



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Kriptokatinėlių rinkos vertės nustatymas ir pritaikymas prekybos strategijai kurti

Baigiamasis magistro projektas

Laura Sirvydytė
Projekto autorė

Prof. dr. Evaldas Vaičiukynas

Vadovas

Doc. dr. Alfreda Šapkauskienė

Vadovė

Vilnius, 2020



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Kriptokatinėlių rinkos vertės nustatymas ir pritaikymas prekybos strategijai kurti

Baigiamasis magistro projektas
Didžiųjų duomenų verslo analitika 6213AX001

Laura Sirvydytė

Projekto autorė

Prof. dr. Evaldas Vaičiukynas

Vadovas

Doc. dr. Alfreda Šapkauskienė

Vadovė

Prof. dr. Audrius Lopata

Recenzentas

Prof. dr. Lina Dagilienė

Recenzentė

Vilnius, 2020



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Laura Sirvydytė

Kriptokatinėlių rinkos vertės nustatymas ir pritaikymas prekybos strategijai kurti

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Lauros Sirvydytės, baigiamasis projektas tema „Kriptokatinėlių rinkos vertės nustatymas ir pritaikymas prekybos strategijai kurti“ 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 mokėjęs.

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

(vardą ir pavardę įrašyti ranka)

(parašas)

Sirvydytė, Laura. Kriptokatinėlių rinkos vertės nustatymas ir pritaikymas prekybos strategijai kurti. Magistro baigiamasis projektas / vadovas prof. dr. Evaldas Vaičiukynas / vadovė doc. dr. Alfreda Šapkauskienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Taikomoji matematika (Matematikos mokslai)

Reikšminiai žodžiai: Kriptokatinėliai, blokų grandinė, mašininis mokymasis, atsitiktinis miškas, reguliarizuota regresija, virtualios kolekcinės vertybės.

Vilnius, 2020. 52 p.

Santrauka

Šio darbo tyrimų objektas yra blokų grandinės technologija paremtas žaidimas *Cryptokitties*. Šiame žaidime galima kolekcionuoti, pirkti, parduoti bei veisti unikalias savybes turinčius objektus – kriptokatinėlius. Šis žaidimas tapo gana populiarus tiek tarp žaidėjų, norinčių smagiai praleisti laiką, tiek tarp investuojančių ir bandančių iš to uždirbti. Buvo bandomas sukonstruoti modelis, kuris įvertintų katinėlio rinkos kainą. Kadangi katinėliai turi labai daug savybių, katinėlio kainos įvertinimas žaidėjui tapo gana sudėtingu uždaviniu. Šiame darbe buvo bandoma prognozuoti katinėlio kainą dviem mašininio mokymo metodais – atsitiktinio miško bei reguliarizuotos regresijos algoritmais. Duomenys imti iš trijų skirtingų šaltinių. Tam, kad gautume geresnius rezultatus, buvo sukurti nauji kintamieji, pasitelktas ir tekstinio lauko vektorizavimo metodas *Sent2Vec*. Geriausias modelis buvo sukurtas su atsitiktinio miško algoritmu. Svarbiausi kintamieji modelyje buvo susiję su eterio ir *JAV* dolerio valiutos kursais, katinėlio atributų bei jų porų dažniais, prabangumo tipais ir kitomis savybėmis. Tačiau prognozavimo rezultatai nėra labai tikslūs, todėl nevertėtų modelio naudoti investavimo strategijai kurti. Šio žaidimo populiarumas pastaruoju metu yra sumažėjęs, tačiau vis daugiau dėmesio yra skiriama virtualioms kolekcinėms vertybėms, kurios paremtos blokų grandinės technologija. Todėl vertėtų toliau tęsti šios srities tyrimus, nes tikėtina, kad sukūrus pakankamai tikslų modelį ir pelningą investavimo strategiją, juos bus galima pritaikyti ir ateityje.

Sirvydytė, Laura. Market valuation of Cryptokitties with application to trading strategy development. Master's Final Degree Project / supervisor prof. dr. Evaldas Vaičiukynas / supervisor doc. dr. Alfreda Šapkauskienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics (Mathematical Sciences).

Keywords: Cryptokitties, blockchain, machine learning, random forest, regularized regression, virtual collectibles.

Vilnius, 2020. 52 pages.

Summary

The object of this research is the game *Cryptokitties*. It is based on blockchain technology. In this game you can collect, buy, sell and breed kitties with unique properties. This game has become quite popular both among players who want to have fun and among those who invest and try to earn money. An attempt was made to construct a model that would estimate the market price of kitties. Since they have so many features, estimating the price of a kitten has become quite a challenge for the player. In this research, we used two machine learning methods - random forest and regularized regression algorithm to predict the price of a kitten. The data was taken from three different sources. In order to obtain better results, new variables were created and the text field vectorization method *Sent2Vec* was used. The random forest model has given the best results. The most important variables in the model were related to ether and US dollar exchange rates, frequencies of kitten attributes and their pairs, fancy types, and other characteristics. However, the forecasting results are not very accurate, so it would not be worth using the model to develop an investment strategy. The popularity of this game has declined recently, but there is an increasing focus on virtual collectibles based on blockchain technology. It is therefore worth pursuing further research in this area, as it is likely that a sufficiently precise model and a profitable investment strategy will allow them to be applied in the future.

Turinys

Lentelių sąrašas	7
Paveikslų sąrašas	8
Įvadas	9
1. Literatūros apžvalga	10
1.1. Kolekcionavimas.....	10
1.2. Virtualus turtas.....	11
1.3. Blokų grandinė ir kriptovaliutos	12
1.4. Virtualus turtas blokų grandinės pagrindu.....	13
1.5. Kriptokatinėlių savybės ir tipai.....	14
1.6. Kriptokatinėlių kainos įvertinimas.....	19
1.7. Kainos vertinimui naudojami algoritmai.....	21
1.8. Kriptokatinėlių kainos ir pardavimų dinamika laike.....	22
1.9. Ateities galimybės virtualioms kolekcinėms vertybėms	23
1.10. Apžvalgos apibendrinimas	24
2. Medžiagos ir tyrimų metodai	25
2.1. Modeliai.....	25
2.1.1. Atsitiktinis miškas.....	25
2.1.2. Reguliarizuota regresija.....	26
2.2. Tekstinio lauko vektorizavimas	27
2.3. Modelio patikrinimas	28
2.4. Modelio vertinimas	28
2.5. Kintamųjų svarba	30
2.6. Naudojami duomenys.....	30
2.7. Duomenų gavimas ir naudojama programinė įranga	31
2.8. Modelių kūrimas su R	31
2.8.1. Atsitiktinis miškas.....	31
2.8.2. Reguliarizuota regresija.....	32
2.9. Kintamieji	32
3. Tyrimų rezultatai ir jų aptarimas	35
3.1. Duomenų paruošimas.....	35
3.2. Išvestiniai kintamieji	37
3.3. Priklausomo kintamojo transformacija	38
3.4. Modeliai.....	39
3.5. Tyrimų rezultatai.....	39
3.5.1. Paklaidų rezultatai	39
3.5.2. Prognozuotos ir tikros reikšmės grafikai	40
3.6. <i>Bland Altman</i> grafikas su paraštinėmis histogramomis	43
3.7. Vektorizuoto tekstinio lauko įtaka paklaidoms	45
3.8. Kintamųjų svarba modelyje.....	46
3.9. Prekybos strategija	48
3.10. Tyrimų ribotumas ir tyrimai ateityje	48
Išvados	50
Literatūros sąrašas	51

