinis standartinis nuokrypis.

$$\text{Log grąža} = x_i = \text{Ln} \left( \frac{c_i + d_i}{c_{i-1}} \right), \quad (2.13)$$

kur  $d_i$  = įprasti dividendai, o  $c_i$  yra uždarymo kaina (angl. *Closing Price*).

Logaritminės grąžos standartinis nuokrypis užrašomas tokiu būdu:

$$x = s_x = \sqrt{\frac{F}{N}} \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2}, \quad (2.14)$$

čia  $\bar{x}$  = dreifas =  $E(x_i)$ .

Pagal Jensen'o nelygybę

$$\sigma = \sqrt{\sigma^2} = \sqrt{E(s^2)} < E(\sqrt{s^2}) = E(s), \quad (2.15)$$

tuomet

$$Kintamumas = \sigma_x = s_x \times \sqrt{\frac{N}{N-1}}. \quad (2.16)$$

Darome prielaidą, kad dreifas yra nulinis, tada:

$$\begin{aligned} Kintamumas_{close\ to\ close} = \sigma_{cc} &= \sqrt{\frac{F}{N-1}} \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} = \\ &= \sqrt{\frac{F}{N-1}} \sqrt{\sum_{i=1}^N \text{Ln}\left(\frac{c_i}{c_{i-1}}\right)}. \end{aligned} \quad (2.17)$$

### **Parkinson matas**

Parkinson matas – tai pirmasis pažangus kintamumo vertinimo matas, sukurtas 1980 m. Užtuot naudojęs uždarymo kainas, jis naudoja didžiausią (angl. *High Price*) ir mažiausią (angl. *Low Price*) kainas. Parkinson matas yra 5,2 karto efektyvesnis nei close-to-close. Nors kitos priemonės yra veiksmingesnės, kai kurie tyrimai parodė, kad tai yra geriausia priemonė tikriems empiriniams duomenims.

$$Kintamumas_{Parkinson} = \sigma_P = \sqrt{\frac{F}{N}} \sqrt{\frac{1}{4\text{Ln}(2)} \sum_{i=1}^N \left(\text{Ln} \frac{h_i}{l_i}\right)^2}. \quad (2.18)$$

Čia  $h_i$  ir  $l_i$  yra atitinkamai didžiausia ir mažiausia kainos.

### **Garman-Klass matas**

Vėliau, 1980 m., buvo sukurtas Garman-Klass nepastovumo vertinimas. Tai yra Parkinson pratėsimas, kuris apima atidarymo (angl. *Opening Price*) ir uždarymo kainas (jei atidarymo kainos nėra, gali būti naudojama ankstesnės dienos uždarymo kaina). Šis matas ignoruoja nakties šuolius, jeigu tokių pasitaiko.

$$\begin{aligned} Kintamumas_{Garman -Klass} &= \sigma_{GK} = \\ &= \sqrt{\frac{F}{N}} \sqrt{\sum_{i=1}^N \frac{1}{2} \left( \text{Ln} \frac{h_i}{l_i} \right)^2 - (2\text{Ln}(2) - 1) \left( \text{Ln} \frac{c_i}{o_i} \right)^2}. \end{aligned} \quad (2.19)$$

### **Rogers-Satchell matas**

Visos ankstesnės pažangiosios nepastovumo vertinimo priemonės numato, kad vidutinė grąža (arba dreifas) yra lygi nuliui. Vertybiniams popieriams, kurie turi dreifą arba nenulinį vidurkį, reikia sudėtingesnių nepastovumo priemonių. 1990 m. pradžioje sukurtas Rogers-Satchell kintamumo vertinimas sugeba tinkamai įvertinti vertybinių popierių (su nenuliniu vidurkiu) svyravimus. Kaip ir ankstesni, taip ir šis matas ignoruoja šuolius.

$$\begin{aligned} Kintamumas_{Rogers -Satchell} &= \sigma_{RS} = \\ &= \sqrt{\frac{F}{N}} \sqrt{\sum_{i=1}^N \text{Ln} \left( \frac{h_i}{c_i} \right) \text{Ln} \left( \frac{h_i}{o_i} \right) + \text{Ln} \left( \frac{l_i}{c_i} \right) \text{Ln} \left( \frac{l_i}{o_i} \right)}. \end{aligned} \quad (2.20)$$

### **Garman-Klass Yang-Zhang pratėsimas**

Yang-Zhang pakeitė Garman-Klass nepastovumo priemonę, kad leistų jai tvarkyti šuolius. Daroma prielaida, jog dreifas yra nulinis, todėl kintamumas bus pervertintas, jeigu vertybiniai popieriai turės nenulinio vidurkio grąžą.

$$\begin{aligned} Kintamumas_{GKYZ} &= \sigma_{GKYZ} = \\ &= \sqrt{\frac{F}{N}} \sqrt{\sum_{i=1}^N \left( \text{Ln} \left( \frac{o_i}{c_{i-1}} \right) \right)^2 + \frac{1}{2} \left( \text{Ln} \left( \frac{h_i}{l_i} \right) \right)^2 - (2\text{Ln}(2) - 1) \left( \text{Ln} \frac{c_i}{o_i} \right)^2}. \end{aligned} \quad (2.21)$$

## Yang-Zhang matas

2000 m. Yang-Zhang sukūrė nepastovumo priemonę, kuri neignoruotų šuolių ir dreifo. Tai yra close-to-open, open-to-close ir Rogers-Satchell kintamumo svertinio vidurkio suma [18].

$$\begin{aligned} Kintamumas_{Yang-Zhang} &= \sigma_{YZ} = \\ &= \sqrt{F} \sqrt{\sum_{i=1}^N \sigma_{close\ to\ open}^2 + k\sigma_{open\ to\ close}^2 + (1-k)\sigma_{RS}^2}, \end{aligned} \quad (2.22)$$

kur

$$k = \frac{0.34}{1.34 + \frac{N+1}{N-1}}$$

$$\sigma_{close\ to\ open}^2 = \frac{1}{N-1} \sum_{i=1}^N \left[ \ln\left(\frac{o_i}{c_{i-1}}\right) - \overline{\ln\left(\frac{o_i}{c_{i-1}}\right)} \right]^2, \quad (2.23)$$

$$\sigma_{open\ to\ close}^2 = \frac{1}{N-1} \sum_{i=1}^N \left[ \ln\left(\frac{c_i}{o_i}\right) - \overline{\ln\left(\frac{c_i}{o_i}\right)} \right]^2. \quad (2.24)$$

## Standartinis nuokrypis

Anksčiau buvo aprašyti pažangūs kintamumo skaičiavimo metodai. Pats lengviausias metodas kintamumui įvertinti yra tiesiog klasikinis standartinis nuokrypis.

$$\sigma = \sqrt{\frac{N}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (2.25)$$

### **2.4. Kintamumo matų paklaidos ir efektyvumas**

Tam, kad galėtume nustatyti, kuris kintamumo vertinimo matas yra tinkamiausias, reikia įvertinti šių matų tikslumą bei efektyvumą. Tikslumas vertinamas naudojant vidutinį absoliutinį nuokrypį (angl. *Mean Absolute Deviation, MAD*):

$$MAD = \frac{1}{N} \sum_{i=1}^N |\bar{\sigma} - \sigma_i|, \quad (2.26)$$

kur  $\sigma_i$  –  $i$ -os dienos kintamumas apskaičiuotas vienu iš aprašytų metodų, o  $\bar{\sigma} = E(\sigma_i)$ . Kintamumo matų efektyvumas vertinamas tokiu būdu:

$$Efektyvumas = \frac{D(\bar{\sigma}_S)}{D(\bar{\sigma}_K)}, \quad (2.27)$$

kur  $\bar{\sigma}_S$  – vidutinis kintamumas, apskaičiuotas naudojant standartinį nuokrypį, o  $\bar{\sigma}_K$  – vidutinis kintamumas, apskaičiuotas naudojant vieną iš pažangių kintamumo vertinimo matų. Jei santykis yra didesnis nei 1, tai reiškia, kad pažangaus mato dispersija yra mažesnė nei standartinio nuokrypio. Kuo didesnis santykis, tuo naudojamas matas yra efektyvesnis [19].

## 2.5. Programinės priemonės

Sparčiai didėjant įmonių, kaupiančių didžiuosius duomenis, skaičiui, o taip pat didėjant ir pačių duomenų kiekiui, vis svarbesnės tampa priemonės, kuriomis būtų galima apdoroti tokius duomenis. Kadangi duomenų apimtys yra milžiniškos, tikrai ne visos programos gali su tuo susitvarkyti. Šiuo metu yra kelios populiarios programavimo kalbos, kurios naudojamos dirbant su didžiais duomenimis:

1. Julia.
2. SAS.
3. Python.
4. R.
5. Scala.
6. HiveQL.
7. Pig Latin.
8. Go.

Plačiau apie kiekvieną iš pateiktų programų galima paskaityti [20].

### **Programavimo kalba R**

Šiame darbe tyrimas atliekamas naudojant programą R. Ši programa pasirinkta neatsitiktinai. Ji kaip ir Python yra viena iš labiausiai paplitusių atviro kodo kalbų dirbant su dideliais ir sudėtingais duomenų rinkiniais. Pastaraisiais metais abi programos labai išpopuliarėjo, be to jas abi palaiko

didelės ir naudingos bendruomenės. Tuo metu, kai Python spinduliuoja paprastumu ir patogumu, R išsiskiria savo pritaikomumu. Jos platus pritaikymas ir galimybė adaptuoti reiškia, kad tikriausiai mes kasdien vykdome kodą, parašytą pasitelkiant R, nes jis buvo naudojamas kuriant „Google“, „Facebook“, „Twitter“ ir daugybės kitų algoritmus [20].

R programavimo kalbos istorija prasidėjo 1976 m. garsiojoje „Bell“ laboratorijoje. Iki tol statistiniams skaičiavimams atlikti dažniausiai buvo naudojama Fortran kalbos biblioteka SCS (angl. *Statistical Computing Subroutines*). Šios bibliotekos galimybių užteko standartiniams skaičiavimams, tačiau net ir pačių paprasčiausių skaičiavimų nebuvo galima atlikti interaktyviai. Ir pagaliau 1993 m. Naujosios Zelandijos Oklando universitete Ross Ihaka ir Robert Gentleman sukūrė programavimo kalbą R, kuri skirta statistiniams skaičiavimams ir grafiniam duomenų atvaizdavimui. Jos pagrindas buvo funkcinio programavimo kalba Scheme (Lisp dialektas), tačiau ji turėjo S kalbos sintaksę. Ilgainiui R perėmė ir daug kitų S kalbos bruožų. Dabar galima teigti, jog R yra S kalbos dialektas arba, kitaip tariant, jos realizacija. Kalbant formaliai, R yra interpretuojama kaip dinaminės tipizacijos funkcinio programavimo kalba su objektinio programavimo galimybėmis. Pagal savo paskirtį R yra orientuota į statistinę duomenų analizę bei duomenų vizualizavimą, tačiau vien tuo neapsiriboja. R galimybes labai išplečia didžiulis bibliotekų, kurios leidžia spręsti įvairius skaičiavimo uždavinius, kiekis. Šiuo metu CRAN (angl. The Comprehensive R Archive Network) turi daugiau nei 8000 bibliotekų. Skirtingai nei S kalba, kuri yra komercinė, R yra GNU projekto dalis, kuri taip pat yra nemokama ir laisvai platinama. Jos vystymu rūpinasi R fondas (angl. *The R Foundation*) [21].

### 3. Tiriamoji dalis

Tiriamojame darbo dalyje atliekamas tyrimas su „Microsoft“ korporacijos akcijų kainų duomenimis ir, remiantis esama teorija, vykdomos darbo užduotys. Duomenys ir gauti rezultatai interpretuojami grafiškai, apskaičiuojamos pagrindinės didžiųjų duomenų charakteristikos, atliekama gautų įverčių analizė, aprašomas kintamumo charakteristikos panaudojimas bei svarba versle.