Lentelių sąrašas

1 lentelė. Skaitiniai kintamieji.....	32
2 lentelė. Kategoriniai kintamieji	33
3 lentelė. Loginiai kintamieji	33
4 lentelė. Modelių rezultatai.....	40
5 lentelė. Tekstinio lauko vektorizavimo rezultatai	45

Paveikslų sąrašas

1 pav. Atvėsimo lygmenys.....	15
2 pav. Atributai.....	16
3 pav. Genomo sandara.....	17
4 pav. Paprastas tipas.....	18
5 pav. Prabangus tipas	18
6 pav. Išskirtinių tipas.....	19
7 pav. Genesis ir jaguarai.....	20
8 pav. Valiutos kursas ETH-USD (duomenų šaltinis finance.yahoo.com)	21
9 pav. Vidutinė kriptokatinėlių pardavimo kaina (pardavimo metu, USD) ir parduotų kačiukų skaičius nuo 2017 m. lapkričio iki 2020 m. sausio	22
10 pav. Atsitiktinių medžių algoritmas.....	26
11 pav. Kainos stačiakampė diagrama prieš išskirčių pašalinimą	36
12 pav. Kainos stačiakampė diagrama po išskirčių pašalinimo	36
13 pav. Pardavimo kainos histograma prieš logaritmovimą	38
14 pav. Pardavimo kainos histograma po logaritmovimo	38
15 pav. Atsitiktinio miško modelis (USD), patikrinimo imties tikros ir prognozuotos reikšmės.....	41
16 pav. Atsitiktinio miško modelis (ETH) su logaritmine transformacija, testavimo imties realios ir prognozuotos reikšmės.....	42
17 pav. Regresijos modelis (USD), testavimo imties realios ir prognozuotos reikšmės.....	42
18 pav. Regresijos modelis (ETH) su logaritmine transformacija, testavimo imties realios ir prognozuotos reikšmės.....	43
19 pav. Skirtumas tarp prognozuojamos ir realios kainos kartu su paraštinėmis histogramomis – atsitiktinis miškas (USD)	44
20 pav. Skirtumas tarp prognozuojamos ir realios kainos kartu su paraštinėmis histogramomis – reguliarizuota regresija (USD).....	45
21 pav. Kintamųjų svarba – atsitiktinis miškas (USD)	46
22 pav. Kintamųjų svarba – reguliarizuota regresija (USD).....	47

Įvadas

Įvairių daiktų kolekcionavimas yra plačiai paplitęs visuomenėje. Tam tikra psichologinė motyvacija skatina žmogų rinkti objektus pagal pasirinktas savybes, keistis objektais ir juos pardavinėti. Virtualaus turto kolekcionavimas nėra sena sąvoka. Panagrinėsime vieną populiariausių virtualaus turto kolekcionavimo žaidimą – *Cryptokitties*.

Kriptokatinėliai (angl. *Cryptokitties*) yra 2017 metais pasirodęs žaidimas, veikiantis *Ethereum* bloku grandinės pagrindu. Jame galima pirkti, parduoti ir veisti skaitmeninius katinėlius. Katinėliai turi savo savininkus ir yra laikomi skaitmeniniu kolekcinio turto. Kiekvienas katinėlis turi savo savybes ir yra unikalus. Atsižvelgiant į jo savybes, paklausą ir pasiūlą yra nustatoma katinėlio rinkos vertė. Susiduriame su problema, kad katinėlių vertę yra sunku nustatyti.

Šio darbo tikslas yra įvertinti kriptokatinėlių kainos modeliavimo perspektyvumą ir tinkamumą prekybos strategijai kurti. Tą bus bandoma pasiekti apžvelgiant literatūrą ir sukuriant modelį, kuris prognozuoja kriptokatinėlių kainą. Jeigu šiame uždavinyje išeitų pasiekti pakankamą tikslumą, tai toks modelis galėtų pasitarnauti automatinės prekybos strategijos sukūrimui. Be to, vis daugėja platformų, kuriose galima kriptovaliutomis įsigyti virtualaus turto ir vis daugiau yra investuojama į technologijas šioje srityje. Tikėtina, kad ateityje gautas tyrimo išvadas bus galima pritaikyti ir kitoms platformoms, panašioms į *Cryptokitties*.

Užsibrėšime šio darbo uždavinius:

1. Išsiaiškinti, kas gali skatinti žmones investuoti į kriptokatinėlius.
2. Sukurti kainos prognozavimo modelius, naudojant porą skirtingų algoritmų ir patikrinti, kurio modelio rezultatai bus geresni.
3. Patikrinti su kuria valiuta (JAV doleriais ar eteriu) sukurti modeliai rodo geresnius rezultatus.
4. Pabandyti pagerinti geriausio modelio rezultatus, įtraukus papildomą tekstinį kintamąjį.
5. Aptarti, kurie kintamieji turėjo daugiausiai svarbos sudarant modelius.
6. Pasvarstyti, ar gauti modeliai gali būti pritaikomi, kuriant prekybos strategiją.

1. Literatūros apžvalga

1.1. Kolekcionavimas

Kolekcionavimas yra gana plačiai paplitęs visuomenėje. Kolekcinis turtas dėl tam tikrų psichologinių ir finansinių priežasčių tapo gan paklausia preke. Yra žmonių, kurie renka pašto ženklus, monetas, beisbolo korteles, statulėles ir kitus objektus, o kolekcinį objektų vis daugėja. Kolekcionierių pagrindinis tikslas yra rinkti pasirinkto tipo vertybes į vieną vietą. Kolekcijai vertės prideda surinktų vertybių kiekis bei reti, kažkuo ypatingi kolekciniai objektai.

Visų pirma, panagrinėkime, patį kolekcionavimo procesą. Kolekcionavimas gali būti apibrėžiamas šiais žingsniais, tačiau žingsniai nebūtinai eina iš eilės ir visi yra vykdomi (Mcintosh & Schmeichel, 2004):

- nusprendžiama kolekcionuoti objektus pagal tam tikrą tipą ar savybę;
- renkama informacija apie objektus;
- išsirenkami objektai, iškeliamas tikslas, kuriuos objektus norima gauti, planuojama kaip juos gauti;
- objektų paieška, įsigijimas, reakcija po įsigijimo;
- objektų įtraukimas į kolekciją, ieškoma kitų objektų.

Matome, jog kolekcionavimas yra ciklinis procesas, kurio ciklą skaičius nebūtinai yra apibrėžtas. Tai priklauso nuo to, ar kolekciijoje gali būti limituotas kiekis objektų, ar ne. Kolekcija gali būti renkama žinant limituotą skaičių objektų ir tik tada įgauti didesnę vertę. Tačiau gali būti, kad siekiama surinkti kuo daugiau objektų, nes visų apibrėžti ir visus gauti būtų labai sunku ar neįmanoma. Kolekcionavimas kai kuriuos žmones įtraukia labai stipriai, jie skiria tam daug pinigų ir laiko. Kyla klausimas, kokios priežastys lemia šį visuomenės reiškinį, kokios jo psichologinės paskatos.

Straipsnyje „Exploring worldwide collecting consumption behaviors” (Lafferty, Matulich, & Liu, 2013) teigiama, kad kolekcionierius pagal kolekcionavimo priežastis galima skirstyti į kelis pagrindinius tipus: kolekcionavimas dėl aistros, investicijų, patiriamo malonumo ir saviraiškos galimybių. Taip pat buvo atliktas tyrimas (Zolfagharian & Cortes, 2011), kuris parodė, jog svarbiausios kolekcionavimo priežastys, kurios lemia kolekcionavimo mastą, yra socialinis priėmimas ir identifikacija grupėje, kolekcinį vertybių estetika. Ir iš tikrųjų, galime pastebėti, kad kolekcionieriai neretai buriasi į tam tikras visuomenės grupes forumuose, socialinių tinklų grupėse bei realybėje. Tikėtina, kad šie kontaktai ne tik skatina pasitenkinimą dėl socialinių priežasčių, bet taip pat yra naudingi tam, kad būtų plečiama kolekcija. Turint daugiau informacijos apie kitus kolekcionierių bendruomenės narius, galima žymiai lengviau surasti norimus įsigyti objektus. Autoriai priduria, kad mažiau, bet kolekcionavimą taip pat motyvuoja malonumas, įsitraukimas, kultūra ir sąžininga kaina (Zolfagharian & Cortes, 2011). Konkretaus kolekcionieriaus priežastims ir motyvacijai nuskaidyti nėra jokios konkrečios formulės, tai priklauso nuo pačio kolekcionieriaus. Tačiau yra svarbu suprasti, kodėl žmonės kaupia kolekcinį turtą tam, kad suprastume, kas skatina leisti pinigus ir laiką kolekcionavimui.

Sritis, ties kuria šią temą nagrinėsime šiek tiek plačiau, tai – kolekcinės vertybės kaip investicija. Kolekcinės vertybės yra perkamos, kai tikimasi, kad vertybės kaina po tam tikro laiko augs ir pardavus bus galima uždirbti pelną. Parduodamos, kai manoma, kad kolekcinė vertybė kaina ateityje gali kristi. Straipsnyje „Investing in Emotional Assets“ (Dimson & Spaenjers, 2014) teigiama, kad ilguoju laikotarpiu investuoti į meno kūrinius, pašto ženklus ir smuikus yra pelningiau nei į Jungtinės Karalystės išdo vekselius, vyriausybės obligacijas ir akcijas (tiriamasis laikotarpis 1900–2012). Reikia atsižvelgti į tai, kad kolekcinės vertybių įsigijimas reikalauja papildomų išlaidų, tokių kaip papildomi mokesčiai ir kitos išlaidos (laikymas saugyklose, transportavimas, draudimas ir kt.). Straipsnyje prie išlaidų taip pat yra įtrauktos išlaidos dėl nelikvidaus turto (išlaidos aukcionuose, ilgai ieškant potencialaus pirkėjo). Reikėtų nepamiršti, kad kolekcionavime yra ir tam tikrų rizikų. Pavyzdžiui, pirkėjų poreikių ar skonio pasikeitimas, klastotės, neadekvatus kainos nustatymas, kainų burbulai, analogiškų kolekcinė vertybių kiekio padidėjimas rinkoje ir kitos rizikos.

Taigi, sužinojome, kas yra kolekcionavimas, kokios gali būti jo priežastys ir rizikos. Tačiau šis skyrelis buvo daugiau apie realius objektus. Toliau apžvelgsime ne realius, o virtualaus turto objektus.

1.2. Virtualus turtas

Kolekcinės vertybės nebūtinai turi būti apčiuopiamos, jos egzistuoja ir virtualiai. Kai kurie teigia (Jung & Pawlowski, 2014), kad virtualaus turto vartojimo naudojimo priežastys yra panašios kaip ir turto realiame gyvenime. Yra pateikiami pavyzdžiai, kai vartotojai įtraukiami į virtualios realybės žaidimus perka tam tikrą turtą, pavyzdžiui, žaidimus, ginklus. Teigiama, kad viena iš svarbiausių priežasčių, kodėl tai yra perkama, yra norėjimas socializuotis, būti įtrauktiems į žaidime esančias vartotojų grupes. Pabrėžiama ir virtualaus turto funkcionalumo ir estetikos svarba (Lehdonvirta, 2009). Funkcionalumo svarba ypač atsiskleidžia kompiuteriniuose žaidimuose. Žaidėjai nori turėti kuo galingesnius veikėjus ir ginklus, kuo greitesnius automobilius. Estetiniai motyvai gali būti reprezentaciniai ar estetiškai patrauklūs pačiam žaidėjui. Žaidimuose dažnai pastebima, kad vartotojams svarbu, kad jų veikėjai atrodytų gražiai, išskirtinai, todėl virtualioms prekėms dažnai būna pridėdama bruožų, kurie nesuteikia funkcionalumo, o tik tam tikrus išvaizdos ar garsų bruožus. Pavyzdžiui, šviečiantys ar ryškūs drabužiai, garsūs ginklai. Tačiau tokie dalykai, kaip virtualūs drabužiai ir kiti objektai, žaidimuose yra lengvai kopijuojami ir pardavinėjami visiems vartotojams, kurie to nori, todėl dažnai yra prarandama virtualaus turto, kaip kolekcinės vertybės, struktūra. Jei objektas yra nesunkiai prieinamas, galima jį kopijuoti neribotą kiekį kartų, tai tampa tiesiog įprasta preke.

Prekės kompiuteriniuose žaidimuose nėra vienintelis virtualaus turto tipas. Kai pradėdame kalbėti apie virtualų turtą, taip pat galime įsivaizduoti muzikos įrašus, filmus, programas ir kitą. Tapo įprasta, kad virtualus turtas yra masinis vartojimo produktas: nors tam tikri virtualūs objektai ir yra unikalūs bei turi emocinę vertę, bet jie turi galybę kopijų ir jais gali naudoti masės žmonių. To pavyzdys yra filmai, dainos, elektroninės knygos. Nors galima teigti, kad kiekvienas filmas, daina ar knyga yra unikalūs, dažniausiai taip nėra, kad tik vienas vartotojas turi prieigą prie šių dalykų. Yra įvairių būdų, kaip gauti šias prekes: mokamos platformos (pavyzdžiui, *Netflix*, *Spotify*), atviros prieigos, piratavimas ir kita. Virtualus turtas tapo masiniu vartojimo produktu, kuris dažnai nėra pritaikytas individualiai ar unikalus kiekvienam vartotojui ir sunku būtų jį priskirti prie kolekcinė vertybių.

Didelė dalis virtualaus turto negali būti priskirti prie kolekcinų vertybių, nes tam reikėtų užtikrinti, kad turtas nebus kopijuojamas ir galės priklausyti tik vienam kolekcionieriui, kas yra gana sudėtinga užduotis virtualioje erdvėje. Toliau aptarsime technologiją, kuri suteikė galimybę funkcionuoti virtualių kolekcinų vertybių rinkai.

1.3. Blokų grandinė ir kriptovaliutos

Blokų grandinė yra viešas elektroninis žurnalas, kuris naudoja *P2P* (angl. *peer-to-peer*) sistemą. *P2P* yra tinklo modelis, kuriame resursų keitimas vyksta tiesiogiai tarp vartotojų. Šiame modelyje nėra kliento ir serverio sąvokų, nes visi susijungę vartotojai yra lygiaverčiai. Kiekvienas iš jų veikia kartu - veikia ir kaip serveris, ir kaip klientas (Mearian, 2019).