#### 3.1. Duomenys

Tyrime naudoti duomenys apima 1986-03-13—2018-03-13 laikotarpį, kuriuo analizuojamos „Microsoft“ kiekvienos dienos atidarymo, didžiausia, mažiausia ir uždarymo akcijų kainos. Duomenys paimti iš „Bloomberg“ duomenų bazės.

„Microsoft Corporation“ — tai JAV programinės ir techninės įrangos gamintojas. Kompanija „Micro-soft“ buvo įkurta 1975 m. Albuquerkeje, Naujosios Meksikos valstijoje. Ją įkūrė Bill Gates ir Paul Allen, prieš tai jau kartu turėję kompaniją „Traf-O-Data“. „Microsoft Corporation“ pavadinimas kilo iš žodžių junginio *Microcomputer Software*. 1976 m. lapkričio 26 d. pavadinimas „Microsoft“ (be brūkšnelio) tapo registruotu prekės ženklu, o nuo 1986 m. kovo 13 d. kompanijos akcijos tapo viešomis [22]. Būtent todėl duomenys turimi nuo šios dienos.

Pagrindinės duomenų aprašomosios statistikos pateiktos žemiau esančioje lentelėje.

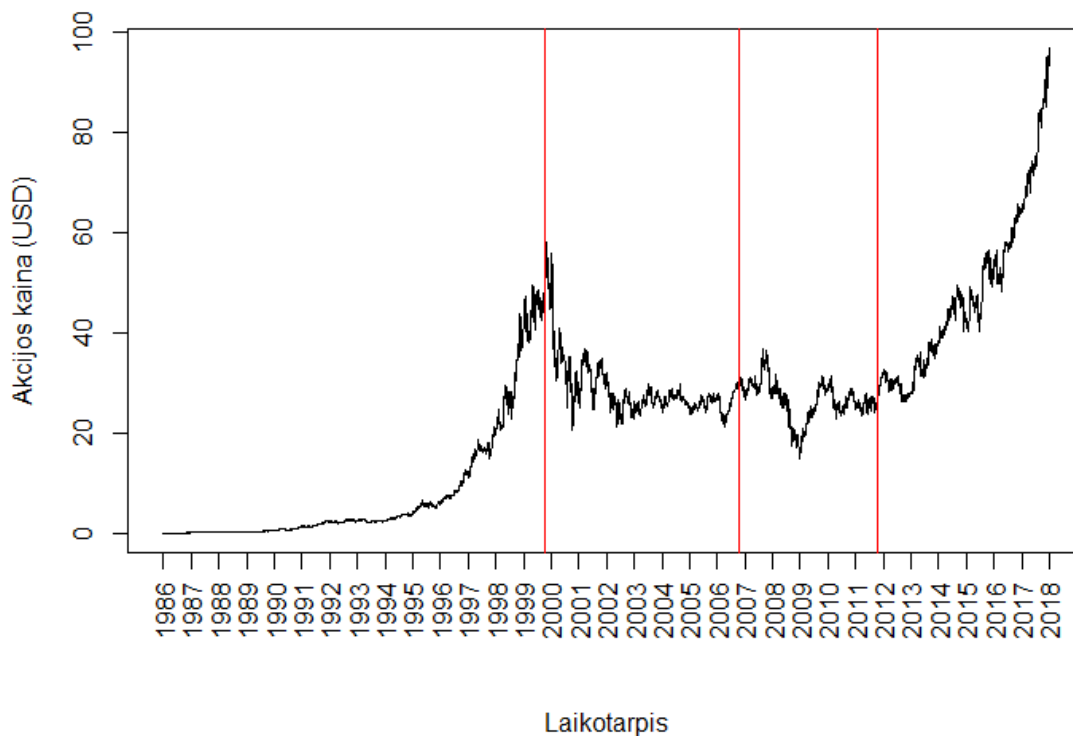
1 lentelė. Aprašomosios duomenų statistikos

Akcijų kainos	Duomenų intervalas	Matavimo vienetai	Duomenų sklaida	Vidurkis	Dispersija	Moda	Mediana
Atidarymo	1986-03-13–2018-03-13	USD	[0,08854 ; 97]	23,08	355,31	2,66	25,73
Didžiausia	1986-03-13–2018-03-13	USD	[0,09201 ; 97,24]	23,34	361,87	27,50	26
Mažiausia	1986-03-13–2018-03-13	USD	[0,08854 ; 96,04]	22,83	348,70	25,50	25,47
Uždarymo	1986-03-13–2018-03-13	USD	[0,09028 ; 96,77]	23,09	355,53	0,36	25,72



Didžiųjų duomenų charakteristikų įvertinimui bus aktualios tiek atidarymo, tiek uždarymo bei didžiausia ir mažiausia akcijų kainos.

Kadangi visos kainos (atidarymo, didžiausia, mažiausia ir uždarymo) tarpusavyje skiriasi ne tiek, jog grafike būtų aiškiai matomi skirtumai, todėl braižomas tik dienos uždarymo akcijų kainų kitimas bėgant laikui.



6 pav. Dienos uždarymo akcijų kainų grafikas

Kaip matome, „Microsoft“ akcijų istorija susideda iš trijų aiškių laikotarpių. Nuo 1980 m. iki 2000 m. technologijų proveržis turėjo įtakos „Microsoft“ akcijų kainų augimui. Per laikotarpį nuo 1987 m. iki 1999 m. naudojimasis kompiuteriu tapo populiarus tiek verslo pasaulyje, tiek asmeniniame gyvenime, o technologinių naujovių kūrimas prisidėjo prie naujų programų kompiuteriams atsiradimo. Ne mažiau svarbus buvo interneto kilimas ir dėl to atsiradęs įrangos, kuri atitiktų duomenų siuntimo ir gavimo tinklais reikalavimus, poreikis.

Po technologijų krizės 2000—2002 m. „Microsoft“ augimo perspektyvos smarkiai sumažėjo. Bendrovė ir toliau rėmėsi savo „Windows“ operacinės sistemos programine įranga ir „Office“ programų paketu, tai padėjo pasiekti pastovias pajamas. Vis dėlto pastangos neatsilikti nuo savo konkurentų nebuvo visiškai sėkmingos ir „Microsoft“ kitoms bendrovėms užleido pirmaujančios

kompanijos vietą, tokiose srityse kaip nešiojamų muzikos grotuvų, išmaniųjų telefonų ir planšetinių kompiuterių gamyba.

Sparčiausio augimo laikotarpiui pasibaigus, „Microsoft“ pradėjo daryti žingsnius, kurie atspindėjo ją kaip labiau subrendusią įmonę. 2003 m. ji pradėjo mokėti dividendus, o laikui bėgant tai žymiai padidino ketvirčio išmokas. „Microsoft“ dividendų istorija rodo, kad kompanija dosniai dalinasi savo kapitalu su investuotojais.

Jau nuo 2012 m. kompanijos situacija ėmė gerėti, o nuo 2013 m. „Microsoft“ įžengė į savo naujausią akcijų kainų istoriją, kuri buvo pelninga jos akcininkams. Akcijų kaina per mažiau nei ketverius metus išaugo daugiau nei dvigubai, o didelė pažanga, kurią padarė „Microsoft“, atsirado dėl noro judėti naujomis kryptimis.

Pavyzdžiui, „Microsoft“ pakeitė ilgalaikį nenorą pasinaudoti mobilia revoliucija ir pradėjo skirti didesnę dėmesį skaičiavimams debesyse ir mobiliuosiuose telefonuose. Viena iš pagrindinių iniciatyvų buvo „Office“ programinės įrangos paketo pritaikymas mobiliam formatui, leidžiantis kitų gamintojų įrenginių vartotojams turėti prieigą prie programinės įrangos produktų. Tuo pat metu „Microsoft“ taip pat stengėsi vystyti „Windows“ link labiau mobiliesiems ir debesims palankios operacinės sistemos. „Windows 10“ pagalba buvo siekiama sukurti bendrą platformą visiems „Microsoft“ įrenginių naudotojams.

Tuo tarpu kai kurie investuotojai skeptiškai žiūrėjo į tai, kad „Microsoft“ stengiasi augti per įsigijimus. Vienas iš naujesnių pirkinių, kuriuos įsigijo JAV technologijų milžinė, buvo socialinis tinklas „LinkedIn“ (2016 m.). Šis pirkinytis kompanijai atsiėjo 26,2 mlrd. JAV dolerių. Ankstesni „Microsoft“ pirkiniai buvo „Nokia“ mobiliųjų telefonų gamybos padalinys (2013 m.), o taip pat ir interneto pokalbių programa „Skype“ (2011 m.). Deja, bet kompanijai šie pirkiniai neatnešė tokio pelno, kurio tikėjosi investuotojai.

Sėkmingiau „Microsoft“ sekasi bendradarbiauti. Bendradarbiaudama su „Samsung“, „Lenovo“ ir kitais mobiliųjų įrenginių kūrėjais, „Microsoft“ pasirūpino, kad pagrindinės programos būtų iš anksto įdiegtos išmaniuosiuose telefonuose ir planšetiniuose kompiuteriuose, kuriuose veikia „Android“ operacinė sistema. Tai netrukdo bendrovėms pasirinkti alternatyvų iš „Google“ padalinio. Tačiau atsižvelgiant į tai, kad „Google“ turėjo konkurencinį pranašumą turėdama operacinę sistemą „Android“ telefonuose, „Microsoft“ gebėjimas mesti iššūkį „Google“ yra geras pavyzdys, kaip „Microsoft“ stengiasi kovoti dėl „vietos po saule“ [23].

## **Laikotarpiai**

Kai kurias charakteristikas įdomu apžvelgti skirtingais „Microsoft“ akcijų kainų augimo arba kritimo laikotarpiais. Dėl šios priežasties iš pradžių duomenys bus išskaidomi į keturis laikotarpius:

- 1) 1986-03-13–1999-12-31. Kaip ir buvo anksčiau minėta, šiuo laikotarpiu sparčiai vystėsi technologijos, o tai turėjo įtakos „Microsoft“ akcijų kainų augimui.
- 2) 2000-01-01–2006-12-31. Būtent po 2000–2002 m. technologijų krizės „Microsoft“ akcijų kainos pradėjo kristi.
- 3) 2007-01-01–2011-12-31. Po technologijų krizės akcijų kainos krito iki pat 2011 m., tačiau būtent 2007–2011 m. yra labai įdomūs ir išskirtiniai, dėl 2007–2008 m. vykusios ekonominės krizės, kuri paveikė visą pasaulį. Šiuo laikotarpiu „Microsoft“ akcijų kainos staigiai nukrito.
- 4) 2012-01-01–2018-03-13. Nuo 2012–2013 m. „Microsoft“ ėmė atsigausti ir akcijų kainos pradėjo sparčiai augti.

Aprašytos charakteristikos vertinamos aukščiau aprašytais laikotarpiais, išskyrus kintamumą. Pastarąją charakteristiką vertinsime ne tik pagal akcijų kainų svyravimus sudarytais laikotarpiais, bet ir laikotarpiais, kurie gauti atsižvelgiant į pačio kintamumo svyravimus. Pirmiausiai apskaičiuosime kasmetinį kintamumą ir pagal jo svyravimus sudarysime tyrimo laikotarpius.

### **3.2. Duomenų spartos vertinimas**

Duomenys tyrimui paimti iš „Bloomberg“ duomenų bazės. „Bloomberg“ – tai informacinė platforma, leidžianti analizuoti rinką, atlikti įvairius tyrimus bei priimti sprendimus, remiantis realaus laiko arba istoriniais duomenimis. „Bloomberg“ platformoje akcijų kainų duomenys yra atnaujinami vos tik kainos pasikeičia. Tai yra labai svarbu investuotojams, kurie kiekvieną sekundę atidžiai seka akcijų kainas, nes nori jas pirkti arba parduoti, esant palankiai kainai. Kitaip sakant, akcijų kainų duomenys yra prieinami realiu laiku. Akcijų rinkoje svarbi kiekviena sekundė, kuri gali lemti tam tikrus sprendimus. Būtent akcijų kainos gali keistis nors ir nežymiai, bet labai greitai. Investuotojams svarbios kainos čia ir dabar. Kokia kaina buvo prieš kelias sekundes jau nėra aktualu, nes ji per tą laiką spėjo pasikeisti.