Taigi blokų grandinės esmė yra ta, kad vienas elektroninis žurnalas yra atviras visiems vartotojams ir visos operacijos, sukurtos jame, yra nekintančios, pažymėtos laiko žyma ir susietos su prieš tai buvusiu įrašu. Kiekvieną kartą, kai yra pridamas operacijų rinkinys, šie duomenys tampa dar vienu grandinės bloku. Pavyzdžiui, viena sandorio šalis inicijuoja procesą sukurdamą bloką. Šis blokas turi būti patvirtintas milijonų kitų kompiuterių, kurie yra bendrame tinkle. Tada tas patvirtintas blokas yra pridamas prie bendros grandinės.

Blokų grandinė gali būti atnaujinama tik susitarus tarpusavyje tarp dalyvių sistemoje. Ir nuo tada, kai nauji duomenys yra pridami, jie negali būti ištrinti. Tai technologija, kurios dėka įrašyti duomenis galima tik vieną kartą, todėl ją galima tik tikrinti, bet nekoreguoti. Ši technologinė savybė leidžia transakcijas verifikuoti po kiekvieno operacijos įrašo.

Kriptovaliutos sąmprata pirmą kartą buvo aprašyta 1998 metais, o pati technologija pirmą kartą buvo panaudota plačiai visuomenėje 2009 metais kartu su bitkoino kriptovaliuta (Bitcoin, be datos). Tai - pirmoji kriptovaliuta, kuri buvo decentralizuota. Ši valiuta išskirtinė tuo, kad ja galima atsiskaityti ją užšifravus ir neatskleidus savininkų tapatybės. Šiuo metu tai yra pirma valiuta pagal populiarumą (Coinmarketcap, 2020). Antra pagal populiarumą šiuo metu yra eterio (angl. ether) valiuta (Coinmarketcap, 2020). Ji buvo pristatyta 2013 metais programuotojo Vitaliko Buterino, kuris tuo metu buvo 19 metų. Eterio valiuta taip pat galima atsiskaityti anonimiškai. Tai – viena iš kriptovaliutos klestėjimo priežasčių.

Kriptovaliutos gali būti skirstomos į 7 kategorijas (Wolfgang Karl Hardle, 2020):

1. Kriptovaliuta kaip transakcijos mechanizmas. Ši kategorija pagrįsta tuo, kad pagrindinis jos tikslas – piniginių pervedimų platforma. To pavyzdys – bitkoino kriptovaliuta.
2. Paskirstytas skaičiavimų žetonas (angl. distributed computation token). Šio tipo kriptovaliutos platformoje yra galimybė paleisti nedideles programas. Eterio kriptovaliuta patenka į šią kategoriją.
3. Naudingumo žetonas (angl. utility token). Turi savybę, kad turtas blokų grandinėje gali būti programuojamas.
4. Vertybinių popierių žetonas (angl. security token) – atstovauja finansinius instrumentus, pavyzdžiui, akcijas ir obligacijas.
5. Pakeičiamas žetonas (angl. fungible token). Šio tipo žetonas gali būti pakeičiamas į kitą tokio pačio tipo žetoną. Pavyzdžiui, valiutų srityje, bet kurį JAV dolerį galima pakeisti kitu JAV doleriu ir dolerio turėtojui tai neturės reikšmės.

6. Nepakeičiamas žetonas (angl. non-fungible token). Šio tipo žetonai negali būti pakeičiami į kitą tokio pačio tipo žetoną.
7. Paskutinė klasė – pastovių kriptovaliutų (angl. stablecoins). Tai kriptovaliutos, susietos su valiutomis, kurių vertę garantuoja ją išleidusi vyriausybė.

Kriptovaliutą įsigyti galima keliais būdais. Ją galima nusipirkti per tam tikras platformas. Vienos platformos leidžia įsigyti kriptovaliutą iškeičiant į kitas kriptovaliutas, kitos platformos leidžia nusipirkti naudojant tokias valiutas, kaip dolerius ar eurus. Kitas būdas yra „kasti“ kriptovaliutas. Efektyviausias būdas tą daryti yra naudojant vaizdo plokščių procesorius. Jos galingiausios ir geba greičiausiai išspręsti matematinės lygtis už kurias gaunama tam tikra kriptovaliutos dalis.

Taigi, kriptovaliuta paprastam vartotojui, neturinčiam daug žinių apie naujausias technologijas, atsiskaityti gali būti sudėtinga. Tam reikia arba įsidiesti papildomas platformas, arba investuoti į techniką, su kuria galima užsidirbti kriptovaliutos.

1.4. Virtualus turtas blokų grandinės pagrindu

Straipsnyje „Is Blockchain Part of the Future of Art?“ (Kampakis, 2019) nagrinėjama kolekcinų vertybių sampratos pokyčius po to, kai kolekcinų vertybių platformos buvo pradėtos diegti blokų grandinės pagrindu. Tiek anksčiau, tiek ir dabar atsiranda kolekcinų vertybių klastočių. Klastotės gali eiti iš rankų į rankas, kol galutinis pirkėjas sužino, kad įsigijo neoriginalų objektą. Tokiu atveju yra sunku surasti asmenį, kuris pradėjo pardavinėti klastotę bei įrodyti, kad būtent nuo jo prasidėjo ši nelegali veikla. Todėl pirkėjai susiduria su rizika, kad įsigys neoriginalų objektą ir dėl to gali patirti didelius nuostolius. Tačiau atsiradus platformų, kurios virtualias kolekcinės vertybes pardavinėja blokų grandinės pagrindu, pirkėjai minėtos rizikos gali išvengti, o kūrėjai žino, kad jų darbas nebus nukopijuotas. Parduodami objektai blokų grandinės dėka tampa autentifikuojami, galima atsekti visą pardavimų istoriją. Autoriaus teigimu, šis vertybių prekybos būdas neabejotinai turės didelę reikšmę ateities menui.

Pažvelkime į porą tokių pavyzdžių, kur galima įsigyti kolekcinų vertybių, iš platformų, sukurtų su blokų grandinės technologija. Pirmasis ir sulaukęs didžiausios sėkmės – Kriptokatinėliai (angl. *Cryptokitties*). Tai yra žaidimas, sukurtas 2017 metais blokų grandinės pagrindu. Jo žaidėjai gali turėti unikalius virtualius objektus - katinėlius, taip pat juos pirkti, parduoti, sukurti naujus iš esamų ir kolekcionuoti. Kitas pavyzdys - puslapis *opensea.io*. Jis siūlo vartotojams pirkti ir parduoti tokias skaitmeninius objektus kaip kolekcinės vertybes, žaidimų objektus, skaitmeninį meną, renginių bilietus, domenų pavadinimus ir net nuosavybės dokumentus fiziniams daiktams (OpenSea, 2020). Platforma įkurta nuo 2017 metų ir 2018 metais sulaukė 2 mln. dolerių investicijų, skelbia puslapis *Coindesk* (Coindesk, 2018). Tiek šiame puslapyje, tiek ir kituose, *Opensea* dažnai yra vadinama *Ebay* kriptokatinėliams.

Šiame darbe nagrinėsime būtent Kriptokatinėlių platformą. Minėtame žaidime vartotojas gali už eterio kriptovaliutą įsigyti virtualius katinėlius, kurie pasižymi unikalumu. Taip pat vartotojas gali katinėlius veisti ir parduoti. Kriptokatinėliai sutraukė daug investuotojų bei kolekcionierių susidomėjimo. 2017 metų gruodį pats pirmasis kriptokatinėlis buvo parduotas už daugiau nei 114 tūkst. dolerių (Cheng, 2017). Čia kyla klausimas, kodėl yra mokami tokie pinigai už, atrodytu, paprastą kačiuko paveikslėlį.

Priežastys iš dalies paaiškėja straipsnyje, kur nagrinėjama, ar kriptokatinėlių žaidimas yra žaidžiamas dėl smagumo, ar spekuliaciniais tikslais (Jaehwan Lee, 2018). Tirti 60 000 vartotojų laikotarpiu nuo 2017 metų gruodžio iki 2018 metų liepos. Buvo tiriamas priklausomas kintamasis katinėlių pardavimai ir nepriklausomi vidiniai kintamieji (vartotojų pirkimai, veisimai ir dovanojimai) ir išoriniai kintamieji (kaina eteriu ir kiekis). Buvo padarytos prielaidos, kad spekuliaciniais motyvais pagrįstas žaidimas bus susijęs su išoriniais kintamaisiais, o smagumu pagrįstas žaidimas – su vidiniais. Gauti rezultatai, kad išoriniai kintamieji yra neigiamai susiję su pardavimų elgsena, o vidiniai kintamieji – teigiamai. Autoriai pateikė išvadą, kad vartotojų elgsena yra susijusi tiek su malonumu žaidžiant, tiek dėl spekuliacinių aspektų.

Ir iš tiesų, kriptokatinėlių vaizdinė sąsaja (jų atvaizdai) yra smagiai pateikta, kad pritrauktų žaidėją. Pačio žaidimo funkcijos yra įtraukiančios: vartotojas gali kurti naujus katinėlius, pirkti ir parduoti. Yra sukurtos papildomos platformos, kur galima dar išplėsti veiklas su jau turimais katinėliais. Taip pat bendruomenės turi savo forumus, organizuojami tam tikri renginiai ir net įvairūs konkursai. Tačiau tai, kad pirkimai ir atsiskaitymai vyksta eterio kriptovaliuta, gali pritraukti vartotojus, kurie nori užsidirbti investuodami. Toliau nagrinėsime, kokias savybes turi katinėliai, kas jiems suteikia vertę ir kodėl už juos mokamos didelės sumos.

1.5. Kriptokatinėlių savybės ir tipai

Pradėkime nuo to, kad kiekvienas katinėlis yra unikalus. Kiekvienas katinėlis turi 256 bitų genomą, todėl yra 4 milijardai įmanomų katinėlių variacijų. Išgauti unikalų katinėlį galima veisiant du kitus katinėlius. Bet kuris katinėlis gali būti ir motina (angl. *Matron*), ir tėvas (angl. *Sire*), katinėliai šiame žaidime neturi lyties.

Kiekvienas katinėlis yra tam tikros kartos. Pirmieji katinėliai, kurie nebuvo išveisti, o atsirado žaidime kūrėjų dėka, yra nulinės kartos. Jų palikuonys bus viena karta aukščiau – pirmos kartos. Pirmos kartos katinėlių palikuonys bus antros kartos ir taip toliau. Jei tėvai yra nevienodos kartos, palikuonis bus viena karta aukščiau už aukštesnės kartos tėvą. Nulinės kartos katinėlių skaičius yra ribotas, nes jų nebegalima išveisti, kitų kartų – neribotas.

Vienas katinėlių veisimas kainuoja 0,008 ETH. Veisimui galima naudoti tiek savo katinėlius, tiek tuos, kuriuos galima pasiskolinti veisimui, sumokėjus tam tikrą kainą kitam žaidėjui. Kaip jau minėjome, bet kuris katinėlis gali būti ir tėvas, ir motina. Po veisimo katinėliai (tėvas ir motina) turi atvėsimo (angl. *cooldown*) laikotarpį, kurio metu negalima naudoti jų kitiems veisimams. Laikotarpis varijuoja ir priklauso nuo kartos (kuo aukštesnė karta, tuo laikotarpis ilgesnis) bei nuo to, kiek kartų katinėlis jau buvo naudotas veisimui (kuo daugiau naudotas, tuo atvėsimo laikotarpis ilgesnis). Toliau pateikta lentelė, parodanti ką tik išveistų katinėlių atvėsimo lygmenis (medium.com, 2018):

Karta	Palikuonys	Atvėsimo laikas ir tipas	
0, 1		1 min	Fast
2, 3		2 min	Swift
4, 5		5 min	
6, 7		10 min	Snappy
8, 9		30 min	
10, 11	0	1 h	Brisk
12, 13	1	2 h	
14, 15	2	4 h	Plodding
16, 17	3	8 h	
18, 19		16 h	Slow
20, 21		24 h	
22, 23		2 dienos	Sluggish
24, 25		4 dienos	
26+		1 savaitė	Catatonic

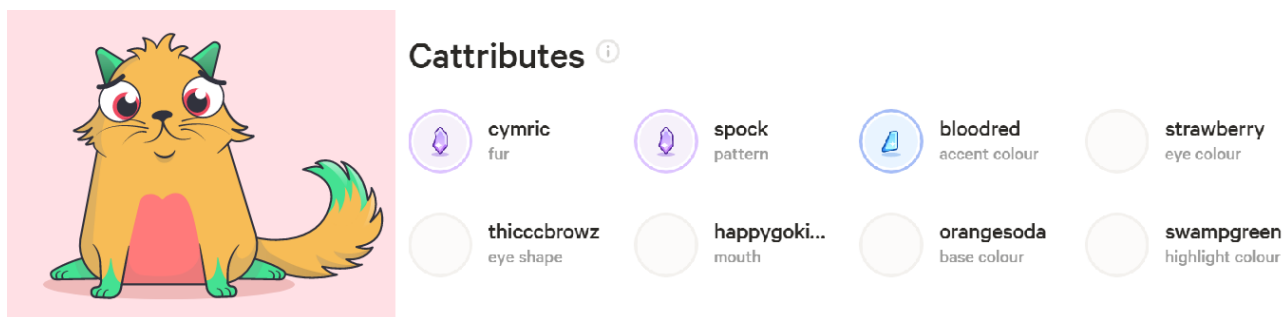
1 pav. Atvėsimo lygmenys

Matome, kad pirmos kartos katinėlis turės 1 minutės atvėsimo laikotarpį, šeštos – 10 minučių. Tam, kad geriau suprastume, kaip atvėsimo laikotarpis keičiasi, turint palikuonių, paimkime pavyzdį, kai katinėlis yra vienuoliktos kartos. Iš pradžių jo atvėsimo laikotarpis bus 1 valanda. Po vieno palikuonio – 2 valandos, antro – 4 valandos, o trečio – 8 valandos. Ilgiausias atvėsimo laikotarpis, kai katinėlis negali turėti palikuonių, yra viena savaitė.

Kaip jau ir minėta anksčiau, kiekvienas katinėlis turi unikalią „DNR“. Ji nulemia ne tik katinėlio išvaizdą, bet ir tai, kaip atrodys ir kokius genus turės palikuonys. Genai slepia ne tik savybes, kurias gali įgauti palikuonis, bet ir tikimybes jas įgauti. Katinėlio genomą yra sudarytas iš blokų, vieną bloką sudaro 4 genai. Kiekvienas katinėlis turi atributus (angl. dar vadinama *cattributes*), kurie didžiąjai daugumai kriptokatinėlių atsispindi išvaizdoje. Pagrindiniai atributai skirstomi į šias kategorijas:

- kailis (angl. fur);
- raštas (angl. pattern);
- akių spalva (angl. eye colour);
- akių forma (angl. eye shape);
- pagrindinė spalva (angl. base colour);
- paryškinių spalva (angl. highlight color);
- akcentų spalva (angl. accent color);
- burna (angl. mouth).