Jeigu iš pradžių tarsi apžvelgėme, per kiek laiko atsinaujina arba atsiranda nauji akcijų kainų duomenys, tai galime apžvelgti ir iš kitos pusės: kiek įrašų (akcijų kainų eilučių, kurios sudarytos iš dienos atidarymo, didžiausios, mažiausios ir uždarymo kainų ) yra užfiksuojama per metus skirtingais laikotarpiais. Šis įrašų skaičius bus lygus prekybos dienų skaičiui. Rezultatai patekti žemiau esančioje lentelėje.

2 lentelė. Vidutinis įrašų skaičius per metus

Laikotarpis	Dienų skaičius	Apytikslis metų skaičius	Įrašų skaičius	Vidutinis įrašų skaičius per metus
1986-03-13– 1999-12-31	5 041	13,81	3 490	252,86
2000-01-01– 2006-12-31	2 556	7	1 759	251,29
2007-01-01– 2011-12-31	1 825	5	1 260	252
2012-01-01– 2018-03-13	2 263	6,2	1 558	251,29
<i>1986-03-13– 2018-03-13</i>	<i>11 688</i>	<i>32</i>	<i>8 067</i>	<i>252,09</i>

Logiška, jog per ilgesnį laikotarpį yra užfiksuojama daugiau įrašų. Kaip matome, visais laikotarpiais paskaičiavus, kiek vidutiniškai yra įrašų (arba prekybos dienų) per metus, buvo gauta maždaug 252 įrašai. Taigi galime sakyti, jog per metus vidutiniškai yra 252 prekybos dienos.

Kaip ir buvo minėta bendrojoje dalyje, pagrindiniai duomenų spartos aspektai yra judantys duomenys, analizė realiu laiku ir naudingumo trukmė. Akcijų kainų duomenys yra greitai besikeičiantys, pasiekiami realiu laiku, o jų naudingumo trukmė iš tiesų labai trumpa, jau praėjus keletui sekundžių duomenis galima vadinti pasenusiais.

### 3.3. Entropijos ir informacijos kiekio įvertinimas

Pirmoji charakteristika, kurią vertinsime yra duomenų apimtis. Šiame darbe ji buvo susieta su entropija ir informacijos kiekiu. Pirmiausia šiuos matavimus apskaičiuosime bendrai visam duomenų rinkiniui neatsižvelgiant į suskaidymą laikotarpiais. Kadangi skaičiuojant entropiją logaritmo pagrindu buvo pasirinktas dvejetas, tai, šiuo atveju, tiek entropijos, tiek informacijos kiekio matavimo vienetai yra bitai. Gauta, jog entropija visiems akcijų tipams yra beveik vienoda ir lygi  $\sim 0,0016$  bito. Apskaičiavus informacijos kiekį gauta, jog dienos atidarymo akcijų kainoms užkoduoti reikia  $\sim 11,6626$  bitų, didžiausioms –  $\sim 11,6303$  mažiausioms –  $\sim 11,6343$  bitų, o uždarymo kainoms  $\sim 11,7425$  bitų. Toliau tie patys matai buvo apskaičiuoti gautais laikotarpiais.

3 lentelė. Entropijos ir informacijos kiekio reikšmės skirtingais laikotarpiais

Laikotarpis	Entropija (bitai)	Informacijos kiekis (bitai)
1986-03-13–1999-12-31	~0,0034	[10,3564 ; 10,5472]
2000-01-01–2006-12-31	~0,0061	[9,7177 ; 9,8400]
2007-01-01–2011-12-31	~0,0082	[9,3634 ; 9,4189]
2012-01-01–2018-03-13	~0,0068	[10,2307 ; 10,2599]

Shannon entropija nurodo, koks yra minimalus bitų skaičius, reikalingas norint užkoduoti vieną elementą dvejetainėje formoje (jei logaritmo pagrindas yra 2) arba kitaip tariant, entropija – tai vieno elemento (vienos akcijų kainos) nešamas informacijos kiekis. Atsižvelgiant į 3 lentelėje pateiktus rezultatus, visiems akcijų kainų tipams didžiausios Shannon entropijos reikšmės gautos laikotarpiu 2007-01-01–2011-12-31 ir lygios maždaug 0,0082 bito. Mažiausios reikšmės lygios ~0,0034, gautos laikotarpiu 1986-03-13–1999-12-31. Apskaičiavus informacijos kiekį, didžiausios reikšmės stebimos 1986–1999 m., t. y., ilgiausiu laikotarpiu, o mažiausios – trumpiausiu (2007–2011 m.). Gauti rezultatai puikiai atspindi tai, jog kuo daugiau reikšmių, tuo daugiau reikia bitų joms užkoduoti ir atvirkščiai.

#### 3.4. Duomenų teisingumo vertinimas

Kita charakteristika, kuri bus įvertinta – tai duomenų teisingumas. Duomenų teisingumas susijęs su tikslumu. Tačiau, kai kalbama apie didžiųjų duomenų tikslumą, tai yra ne tik pačių duomenų kokybė, bet ir duomenų šaltinio patikimumas. Labai svarbu, kad duomenys būtų kuo kokybiškesni, t. y., kad juose būtų kuo mažiau dubliavimosi, įvedimo klaidų, praleistų reikšmių ir panašiai. „Nešvarių“ duomenų analizę dažniausiai lydi klaidingos išvalgos ir dėl to blogai suformuoti sprendimai. Visa tai gali įmonei sukelti rimtų problemų, kurių sprendimui reikės papildomų išlaidų, todėl svarbu iš anksto pasirūpinti analizuojamų duomenų kokybe.

Šiame darbe nagrinėjamų duomenų teisingumą įvertinti yra gana sudėtinga. Tačiau akivaizdu, jog nors laikotarpio nuo 1986-03-13 iki 2018-03-13 duomenys yra dieniniai, bet akcijų kainos yra fiksuotos tikrai ne kasdien. Toliau vertinant duomenų spartą, buvo apskaičiuota, jog per metus yra maždaug 252 prekybos dienos iš 365. Tai reiškia, jog vidutiniškai tiek akcijų kainų įrašų yra užfiksuojama per metus, na o 113-114 dienų akcijų kainos nėra fiksuojamos.

Versle prekybos diena yra laikotarpis, per kurį veikia konkreti vertybinių popierių birža. Prekybos dienos paprastai yra nuo pirmadienio iki penktadienio. Pasibaigus prekybos dienai, visi akcijų prekybos rezultatai uždaromi ir yra įšaldomi iki kitos prekybos dienos pradžios. Yra keletas kitų ypatingų aplinkybių, dėl kurių prekybos diena yra sutrumpinama arba prekyba iš vis nevykdoma.

Skelbiama, jog daugelyje vertybinių popierių biržų vidutiniškai yra apie 252 prekybos dienos per metus (tokį patį rezultatą gavome apskaičiavę vidutinį „Microsoft“ korporacijos akcijų kainų įrašų skaičių per metus). Tai yra:  $365,25$  (vidutinis dienų skaičius per metus) \*  $5/7$  (proporcingai darbo dienos per savaitę) =  $260,89$  - 9 (šventės) =  $251,89 \sim 252$ . Jungtinėse Amerikos Valstijose šventinių dienų, per kurias neveikia vertybinių popierių biržos, ir sutrumpintų prekybos dienų datos pateikiamos 3 priede [24].

Be praleistų dėl šventinių dienų akcijų kainų įrašų, nagrinėjamuose duomenyse yra ir keletas dienų, kuriomis akcijų kainos nebuvo užfiksuotos dėl nežinomų priežasčių. Tokios praleistos reikšmės jau kiek iškreipia duomenų kokybę. Taip pat duomenyse gali būti įvedimo ar kitokių klaidų ir netikslumų susijusių su užfiksuotomis dienos atidarymo, didžiausia, mažiausia bei uždarymo kainomis, tačiau tai patikrinti yra sudėtinga.

### **3.5. Duomenų kintamumo vertinimas**

#### **Kintamumo vertinimas laikotarpiais, sudarytais pagal akcijų kainų svyravimus**

Kaip ir buvo minėta anksčiau, šiame darbe daugiausiai dėmesio yra skiriama kintamumo charakteristikai. Vertinant „Microsoft“ akcijų kainų kintamumą, atsižvelgiama į duomenų aprašyme nurodytus laikotarpius, t. y. ši didžiųjų duomenų charakteristika įvertinama keturiais skirtingais laikotarpiais. Kuo didesnis kintamumas, tuo akcijų kainos yra nepastovesnės, o tai investuotojams reiškia didesnę riziką. Įsigydami akcijas investuotojai prisiima riziką patirti nuostolius dėl nepalankių akcijų kainos pokyčių rinkoje. Akcijų kainos kritimą gali lemti tiek neigiami bendrovės turto vertės ar pelningumo pokyčiai, tiek bendros akcijų rinkos tendencijos regione ir pasaulyje. Todėl labai svarbu žinoti akcijų kintamumą esant kainų kilimui ir kritimui. Žinodami šią informaciją, investuotojai yra pasiruošę netikėtiems kainų šuoliams, esant tam tikroms tendencijoms.

Pirmiausia kintamumas ištirtas laikotarpiu nuo 1986-03-13 iki 1999-12-31, tuomet 2000-01-31–2006-12-31, 2007-01-01–2011-12-31 ir galiausiai 2012-01-01–2018-03-13. Kintamumas buvo vertinamas septyniais skirtingais matais, tarp jų: standartinis nuokrypis, close-to-close, Garman-Klass, Parkinson, Roger-Satchell, Garman-Klass Yang-Zhang pratęsimas ir Yang-Zhang matai. Iš tiesų, standartinis nuokrypis yra skaičiuojamas tam, kad būtų galima apskaičiuoti pažangesnių kintamumo matų efektyvumą. Gautos kintamumo reikšmės išreiškiamos procentais. Rezultatai pateikiami lentelėse, bet kadangi visiems nagrinėjamiems laikotarpiams lentelės yra analogiškos tik su skirtingomis reikšmėmis, dėl to žemiau pateikiama tik anksčiausio laikotarpio rezultatų lentelė. Likusias lenteles su rezultatais galima rasti 4 priede.