Pažvelkime, kaip šie atributai atsispindi katinėlio išvaizdoje. Paimkime vieno katinėlio pavyzdį (<https://www.cryptokitties.co/>, 2019):



2 pav. Atributai

Matome, kad katinėlio išvaizda yra tarsi sulipdyta pagal šiuos atributus. Šiuo atveju, pagrindinė spalva yra *orangesoda* ir matome, kad katinėlis tikrai yra oranžinio limonado spalvos. Akių forma yra *thiccbrowz*, tai atsispindi matant storus katinėlio antakius. Tačiau katinėlio išvaizdai atributai daro įtaką tik tuo atveju, kai katinėlis yra paprasto tipo. Katinėlių tipai bus apibūdinti šiek tiek vėliau.

Taip pat yra ir papildomi atributai. Nėra iki galo aišku, kokią informaciją jie saugo, nors internete yra įvairių spėliojimų (pavyzdžiui, atributas *Wild*).

Katinėlių savybės kiekviename geno bloke yra skirstomos į:

- Bazinės – savybės, randamos nulinės kartos katinėliuose. Jas perduoti palikuoniui yra 75 % tikimybė.
- Paslėpti genai – savybės, kurios nedaro įtakos katinėlio išvaizdai, tačiau gali atsispindėti palikuonyse. Jas perduoti palikuoniui yra 25 % tikimybė.

Viena savybių rūšis į geną nėra įtraukta. Tai – mutacijos savybės. Jos atsiranda naujai išveistam katinėliui, nors nei tėvo, nei motinos genuose jų nėra.

Toliau pademonstruosime pavyzdį kaip atrodo katinėlio geno sandara (Alanfalcon & Poopie.cat, 2019).



3 pav. Genomo sandara

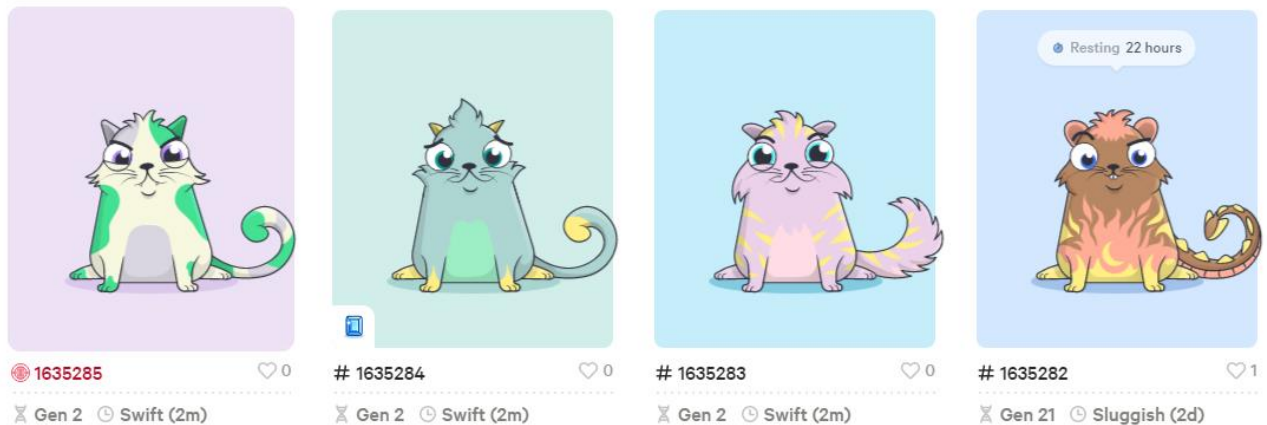
Matome, kad genomą sudaro 12 atributų blokų, o kiekvieną bloką sudaro 4 genai. Paimkime bloką „Eye Shape“. Šiuo atveju bazinė savybė (P) yra wiley. Paslėpti genai yra baddate (H1) ir asif (H2, H3).

Oficialiame tinklapyje katinėlio genomo ir jo sandaros savybių nepamatysime, tačiau prie kiekvieno katinėlio yra pateikiama nuoroda į minėtą puslapį. Tai yra tik papildomas instrumentas, norint daugiau sužinoti apie galimas katinėlių palikuonių savybes.

Kriptokatinėlių tinklapyje galime pastebėti, kad, nors ir visi katinėliai yra unikalūs, dalis katinėlių atrodo kitaip nei visi. Taip yra todėl, kad jie yra skirtingų tipų. Kriptokatinėlius galime skirstyti į tris skirtingus tipus:

- Paprasti (angl. Normal)

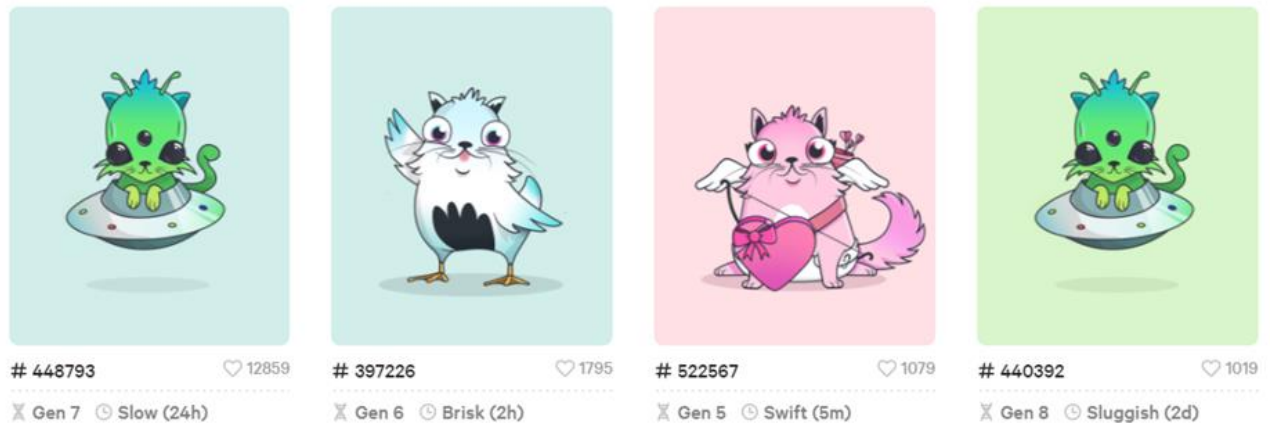
Jų yra didžioji dalis. Šio tipo katinėlio išvaizdą lemia jo genetiniai atributai. Matome, kad šie katinėliai yra sudėlioti iš tam tikrų atributų tipų. Pateiksime pavyzdžius iš oficialaus puslapio (<https://www.cryptokitties.co/>, 2019):



4 pav. Paprastas tipas

- Prabangūs (angl. Fancy)

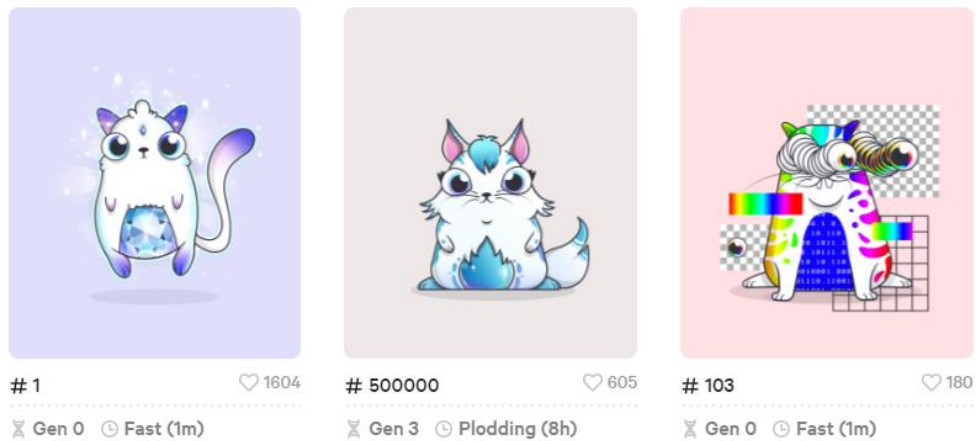
Šio tipo katinėlius įmanoma gauti, jei tėvų genai atitinka tam tikras kombinacijas. Yra limituotas skaičius, kiek tokio tipo kiekvienos rūšies katinėlių galima išveisti. Platformoje dažnai pranešama, kiek konkretaus tipo katinėlių galima išveisti. Taip pat nurodomas laikotarpis, kada tokie katinėliai gali būti gaunami. Šie katinėlių išskirtinumas yra tas, kad jie pasižymi kitokia grafika negu paprasti. Tačiau patys jų paveikslėliai gali būti vienodi, skirtingai nei paprastų katinėlių. Pavyzdžiai (<https://www.cryptokitties.co/>, 2019):



5 pav. Prabangus tipas

- Išskirtiniai (angl. Exclusive)

Šio tipo katinėliai yra panašūs į prabangius, tačiau žymiai retesni, dažnai būna tik po vieną vienetą. Išskirtinių katinėlių išveisti negalima. Tai būna proginiai ir tam tikram tikslui skirti katinėliai. Šiam tipui priklauso pirmasis katinėlis Genesis, 500.000-asis katinėlis ir kiti. Pateiksime kelis pavyzdžius (<https://www.cryptokitties.co/>, 2019):



6 pav. Išskirtinių tipas

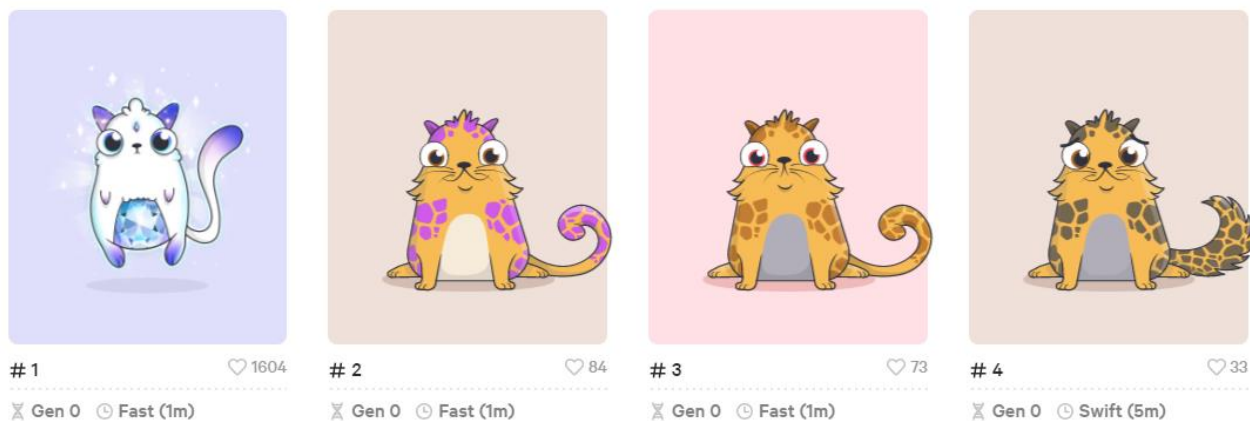
1.6. Kriptokatinėlių kainos įvertinimas

Pardavinėdamas katinėlius, jų turėtojas pats įvertina už kokią kainą nori juos parduoti. Pirkėjas vertina, ar kaina jam priimtina, ar kaina žemesnė / aukštesnė nei rinkos vertė. Čia susiduriama su gana sunkia užduotimi – įvertinti, kiek katinėlis yra vertas. Katinėlių veisėjai susiduria su panašia užduotimi ir sprendžia, kuriuos tėvus panaudoti, kad palikuonis turėtų kuo didesnę vertę. Taigi, šioje dalyje išsiaiškinsime, kas daro įtaką katinėlio kainai.

Yra pateikiamos trys pagrindinės sritys, kurios suteikia katinėliui vertę: retumas, naudingumas ir išvaizda (The many ways to value a kitty, 2018). Toliau aptarsime kiekvieną sritį atskirai ir pamąstysime, kokie kintamieji galėtų lemti katinėlio kainą.

Pirmiausiai pakalbėsime apie retumą. Į šią savybę daugiausiai atsižvelgiama iš kolekcinės perspektyvos. Kolekcionieriai siekia turėti pačias rečiausias kolekcinės vertybes, todėl retumas daro gana stiprią įtaką kainoms. Kai kurie katinėliai yra limituoto leidimo, pavyzdžiui, prabangaus ir išskirtinio tipai. Kitiems katinėliams tam tikrus jų bruožus yra labai sunku išveisti, todėl jie priskiriami retiems.

Kiekvienas katinėlis turi savo identifikacinį numerį. Pirmieji katinėliai (su mažu identifikaciniu numeriu) yra vertinami aukštesne rinkos verte. Ypač tai pastebima su pirmu 100 katinėlių, kur pirmasis yra Genesis, o kiti 99 – jaguarai, pirmieji sukurti katinėliai ir vieninteliai tokiais raštais.



7 pav. Genesis ir jaguarai

Vertė taip pat priklauso nuo katinėlio kartos (kuo žemesnė karta, tuo brangesni katinėliai). Nulinės kartos katinėlių yra limituotas skaičius – 50 000, tai nulemia didesnę retumą ir didesnę kainą. Yra atsižvelgiama ir į mutacijas, kurios vertingiausios, kai įvyksta kuo žemesnėje kartoje.

Naudingumas yra taip pat svarbi savybė. Tai yra ypač aktualu veisiant katinėlius, nes yra bandoma su tam tikrų savybių ir genų tėvais išvesti kuo aukštesnės vertės katinėlį. Čia vėl yra vertinga jau anksčiau paminėta katinėlio karta. Palikuonio karta bus vienu lygiu didesnio skaičiaus, nei maksimalios kartos tarp tėvo ir mamos. Todėl tėvus naudingiausia rinktis kuo žemesnės kartos.

Atsižvelgiama yra ir į tėvų genus. Katinėliai, turintys genus, kurių reikia išvesti prabangaus tipo katinėlius, yra labai paklausūs. Taip pat žiūrima į mutacijas. Katinėlio naudingumas priklauso tiek nuo bazinių, tiek nuo paslėptų genų. Nors paslėpti genai neatsispindi katinėlio išvaizdoje, tačiau tai duoda tikimybę išvesti katinėlius, kurie turėtų paslėpto geno savybes.