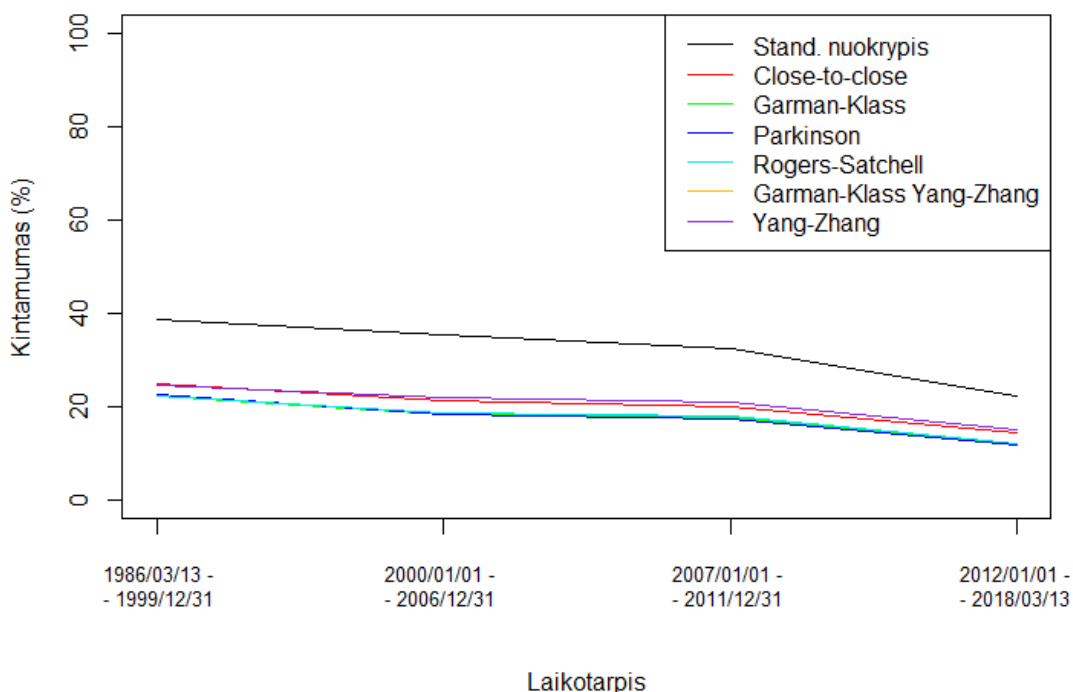
4 lentelė. Kintamumo įvertinimas laikotarpiu 1986-03-13–1999-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986-03-13–1999-12-31	38,66 %	24,89 %	22,33 %	22,69 %	22,25 %	24,49 %	24,62 %	RS, GK, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986-03-13–1999-12-31	0,1798	0,0819	0,0528	0,0566	0,0528	0,0601	0,0609	RS, GK, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986-03-13–1999-12-31		4,02	9,30	8,30	9,69	7,80	7,77	RS, GK, PARK, GKYZ, YZ, CC

Kaip matome iš rezultatų, pateiktų lentelėse, visais atvejais didžiausias kintamumas buvo gautas naudojant standartinį nuokrypį. Taip yra todėl, kad būtent šis kintamumo vertinimo metodas yra pats paprasčiausias ir mažiausiai tikslus. Pastaroji savybė atsispindi apskaičiuavus vidutinį absoliutinį nuokrypį (MAD). Visais keturiais atvejais standartinio nuokrypio MAD yra didžiausias. Taip pat galime pastebėti, jog mažiausios kintamumo reikšmės yra gautos naudojant tris matus – Rogers-Satchell, Parkinson ir Garman-Klass. Būtent šie trys matai pakaitomis dalinasi tris pirmąsias vietas. Tai rodo tiek gauti kintamumo įvertinimai, tiek MAD ir, žinoma, efektyvumas. Šie trys matai efektyvumu pastebimai lenkia kitus.

Lyginant tarpusavyje skirtingais laikotarpiais gautus rezultatus, galime pastebėti, jog didžiausias kintamumas buvo gautas laikotarpiu nuo 1986-03-13 iki 1999-12-31. Tokį rezultatą galėjo lemti technologijų proveržis, didelis kompiuterių poreikis ir „Microsoft“ vystymasis, todėl „Microsoft“ akcijų kainos per šiuos metus žymiai pasikeitė. Kompanijos akcijų kainos augimui įtakos turėjo „MS-DOS“ (angl. *Microsoft Disk Operating System*) operacinės sistemos sukūrimas. Tačiau žymesnis kainos padidėjimas pastebimas po to, kai buvo sukurtos „Windows 3.0“ (1990 m.), „Windows 95“ (1995 m.) ir „Windows 98“ (1998 m.) operacinės sistemos. Šios operacinės sistemos pasirodė būtent tokiu laikotarpiu, kai internetas bei kompiuterių naudojimas buvo itin populiarus, todėl šie „Microsoft“ produktai labai pasiteisino.

Abiem nuosmukio — technologijų krizės ir ekonominės krizės — laikotarpiais kintamumo reikšmės labai panašios ir gerokai didesnės, lyginant su „Microsoft“ atsigavimo laikotarpiu (2012–2018 m.). Kainos pradėjo kristi ir tapo nepastoviomis dėl ekonominės krizės metu sulėtėjusios viso pasaulio ekonomikos, o taip pat suprastėjo bendrovių veiklos perspektyvos ir galimybės gauti anksčiau prognozuotą pelną. Tokia situacija nulėmė tai, kad nemaža dalis investuotojų traukėsi iš vertybinių popierių rinkos, o dėl to akcijų kainos krito dar labiau. Investuoti tapo labai rizikinga ir nepalanku. Tai įvertinę investuotojai ieškojo palankesnių rinkų investuoti. Žemiau pateikiamas grafikas, kuriame atvaizduojamos skirtingais kintamumo matais visais laikotarpiais gautos kintamumo reikšmės.

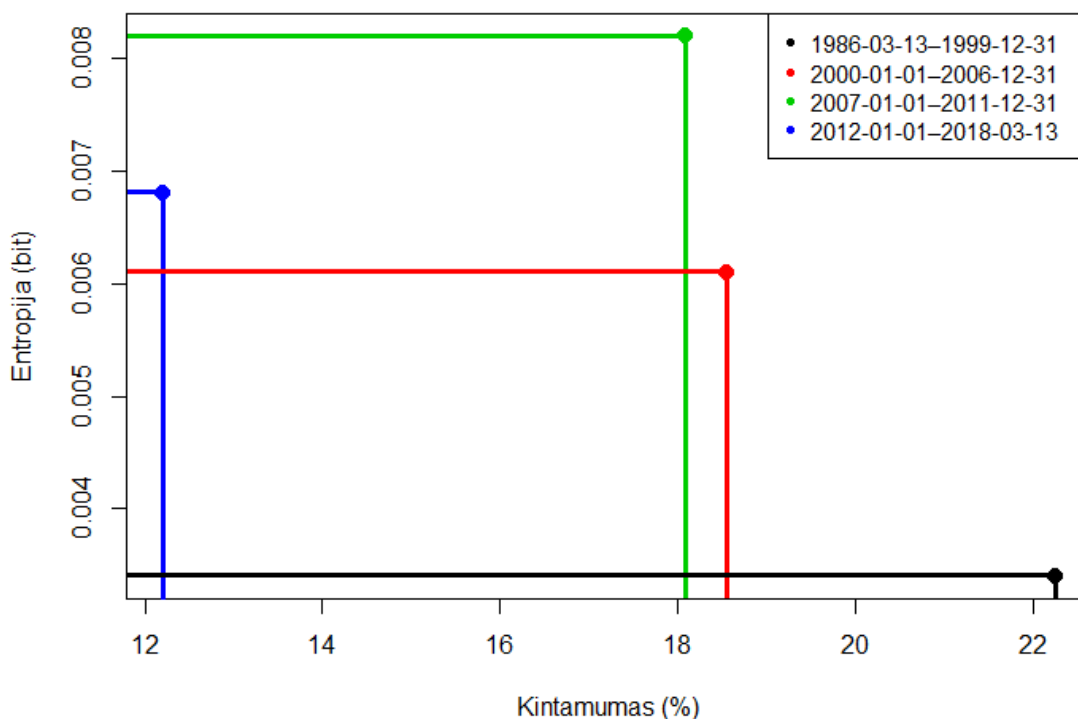


7 pav. Kintamumas keturiais skirtingais laikotarpiais

Iš grafiko matome, kad visais laikotarpiais skirtingais kintamumo matais gauti įverčiai elgiasi vienodai, t. y. visi kartu sumažėja arba padidėja, todėl galima sakyti, jog matai yra suderinti. Taip pat iš grafiko bei rezultatų, pateiktų 4 priede 11 lentelėje, matome, kad „Microsoft“ korporacijos atsigavimo laikotarpiu (nuo 2012 m.) kintamumo ir MAD reikšmės yra pačios mažiausios. Tai rodo, jog kainos tapo ne tokios „šokinėjančios“, bet jos tendencingai ir sparčiai augo, kad per mažiau nei ketverius metus akcijų kaina išaugo daugiau nei dvigubai ir toliau nenustoja augti.

Taip pat buvo nubraižytas kombinuotas dviejų charakteristikų grafikas, kur  $x$  ašyje yra pateikiamas kintamumas nagrinėjamais laikotarpiais, o  $y$  ašyje – atitinkamos entropijos reikšmės.





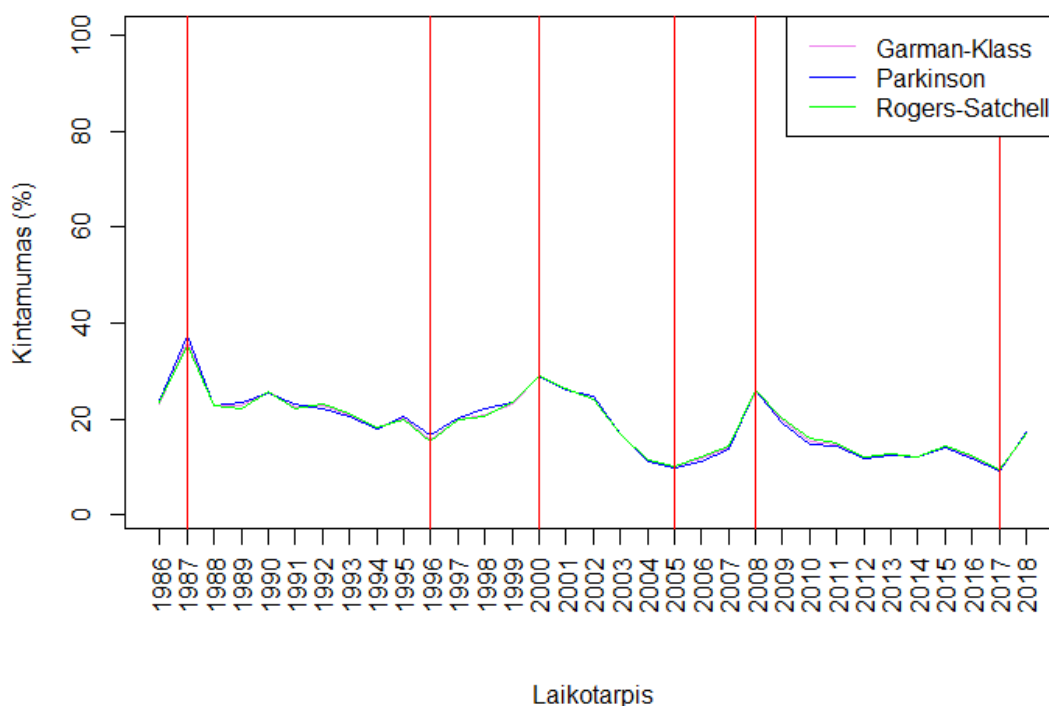
8 pav. Kombinuotas kintamumo ir entropijos grafikas

Kaip matome iš 8 paveiksle pateikto grafiko, esant didžiausiam kintamumui, entropijos reikšmė yra mažiausia, o esant mažiausiam kintamumui, entropijos reikšmė – viena iš didžiausių. Tai reiškia, jog ilgiausiu laikotarpiu kintamumas yra didžiausias, o vienu iš trumpiausių laikotarpių priešingai – kintamumas yra mažiausias. Šis grafikas dar kartą atspindi anksčiau gautus rezultatus.

### **Kintamumo vertinimas laikotarpiais, sudarytais pagal kasmetinius kintamumo svyravimus**

Kaip ir buvo minėta anksčiau, kintamumo charakteristika bus vertinama ne tik laikotarpiais, sudarytais pagal korporacijos akcijų kainų istorinį kitimą, bet ir laikotarpiais, sudarytais pagal pačio kintamumo svyravimus.

Taigi, toliau sudarysime naujus laikotarpius, kuriais nagrinėsime kintamumą. Pirmiausia yra įvertinamas kasmetinis kintamumas. Tai atliksime naudodami tris metodus – Garman-Klass, Parkinson bei Rogers-Satchell, kurie prieš tai atlikto tyrimo metu parodė geriausius rezultatus. Gauti rezultatai pateikti lentelėje, esančioje 5 priede, o rezultatų grafikas pateikiamas 7 paveiksle.



9 pav. Kasmetinis kintamumas gautas naudojant dieninius duomenis (1986–2018 m.)

Galime pastebėti, jog trimis skirtingais metais gautos kasmetinės kintamumo reikšmės tarpusavyje beveik nesiskiria. Taip pat matome, jog yra aiškiai mažėjanti kintamumo tendencija. Pagal grafike stebimus kitimo svyravimus (augimą bei kritimą) galime išskirti tokius laikotarpius:

- 1) 1986–1987 m. – augimas.
- 2) 1988–1996 m. – kritimas.
- 3) 1997–2000 m. – augimas.
- 4) 2001–2005 m. – kritimas.
- 5) 2006–2008 m. augimas.
- 6) 2009–2017 m. – kritimas.
- 7) 2018 m. – augimas.

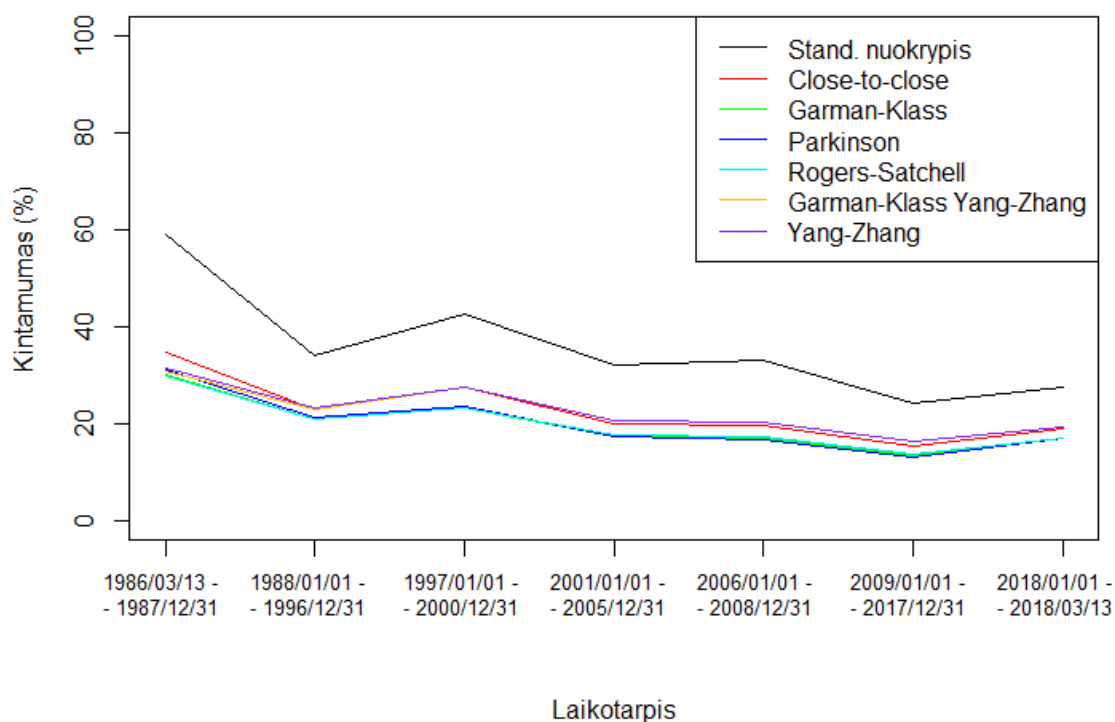
Kaip matome, gauti laikotarpiai skiriasi nuo anksčiau pagal akcijų kainų svyravimus sudarytų laikotarpių. Tokie rezultatai nestebina, nes vertinama pagal skirtingus kriterijus: vienu atveju vertinama pagal akcijų kainas (jų svyravimą), kitu – pagal tų pačių kainų kintamumo kitimą. Iš gautų rezultatų matyti, kad 2018 m. stebimas išaugęs kintamumas lyginant su prieš tai buvusiais metais. Tačiau reikia turėti omenyje, jog 2018 m. turimi ne visų metų dieniniai duomenys, o tik dviejų su puse mėnesių duomenys, todėl per likusius metus situacija gali pasikeisti.

Toliau gautais laikotarpiais įvertinamas akcijų kainų kintamumas. Kaip ir prieš tai buvusi atveju, vertinama septyniais skirtingais matais, apskaičiuojamas vidutinis absoliutinis nuokrypis bei naudotų matų efektyvumas. Anksčiausio laikotarpio lentelė su gautais rezultatais pateikiama žemiau, o likusių laikotarpių rezultatai – 6 priede.

5 lentelė. Kintamumo įvertinimas laikotarpiu 1986-03-13–1987-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986-01-01–1987-12-31	58,94 %	34,95 %	30,09 %	31,25 %	30,01 %	31,02 %	31,54 %	RS, GK, GKYZ, PARK, YZ, CC, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986-01-01–1987-12-31	0,2566	0,1389	0,0964	0,1026	0,0930	0,1038	0,1028	RS, GK, PARK, YZ, GKYZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986-01-01–1987-12-31		2,72	6,32	5,61	7,16	5,52	5,76	RS, GK, YZ, PARK, GKYZ, CC

Iš gautų rezultatų galime pastebėti, jog didžiausias kintamumas buvo gautas anksčiausiu laikotarpiu, o mažiausias – laikotarpiu 2009–2017 m. Pirmu atveju, įtakos turėjo tai, kad „Microsoft“ korporacija buvo labai „jauna“ ir besivystanti, taip pat šios korporacijos akcijos vertybinių popierių biržoje pasirodė tik 1986 m. o tai stipriai prisidėjo prie kainų svyravimų. Antruoju atveju, įtaką darė, tai, kad pasaulis pradėjo atsigauti po ekonominės krizės, o pati „Microsoft“ korporacija šiuo laikotarpiu priėmė svarbius sprendimus, kurie lėmė spartų akcijų kainų augimą ir nedidelį jų svyravimą (kritimą ir augimą). Panašūs rezultatai buvo gauti ir tiriant kintamumą, duomenis suskirsčius į laikotarpius pirmuoju būdu. Čia didžiausias kintamumas buvo gautas 1986–1999 m., o mažiausias 2012–2018 m. Toliau grafiškai atvaizduojamos skirtingais matais gautos kintamumo reikšmės.

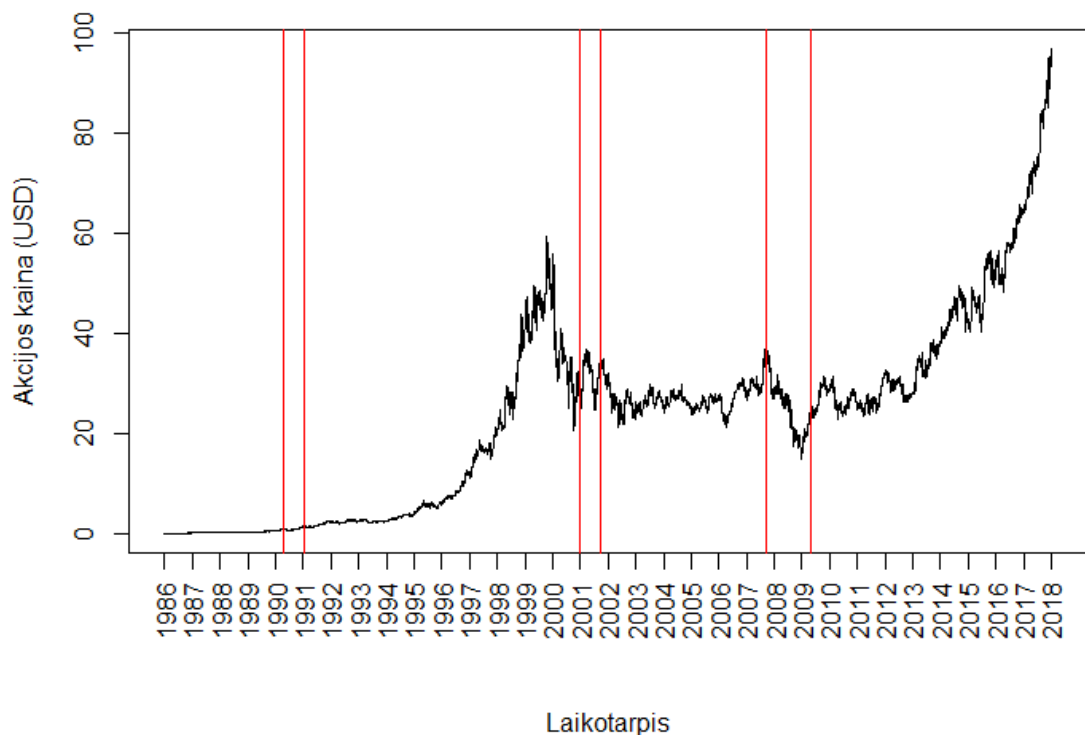


10 pav. Kintamumas skirtingais laikotarpiais

Tiek iš grafiko, tiek iš rezultatų, pateiktų lentelėse, matome, jog beveik visais laikotarpiais mažiausias kintamumas buvo gautas Garman-Klass, Parkinson ir Rogers-Satchell matais. Taip pat dažniausiai šių matų vidutiniai absoliutiniai nuokrypiai buvo mažiausi ir patys matai efektyviausi. Grafike ryškiai išsiskiria standartinio nuokrypio pagalba apskaičiuotas kintamumas, kuris gerokai didesnis lyginant su kitais matais gautais įvertinimais.

### **Kintamumo vertinimas laikotarpiais, sudarytais atsižvelgiant į Amerikos ekonominius ciklus**

Prieš tai kintamumas buvo vertinamas atsižvelgus į pačių akcijų bei kasmetinius kintamumo svyravimus. Tačiau įdomu ištirti koks kintamumas buvo tam tikrais ekonominiais periodais. Taigi šiuo atveju, kintamumo charakteristika bus tiriama atsižvelgiant į Jungtinių Amerikos Valstijų ekonominius ciklus. Žemiau pateiktas akcijų kainų grafikas, kuriame pažymėti laikotarpiai pagal ekonominius ciklus.



11 pav. Dienos uždarymo akcijų kainų ir ekonominių ciklų grafikas

Ekonominiai ciklai JAV:

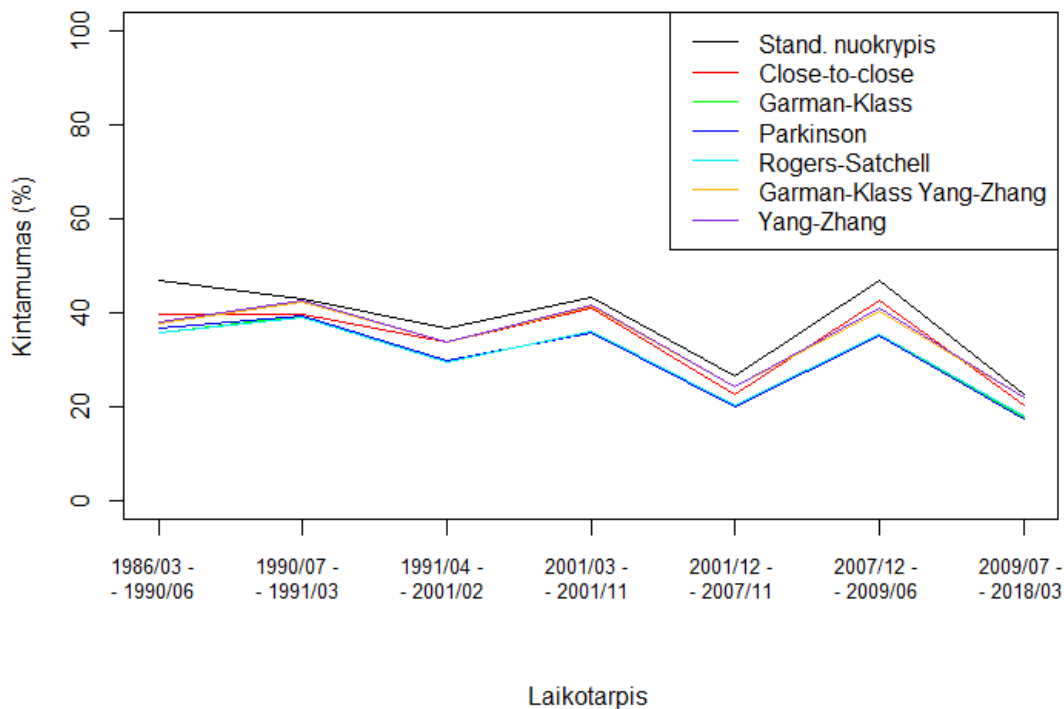
- 1982 m. gruodis – 1990 m. birželis – kai R. Reagan sumažino mokesčius ir padidino karines išlaidas, ekonomika ėmė augti.
- 1990 m. liepa – 1991 m. kovas – kai 1989 m. JAV ištiko taupymo ir paskolų krizė, ekonomika ėmė smukti.
- 1991 m. balandis – 2001 m. vasaris – interneto bendrovių (aukštųjų technologijų bendrovių, „Dot-com“) burbulas, apimantis 1995–2000 metus, turėjo didelę įtaką ekonomikos augimui.
- 2001 m. kovas – 2001 m. lapkritis – recesija<sup>2</sup>, kuri kilo dėl didelių palūkanų normų bei akcijų biržos griūties, kilusios po rugsėjo 11 d. išpuolių.
- 2001 m. gruodis – 2007 m. lapkritis – nekilnojamojo turto burbulo sprogdimas JAV, kuris turėjo lemiamą įtaką ekonomikos augimui.
- 2007 m. gruodis – 2009 m. birželis – 2008 m. finansų krizė ir Didžioji recesija.
- 2009 m. liepa – dabar – kiekybinis sušvelninimas ir JAV atsigavimas po krizės [25].

<sup>2</sup> Recesija – situacija, kai šalies bendrasis vidaus produktas mažėja du ketvirčius iš eilės.

Taigi, atsižvelgus į ekonominius ciklus JAV, buvo sudaryti septyni skirtingi laikotarpiai. Toliau šiais gautais laikotarpiais buvo įvertintas akcijų kainų kintamumas. Žemiau pateikiama lentelė su anksčiausiu laikotarpiu gautais rezultatais. Likusių laikotarpių rezultatai pateikiami 7 priede.

6 lentelė. Kintamumo įvertinimas laikotarpiu 1986 m. kovas – 1990 birželis

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986 m. kovas – 1990 m. birželis	46,90 %	39,74 %	35,80 %	36,68 %	35,65 %	37,67 %	38,06 %	RS, GK, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986 m. kovas – 1990 m. birželis	0,2083	0,1516	0,0947	0,1039	0,0928	0,0992	0,0989	RS, GK, GKYZ, YZ, PARK, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1986 m. kovas – 1990 m. birželis		1,57	3,79	3,33	4,17	3,41	3,49	RS, GK, YZ, GKYZ, PARK, CC



12 pav. Kintamumas skirtingų JAV ekonominių ciklų metu

Gauti rezultatai rodo, jog didžiausios kintamumo reikšmės gautos šiais laikotarpiais: 1986 06 – 1990 06, 2001 03 – 2001 11 ir 2007 12 – 2009 06. Pirmuoju atveju R. Reagan sumažino mokesčius, ko pasekoje kilo ekonomika, o tuo pačiu metu „Microsoft“ korporacija buvo jauna kompanija, kuri dideliais tempais augo. Svarbu tai, kad šios bendrovės akcijomis viešai pradėta prekiauti tik 1986 m. Tai buvo nauja ir daug žadanti įmonė vertybinių popierių biržoje, dėl to vyko aktyvi prekyba jos akcijomis, o tai turėjo įtaką akcijų kainų svyravimui. 2001 m. kovą – 2001 m. lapkritį JAV patyrė recesiją, kuri kilo dėl akcijų biržos griūties bei didelių palūkanų normų. Nors šis periodas buvo trumpas, tačiau jo metu akcijų kainos buvo gana nepastovios. 2007 m. gruodį – 2009 m. birželį akcijų kainų svyravimams įtakos turėjo 2008 m. finansų krizė ir Didžioji recesija, kurių metu akcijų kainos ėmė sparčiai kristi.

11 paveiksle matome, jog taip pat kaip ir prieš tai buvusiais atvejais, visų 7 kintamumo matų svyravimai sutampa. Tai rodo, jog matai suderinti ir, kad šių duomenų kintamumui vertinti galime pasirinkti bet kurį iš nagrinėjamų matų.

Panašus kintamumo tyrimas buvo atliktas dviejų Nigerijos mokslininkų O. B. Oyediran ir E. E. Sambo. Tyrimui buvo naudojami vienuolikos Nigerijos bankų 2006–2008 m. duomenys, paimti iš Nigerijos vertybinių popierių biržos. Apie tyrimą ir gautus rezultatus galima paskaityti [26]. Kaip ir minėtame, taip ir mūsų atliktame tyrime, geriausi rezultatai buvo gauti naudojant Garman-Klass, Parkinson bei Rogers-Satchell, t. y. beveik visais atvejais būtent šių matų vidutiniai absoliutiniai nuokrypiai buvo mažiausi, o efektyvumas didžiausias. Taip pat abiejuose tyrimuose gauta, jog standartinis nuokrypis pasižymėjo didžiausiomis pačio kintamumo bei MAD reikšmėmis. Šis matas gana dažnai naudojamas vertinant akcijų kainų nepastovumą, tačiau abu tyrimai paneigia jo tikslumą lyginant su kitais matais.

### **Kintamumo taikymas**

Kintamumas labai svarbus, vertinant investavimo riziką. Jei kaina išlieka santykinai stabili, tuomet kintamumas yra mažas, todėl maža ir rizika. Labai nepastovios akcijų kainos yra tos, kurios staigiai pasiekia aukštumas ir taip pat staigiai nukrenta. Tokiu atveju labai išauga rizika, t. y. tikimybė, kad tikėtina investicinių priemonių portfelio grąža nebus gauta. Kadangi žmonės labiau išgyvena nuostolių skausmą nei naudą džiaugsmą, nepastovios akcijos gali atrodyti kaip bereikalingas ir rizikingas pasiūlymas. Tačiau patyrę investuotojai žino, kad nepastovios akcijų kainos gali atnešti nemažus pinigus. Investavimas iš esmės susijęs su rizika, tačiau rizika veikia dviem būdais, t. y. kiekviena prekyba kelia riziką ir sėkmei, ir nesėkmei.

Kintamumas gali būti naudingas investuotojams. Daug konservatyvių prekybininkų pritaria ilgalaikiai strategijai „pirkti-laikyti“. Tokiu būdu akcijos įsigyjamos ir vėliau laikomos ilgą laiką,

dažnai daugelį metų, siekiant gauti naudos iš įmonės augimo. Ši strategija remiasi prielaida, kad, nors rinkoje gali būti svyravimų, ilginiui vis tiek gaunamas pelnas. Labai nepastovios akcijos gali būti nerimą keliantis pasirinkimas tokiai strategijai, o nedaug kintančios gali iš tikrųjų reikšti didesnę pelną. Kadangi kainos svyruoja, tai suteikia investuotojams galimybę įsigyti akcijų, kai kaina yra labai maža, o tada palaukti augimo. Trumpojo laikotarpio prekybininkams kintamumas yra dar svarbesnis. Tokie prekybininkai dirba su pasikeitimais, kurie vyksta kas sekundę. Jei nėra kainų pokyčių, nėra ir pelno. Vidutinės trukmės, arba kitaip dar vadinamiems „Swing“, prekybininkams aktualus šiek tiek ilgesnis laikas, dažniausiai dienos ar savaitės, tačiau rinkos kintamumas vis tiek yra jų strategijos kertinis akmuo.

Kintamumas yra rinkos rizikos matas, kuris tinkamai naudojamas, gali padidinti prekybos pelną ir sumažinti investavimo riziką. Kintamumas tradiciškai yra susijęs su chaosu ir nestabilumu. Žinoma, yra labai mažai reiškinių, kurie yra pakankamai nuoseklūs ir beveik nekinta. Istorinį kintamumą galima laikyti pagrindine akcijų kainos pokyčio norma. Kuo didesnis istorinis kintamumas, tuo labiau akcijos judėjo ir buvo nepastovesnės, todėl teoriškai jos gali labiau judėti ir ateityje. Teigiama, kad nepastovumas yra viena iš paprasčiausių statistinės rizikos vertinimo priemonių. Jis gali būti naudojamas vertinant finansinio instrumento arba jų portfelio riziką. Brandt ir Kingsley teigė, kad kintamumo įvertinimas yra labai svarbus rizikos valdymui, kainodarai ir portfelio sudarymui. Kintamumas dažniausiai kyla dėl naujos informacijos atsiradimo. Pavyzdžiui, kai investuotojai gauna naujienų apie įmonių pelną, palūkanų normas, dividendus ar ekonomiką, jie naudoja tą informaciją pirkimo ir pardavimo sprendimams priimti. Istorinis kintamumas apskaičiuojamas naudojant ankstesnes akcijų kainas. Kiekvienas finansų ekonomistas ar ekspertas turi skirti didžiulį dėmesį kintamumo vertinimo ir prognozavimo metodams, nes tai nusako investicijų į akcijas riziką. Tai yra svarbu optimalaus portfelio sudarymui. Neabejotina, kad nepastovumas yra svarbi finansų teorijos ir taikymo koncepcija. Todėl tinkamas kintamumo modeliavimas visada naudingas tiek praktikams, tiek mokslininkams [26].



## Išvados

Išanalizavus „Microsoft“ korporacijos dienos atidarymo, didžiausią, mažiausią ir uždarymo akcijų kainas, duomenys buvo suskirstyti į laikotarpius, atsižvelgiant į kainų augimo ir kritimo tendencijas. Tuomet kai kurios charakteristikos, tokios kaip sparta, apimtis ir kintamumas buvo vertinamos kiekvienu iš gautų laikotarpių, o teisingumas ir įvairovė įvertinti bendrai visam duomenų rinkiniui. Kintamumo charakteristika buvo vertinama ne tik anksčiau įvardintais laikotarpiais, tačiau ir laikotarpiais, gautais atsižvelgus į pačio kintamumo kasmetinius svyravimus, bei laikotarpiais, sudarytais atsižvelgiant į JAV ekonominius ciklus. Atlikus tyrimą, buvo suformuotos pagrindinės išvados:

1. Nubraižius „Microsoft“ akcijų kainų grafiką, aiškiai išsiskyrė trys laikotarpiai: 1986–1999 m. (augimas), 2000–2011 m. (kritimas) ir 2012–2018 m. (augimas). Kadangi 2008 m. pasaulį užklupo ekonominė krizė, atsirado poreikis atskirai išskirti šį laikotarpį, todėl akcijų kainos kritimo laikotarpis dar buvo suskaidytas tokiu būdu: 2000–2006 m. ir 2007–2011 m. Taip buvo gauti tyrimo laikotarpiai.
2. Įvertinus duomenų spartos charakteristiką, buvo nustatyta, jog tyrimo duomenys „Bloomberg“ platformoje (iš kur ir buvo paimti) yra atnaujinami realiu laiku, tačiau jų naudingumo trukmė yra labai trumpa, nes kainai atsinaujinus, prieš tai buvusi kaina jau yra nebeaktuali. Todėl tokie duomenys yra greitai „senstantys“. Taip pat nustatyta, jog vidutinis akcijų kainų įrašų skaičius per metus yra 252.
3. Atlikus duomenų kiekio, kuris šiame darbe susietas su informacijos kiekiu ir Shannon entropija, vertinimą, buvo nustatyta, jog didžiausios Shannon entropijos reikšmės gautos laikotarpiu 2007-01-01–2011-12-31 ir lygios maždaug 0,0082 bito. Mažiausios reikšmės lygios ~0,0034, gautos laikotarpiu 1986-03-13–1999-12-31. Taip pat nustatyta, jog didžiausios informacijos kiekio reikšmės stebimos ilgiausiu laikotarpiu (1986–1999 m.), o mažiausios atitinkamai trumpiausiu (2007–2011 m.).
4. Septynių skirtingų matų pagalba įvertinus kintamumą laikotarpiais, sudarytais atsižvelgiant į akcijų kainų svyravimus, nustatyta, kad didžiausios kintamumo reikšmės stebimos laikotarpiu 1986-03-13–1999-12-31. Didžiausia kintamumo reikšmė buvo gauta standartinio nuokrypio pagalba ir ji siekė 38.66 %. Mažiausios kintamumas stebimas laikotarpiu 2012-01-01–2018-03-13. Čia mažiausia reikšmė gauta Parkinson matu ir ji yra 11,92 %.
5. Įvertinus kintamumo charakteristiką laikotarpiais, gautais atsižvelgus į kasmetinį kintamumo svyravimą, matyti, jog didžiausios kintamumo reikšmės stebimos laikotarpiu 1986-03-13–

–1987-12-31. Didžiausias kintamumas, gautas pasitelkiant standartinį nuokrypį, siekia 58,94 %. Mažiausios kintamumo reikšmės stebimos 2009-01-01–2017-12-31. Šiuo laikotarpiu mažiausia kintamumo reikšmė gauta Parkinson matu ir ji yra 13,20 %.

6. Taip pat atsižvelgiant į JAV ekonominius ciklus, buvo sudaryti nauji laikotarpiai, kuriais buvo vertinamas kintamumas. Gauta, jog didžiausias kintamumas stebimas laikotarpiu, kai buvo sumažinti mokesčiai ir ekonomika augo, recesijos, kuri kilo dėl akcijų biržos griūties, metu bei finansinės krizės laikotarpiu.
7. Atskirais laikotarpiais paskaičiavus kiekvieno kintamumo mato efektyvumą bei vidutinį absoliutinį nuokrypį (MAD), nustatyta, jog dažniausiai didžiausiu efektyvumu ir mažiausiu MAD pasižymėjo Parkinson, Garman-Klass ir Rogers-Satchell matai. Na, o priešingomis savybėmis visais atvejais pasižymėjo standartinis nuokrypis. Nors matų efektyvumas bei tikslumas skiriasi, tačiau visi matai vienodai atspindi kintamumo tendencijas, todėl vertinant „Microsoft“ akcijų kainų kintamumą visi matai yra tinkami.

## Literatūros sąrašas

- [1] Kalicky, A. Master Thesis – High Performance Analytics. Czech Republic: Charles University, 2013.
- [2] Sippe, R. Master Thesis - The Relation between Big Aata and Informational Privacy in The Context of The Healthcare. Netherlands: Delft University of Technology, 2015.
- [3] Abdelhameed, M. Master Thesis – The Effect of In-Memory Computing and Big Data on Enterprise Software. Czech Republic: Masaryk University, 2016.
- [4] Shafer, T. The 42 V's of Big Data and Data Science. April 1, 2017 [žiūrėta 2018-02-19]. Prieiga per internetą:  
<https://www.elderresearch.com/company/blog/42-v-of-big-data>
- [5] Kitchin, R., and. G. McArdle. Research Article – What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. February 17, 2016 [žiūrėta 2018-05-27]. Prieiga per internetą:  
<http://journals.sagepub.com/doi/10.1177/2053951716631130>
- [6] Vorhies, W. How Many „V's“ in Big Data? The Characteristics that Define Big Data. October 31, 2014 [žiūrėta 2018-02-21]. Prieiga per internetą:  
<https://www.datasciencecentral.com/profiles/blogs/how-many-v-s-in-big-data-the-characteristics-that-define-big-data>
- [7] Jewell, D., Diederichs, S. and others. Performance and Capacity Implications for Big Data. IBM Redbooks, 2014. ISBN 9780738453583.
- [8] Big data forms. 2014 [žiūrėta 2018-04-08]. Prieiga per internetą:  
<https://e-skillsbusinessstoolbox.webnode.nl/big-data/voorbij-de-hype/>
- [9] Data and Data Warehouse: Scope, Challenges, Future. 2015 [žiūrėta 2018-04-08]. Prieiga per internetą:  
<http://kaurchanpreet.blogspot.lt/2015/02/data-and-data-warehouse-scope.html>
- [10] A Beginner's Guide to Big Data! 2016 [žiūrėta 2018-04-08]. Prieiga per internetą:  
<http://upxacademy.com/beginners-guide-to-big-data/>
- [11] Sadahoru, K. Master Thesis – Major Challenges and Solutions for Utilizing Big Data in The Maritime Industry. Sweden: World Maritime University, 2015.
- [12] Lukoianova, T. and V. Rubin. Veracity Roadmap: Is Big Data Objective, Truthful and Credible? November 2013.
- [13] International Controller Association: Having projection on boom and crisis | To design controlling to withstand volatility! 2013 [žiūrėta 2018-03-25]. Prieiga per internetą:  
[https://www.icvcontrolling.com/fileadmin/Assets/Content/AK/Ideenwerkstatt/Files/ICV\\_IW\\_DreamCar\\_Volatilit%C3%A4t\\_engl.pdf](https://www.icvcontrolling.com/fileadmin/Assets/Content/AK/Ideenwerkstatt/Files/ICV_IW_DreamCar_Volatilit%C3%A4t_engl.pdf)

- [14] NIST Big Data Public Working Group: DRAFT NIST Big Data Interoperability Framework: Volume 1, Definitions. National Institute of Standards and Technology, 2017.
- [15] Normandeau, K. Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity. September 12, 2013 [žiūrėta 2018-03-25]. Prieiga per internetą:  
<https://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
- [16] Kajackas, A. Ryšių teorijos pagrindai: mokomoji knyga. Vilnius: Technika, 2001.
- [17] Paryani, J. Bachelor Thesis – A Case Study on Determining The Big Data Veracity: A Method To Compute The Relevance of Twitter Data. India: Pune University, 2012.
- [18] Bennett, C. and Miguel A. Gil. Measuring Historical Volatility: Close-to-Close, Exponentially Weighted, Parkinson, Garman-Klass, Rogers-Satchell and Yang-Zhang Volatility. Madrid, February 3, 2012.
- [19] Brandt, M. W. and J. Kinlay. Estimating Historical Volatility. The Fuqua School of Business Duke University, 2005.
- [20] Gupta, K. Best Programming Languages for Big Data. July 20, 2017 [žiūrėta 2018-05-09]. Prieiga per internetą:  
<https://www.freelancinggig.com/blog/2017/07/20/best-programming-languages-big-data/>
- [21] What is R? [žiūrėta 2018-04-19]. Prieiga per internetą:  
<https://www.r-project.org/about.html>
- [22] „Microsoft Corporation“ internetinis puslapis [žiūrėta 2018-04-20]. Prieiga per internetą:  
<https://news.microsoft.com/facts-about-microsoft/#About>
- [23] Caplinger, D. Microsoft Stock History: How the Tech Giant Made Shareholders Rich. 2016 [žiūrėta 2018-04-24]. Prieiga per internetą:  
<https://www.fool.com/investing/2016/09/06/microsoft-stock-history-how-the-tech-giant-made-sh.aspx>
- [24] Internetinė svetainė [žiūrėta 2018-05-04]:  
<http://business.nasdaq.com/discover/events/trading-hours/index.html>
- [25] Amadeo, K. Boom and Bust Cycle: Causes & History. October 18, 2017 [žiūrėta 2018-05-26]. Prieiga per internetą:  
<https://www.thebalance.com/boom-and-bust-cycle-causes-and-history-3305803>
- [26] Oyediran, O. B. And E. E. Sambo. Comparative Analysis of Some Volatility Estimators: An Application to Historical Data from the Nigerian Stock Exchange Market. National Mathematical Centre, 2017.

## Priedai

1 priedas

### Duomenų klasifikavimas

7 lentelė. Duomenų tipų klasifikavimas [3]

Duomenys	Tipas	Apimtis	Aprašymas	Pavyzdžiai
Pagrindiniai	Struktūrizuoti	Maža	Įmonės lygmens duomenų subjektai, kurie yra organizacijos strateginės vertės. Paprastai nepastovūs ir nesusiję su sandoriais.	Klientai, produktai, tiekėjai ir vietos.
Sandoriai	Struktūrizuoti arba pusiau struktūrizuoti	Vidutinė – didelė	Verslo sandoriai, kurie yra užfiksuoti verslo operacijų ir procesų metu.	Pirkimo įrašai, užsakymai ir apmokėjimai.
Nuorodos	Struktūrizuoti arba pusiau struktūrizuoti duomenys	Maža – vidutinė	Iš vidaus valdomi išorės faktai, lemiantys organizacijos gebėjimą efektyviai apdoroti sandorius, valdyti pagrindinius duomenis ir teikti sprendimų palaikymo funkcijas.	Geografiniai ir pardavimų duomenys.
Metaduomenys	Struktūrizuoti	Maža	Apibrėžiami kaip duomenys apie duomenis. Tai yra informacija, kuri padeda suprasti išdėstymo struktūrą, tarpusavio ryšius ir teisingai interpretuoti pateikiamus duomenis.	Duomenų pavadinimai, matmenys arba vienetai, duomenų subjektų apibrėžimai arba skaičiavimo formulės.
Analitiniai duomenys	Struktūrizuoti	Vidutinė – didelė	Išvestiniai verslo operacijų ir sandorių duomenys, naudojami ataskaitų teikimo ir analizės poreikiams tenkinti.	Duomenys, esantys duomenų saugyklose ar duomenų bazėse.
<b>Didieji duomenys</b>	<b>Stuktūrizuoti, pusiau struktūrizuoti ir nestructūrizuoti</b>	<b>Didelė</b>	<b>Dideli duomenų rinkiniai, kuriuos sudėtinga saugoti, bendrinti, vizualizuoti, analizuoti ir jų ieškoti.</b>	<b>Vartotojo ir kompiuterio sukurtas turinys socialinės žiniasklaidos priemonių, žiniatinklio, programinės įrangos, informacinių jutiklių ar mobiliųjų įrenginių pagalba.</b>

**Išorinės nuorodos**

8 lentelė. Nuorodos [17]

<b>Nr.</b>	<b>Nuorodos</b>
1.	Blob classes, <a href="http://textblob.readthedocs.io/en/dev/api_reference.html">http://textblob.readthedocs.io/en/dev/api_reference.html</a>
2.	Pattern Module, <a href="https://www.clips.uantwerpen.be/pattern">https://www.clips.uantwerpen.be/pattern</a>
3.	TextBlob Documentation, <a href="https://media.readthedocs.org/pdf/textblob/dev/textblob.pdf">https://media.readthedocs.org/pdf/textblob/dev/textblob.pdf</a>
4.	TextBlob Sentiment Analyzers, <a href="http://textblob.readthedocs.io/en/dev/api_reference.html#textblob.en.sentiments.PatternAnalyzer">http://textblob.readthedocs.io/en/dev/api_reference.html#textblob.en.sentiments.PatternAnalyzer</a>

### Švenčių ir sutrumpintos prekybos dienos JAV

9 lentelė. Šventinės ir sutrumpintos dienos JAV

<b>Diena</b>	<b>Šventės/įvykiai</b>	<b>Rinkos statusas</b>
Sausio 1 d.	Pirma Naujųjų metų diena	Uždaryta
Trečias sausio pirmadienis	Martin Luther King Jr. diena	Uždaryta
Trečias vasario pirmadienis	Prezidento diena	Uždaryta
Paskutinis penktadienis prieš Velykas	Didysis penktadienis	Uždaryta
Paskutinis gegužės pirmadienis	Atminimo diena	Uždaryta
Diena prieš Nepriklausomybės dieną	Sutrumpintas darbo laikas	Darbo laikas 09:30-13:00
Liepos 4 d.	Nepriklausomybės diena	Uždaryta
Pirmas rugsėjo pirmadienis	Darbo diena	Uždaryta
Ketvirtas lapkričio ketvirtadienis	Padėkos diena	Uždaryta
Diena prieš Kūčias	Sutrumpintas darbo laikas	Darbo laikas 09:30-13:00
Gruodžio 24 d.	Kūčios (sutrumpintas darbo laikas)	Darbo laikas 09:30-13:00
Gruodžio 25 d.	Kalėdos	Uždaryta

**Kintamumo įvertinimas**

10 lentelė. Kintamumo įvertinimas laikotarpiu 2000-01-01–2006-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2000-01-01–2006-12-31	35,55 %	21,34 %	18,44 %	18,32 %	18,55 %	21,96 %	22,03 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2000-01-01–2006-12-31	0,1838	0,1095	0,0736	0,0754	0,0733	0,0948	0,0957	RS, GK, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2000-01-01–2006-12-31		3,30	8,27	7,97	8,29	4,68	4,60	RS, GK, PARK, GKYZ, YZ, CC

11 lentelė. Kintamumo įvertinimas laikotarpiu 2007-01-01–2011-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2007-01-01–2011-12-31	32,44 %	20,09 %	17,79 %	17,48 %	18,08 %	20,97 %	21,12 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2007-01-01–2011-12-31	0,1613	0,0867	0,0587	0,0592	0,0603	0,0732	0,0758	GK, PARK, RS, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2007-01-01–2011-12-31		3,89	9,07	8,75	8,40	6,21	5,75	GK, PARK, RS, GKYZ, YZ, CC



12 lentelė. Kintamumo įvertinimas laikotarpiu 2012-01-01–2018-03-13

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2012-01-01–2018-03-13	22,31 %	14,38 %	12,08 %	11,92 %	12,20 %	15,13 %	15,17 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2012-01-01–2018-03-13	0,1094	0,0535	0,0255	0,0267	0,0261	0,0448	0,0458	GK, RS, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2012-01-01–2018-03-13		4,29	20,10	18,81	19,48	6,16	5,93	GK, RS, PARK, GKYZ, YZ, CC

**Kasmetinis kintamumo įvertinimas**

13 lentelė. Kasmetinis kintamumo įvertinimas laikotarpiu 1986–2018 m.

<b>Metai</b>	<b>Garman-Klass</b>	<b>Parkinson</b>	<b>Rogers-Satchell</b>
1986	23,08 %	23,69 %	23,51 %
1987	35,95 %	37,46 %	35,52 %
1988	22,66 %	22,63 %	22,71 %
1989	22,75 %	23,50 %	22,27 %
1990	25,59 %	25,41 %	25,75 %
1991	22,40 %	23,11 %	22,04 %
1992	22,66 %	22,25 %	22,95 %
1993	20,89 %	20,37 %	21,05 %
1994	18,10 %	17,88 %	18,25 %
1995	20,05 %	20,54 %	19,87 %
1996	15,78 %	16,60 %	15,35 %
1997	19,77 %	20,03 %	19,76 %
1998	20,91 %	21,98 %	20,42 %
1999	23,21 %	23,41 %	23,35 %
2000	29,00 %	28,94 %	28,86 %
2001	26,20 %	25,93 %	26,38 %
2002	24,05 %	24,58 %	24,12 %
2003	17,01 %	17,08 %	16,84 %
2004	11,32 %	11,00 %	11,51 %
2005	9,98 %	9,70 %	10,12 %
2006	11,58 %	11,12 %	12,07 %
2007	13,91 %	13,54 %	14,17 %
2008	25,74 %	25,9 %	26,00 %
2009	19,77 %	19,33 %	20,06 %
2010	15,38 %	14,78 %	15,81 %
2011	14,69 %	14,3 %	14,95 %
2012	11,91 %	11,76 %	11,94 %
2013	12,51 %	12,34 %	12,62 %
2014	11,99 %	11,97 %	12,01 %
2015	14,12 %	14,08 %	14,17 %
2016	12,18 %	11,88 %	12,47 %
2017	9,22 %	8,97 %	9,37 %
2018	16,96 %	17,15 %	16,93 %

**Kintamumo įvertinimas**

14 lentelė. Kintamumo įvertinimas laikotarpiu 1988-01-01–1996-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1988-01-01–1996-12-31	33,99 %	22,85 %	21,10 %	21,25 %	21,03 %	23,08 %	23,16 %	RS, GK, PARK, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1988-01-01–1996-12-31	0,1642	0,0693	0,0452	0,0472	0,0455	0,0525	0,0532	GK, RS, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1988-01-01–1996-12-31		5,75	14,07	13,19	13,57	10,57	10,32	GK, RS, PARK, GKYZ, YZ, CC

15 lentelė. Kintamumo įvertinimas laikotarpiu 1997-01-01–2000-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1997-01-01–2000-12-31	42,59 %	27,70 %	23,26 %	23,60 %	23,17 %	27,51 %	27,54 %	RS, GK, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1997-01-01–2000-12-31	0,2038	0,0885	0,0508	0,0520	0,0515	0,0714	0,0720	GK, RS, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1997-01-01–2000-12-31		4,76	17,03	16,68	16,91	7,34	7,18	GK, RS, PARK, GKYZ, YZ, CC

16 lentelė. Kintamumo įvertinimas laikotarpiu 2001-01-01–2005-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001-01-01–2005-12-31	32,11 %	19,92 %	17,55 %	17,49 %	17,64 %	20,60 %	20,67 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001-01-01–2005-12-31	0,1690	0,0941	0,0641	0,0659	0,0643	0,0796	0,0810	GK, RS, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001-01-01–2005-12-31		3,96	8,90	8,37	8,69	5,95	5,77	GK, RS, PARK, GKYZ, YZ, CC

17 lentelė. Kintamumo įvertinimas laikotarpiu 2006-01-01–2008-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2006-01-01–2008-12-31	33,03 %	19,66 %	17,01 %	16,79 %	17,34 %	20,24 %	20,47 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2006-01-01–2008-12-31	0,1657	0,0977	0,0659	0,0669	0,0869	0,0675	0,0896	GK, PARK, GKYZ, RS, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2006-01-01–2008-12-31		3,44	7,52	7,22	7,00	4,88	4,53	GK, PARK, RS, GKYZ, YZ, CC

18 lentelė. Kintamumo įvertinimas laikotarpiu 2009-01-01–2017-12-31

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2009-01-01–2017-12-31	24,39 %	15,50 %	13,46 %	13,20 %	13,64 %	16,50 %	16,55 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2009-01-01–2017-12-31	0,1201	0,0595	0,0348	0,0350	0,0359	0,0514	0,0525	GK, PARK, RS, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2009-01-01–2017-12-31		4,32	13,39	13,50	12,49	6,42	6,19	PARK, GK, RS, GKYZ, YZ, CC

19 lentelė. Kintamumo įvertinimas laikotarpiu 2018-01-01–2018-03-13

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2018-01-01–2018-03-13	27,58 %	19,15 %	16,96 %	17,16 %	16,93 %	19,39 %	19,45 %	RS, GK, PARK, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2018-01-01–2018-03-13	0,1367	0,0805	0,0581	0,0608	0,0586	0,0574	0,0609	GK, RS, GKYZ, PARK, YZ, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2018-01-01–2018-03-13		4,60	9,12	8,19	9,04	9,15	8,08	GKYZ, GK, RS, PARK, YZ, CC

**Kintamumo įvertinimas**

20 lentelė. Kintamumo įvertinimas laikotarpiu 1990 m. liepa – 1991 m. kovas

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1990 m. liepa – 1991 m. kovas	43,12 %	39,75 %	39,24 %	39,32 %	39,14 %	42,33 %	42,55 %	RS, GK, PARK, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1990 m. liepa – 1991 m. kovas	0,1847	0,1152	0,0825	0,0854	0,0847	0,0877	0,0902	GK, RS, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1990 m. liepa – 1991 m. kovas		2,36	5,33	5,28	4,92	4,88	4,69	GK, PARK, RS, GKYZ, YZ, CC

21 lentelė. Kintamumo įvertinimas laikotarpiu 1991 m. balandis – 2001 m. vasaris

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1991 m. balandis – 2001 m. vasaris	36,91 %	33,90 %	29,72 %	29,97 %	29,63 %	33,93 %	33,97 %	RS, GK, PARK, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1991 m. balandis – 2001 m. vasaris	0,1799	0,1086	0,0664	0,0687	0,0675	0,0884	0,0894	GK, RS, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
1991 m. balandis – 2001 m. vasaris		2,62	7,58	7,38	7,40	3,86	3,80	GK, PARK, RS, GKYZ, YZ, CC

22 lentelė. Kintamumo įvertinimas laikotarpiu 2001 m. kovas – 2001 m. lapkritis

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001 m. kovas – 2001 m. lapkritis	43,41 %	40,90 %	35,88 %	35,70 %	35,99 %	41,49 %	41,58 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001 m. kovas – 2001 m. lapkritis	0,2044	0,1298	0,0773	0,0792	0,0777	0,0912	0,0900	GK, RS, PARK, YZ, GKYZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001 m. kovas – 2001 m. lapkritis		3,32	9,58	9,05	9,08	5,99	6,04	GK, RS, PARK, YZ, GKYZ, CC

23 lentelė. Kintamumo įvertinimas laikotarpiu 2001 m. gruodis – 2007 m. lapkritis

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001 m. gruodis – 2007 m. lapkritis	26,46 %	22,78 %	20,25 %	20,07 %	20,48 %	24,16 %	24,28 %	PARK, GK, RS, CC, GKYZ, YZ, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001 m. gruodis – 2007 m. lapkritis	0,1366	0,1059	0,0643	0,0678	0,0640	0,0890	0,0907	RS, GK, PARK, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2001 m. gruodis – 2007 m. lapkritis		2,02	5,81	5,15	5,69	3,01	2,87	GK, RS, PARK, GKYZ, YZ, CC

24 lentelė. Kintamumo įvertinimas laikotarpiu 2007 m. gruodis – 2009 m. birželis

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2007 m. gruodis – 2009 m. birželis	46,89 %	42,75 %	35,02 %	34,99 %	35,35 %	40,38 %	40,91 %	PARK, GK, RS, GKYZ, YZ, CC, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2007 m. gruodis – 2009 m. birželis	0,2496	0,1784	0,1104	0,1126	0,1152	0,1309	0,1372	GK, PARK, RS, GKYZ, YZ, CC, Sand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2007 m. gruodis – 2009 m. birželis		2,45	6,18	5,95	5,52	4,66	4,19	GK, PARK, RS, GKYZ, YZ, CC

25 lentelė. Kintamumo įvertinimas laikotarpiu 2009 m. liepa – 2018 m. kovas

Kintamumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2009 m. liepa – 2018 m. kovas	22,58 %	20,42 %	17,86 %	17,52 %	18,10 %	22,03 %	22,08 %	PARK, GK, RS, GKYZ, YZ, CC, Stand. nuokrypis
Vidutinis standartinis nuokrypis (MAD)								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2009 m. liepa – 2018 m. kovas	0,1100	0,0725	0,0402	0,0408	0,0416	0,0635	0,0648	GK, PARK, RS, GKYZ, YZ, CC, Stand. nuokrypis
Efektyvumas								
Laikotarpis	Stand. nuokrypis	CC	GK	PARK	RS	GKYZ	YZ	Reitingavimas
2009 m. liepa – 2018 m. kovas		2,38	8,02	8,08	7,28	3,13	2,99	PARK, GK, RS, GKYZ, YZ, CC