Trečia sritis, kurią aptarsime, yra katinėlio išvaizda. Dažniausiai katinėlio savininkui ši savybė yra labai svarbi. Tai yra katinėlio, kaip kolekcinės vertybės, prekinė išvaizda, kuri daro didelę įtaką kainai. Ši sritis yra labiausiai subjektyvi iš visų jau minėtų, nes skirtingas vartotojas gali skirtingai įvertinti išvaizdą.

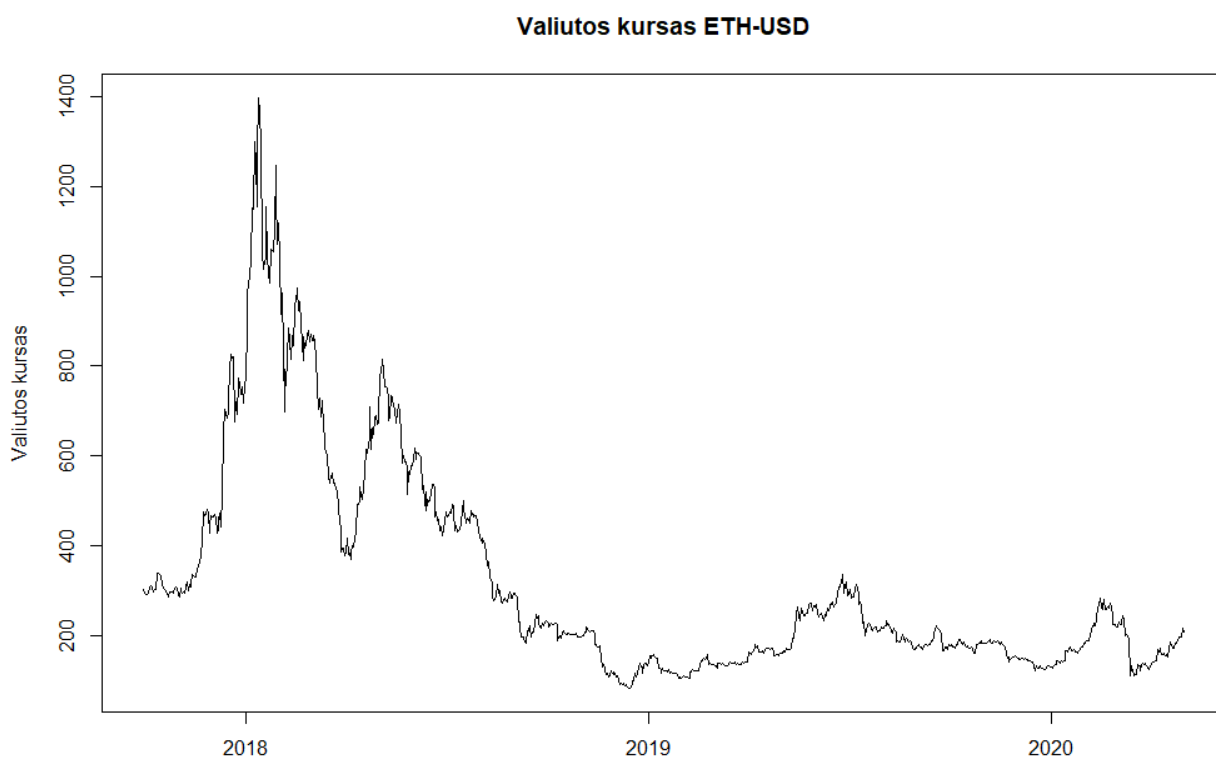
Čia, žinoma, vėl daug įtakos daro baziniai genai. Tačiau šioje srityje yra svarbu ne tik atskiri genai, bet ir jų kombinacijos. Spalvų ir savybių deriniai turėtų atrodyti estetiškai ir derėti.

Dar vienas išvaizdos kriterijus yra katinėlių brangakmeniai (angl. Jewels). Jie parodo kilmingumą. Juos gauna tam tikras skaičius pirmųjų savybes įgavusių katinėlių ir brangakmeniai taip pat yra perduodami palikuoniams.

Tačiau, ypač ties dėl katinėlio išvaizda, yra labai didelis neapibrėžtumas, kaip turėtume įvertinti katinėlio kainą. Kadangi išvaizdos vertinimas yra subjektyvus dalykas, skirtingi žmonės gali įvertinti tą labai skirtingai. Nereikėtų pamiršti, kad ties išvaizdos vertinimu slypi ne tik estetika, bet ir vartotojo sentimentai, jausmai, katinėlio nuotaika bei sąsajos ir panašumai su kažkuo.

Visgi kriptokatinėlių kainai įtakos taip pat gali turėti ir išoriniai kintamieji. Ypač turėtume atkreipti dėmesį į tai, kad katinėliai pardavinėjami už kriptovaliutą. Paimkime pavyzdį populiaros kriptovaliutos – bitkoinų. Bitkoinų kriptovaliuta yra stipriai nepastovi ir jų kainos dinamika yra jautri ir stipriai priklauso tiek nuo investuotojų nuotaikų pasikeitimų, tiek nuo įstatyminių pozicijų

(Dong, 2018). Taip pat minima, kad bitkoinų tipo kriptovaliutos yra rizikingos dėl galimų kainų burbulų. Tai padidina riziką investicijoms, kurios būna susijusios su kriptovaliuta. Kriptokatinėlių platforma yra paremta eterio valiuta. Pažvelkime, kaip eterio valiutos kursas kinta JAV dolerio atžvilgiu.



8 pav. Valiutos kursas ETH-USD (duomenų šaltinis finance.yahoo.com)

Pasitelkus viršuje esantį grafiką, galima teigti, kad valiutos kursas buvo gana stipriai kintantis. 2018 metų pradžioje buvo gana stipriai šoktelėjęs į viršų, o metų pabaigoje sumažėjo maždaug 14 kartų. Todėl tikėtina, kad valiutos kurso svyravimai gali gana smarkiai nulemti katinėlių vertę.

1.7. Kainos vertinimui naudojami algoritmai

Iš tiesų, informacijos, kaip įvertinti kriptokatinėlio kainą, yra labai nedaug. Sudarytų modelių, kurie prognozuotų kainą, taip pat nebuvo rasta. Todėl apžvelgsime, kokie algoritmai yra naudojami prognozuojant nekilnojamo turto kainą. Ši sritis pasirinkta, nes yra atlikta pakankamai nemažai kainos vertinimo tyrimų ir kainos įvertinime turi panašių aspektų. Abiejų sričių objektų kaina priklauso nuo estetikos, naudingumo. Taip pat, abiejų sričių objektai yra unikalūs. Jau minėjome, kad kiekvienas kriptokatinėlis yra skirtingas, tačiau nekilnojamo turto objektai dažniausiai taip pat kiekvienas skiriasi. Net jei nekilnojamo turto projektai yra vienodi, gali skirtis vieta, aukštas ir kitos savybės.

Mašininio mokymo algoritmai buvo taikomi, kuriant modelį, kuris skirtas įvertinti Santjago mieste esantį nekilnojamą turtą (V.H. Masías, 2016). Buvo naudojami keturi algoritmai: atsitiktinis miškas, dirbtiniai neuroniniai tinklai, atraminių vektorių metodas bei tiesinė regresija. Tiksliausius rezultatus čia davė atsitiktinio miško algoritmas, taip pat geri rezultatai pasiekti atraminių vektorių

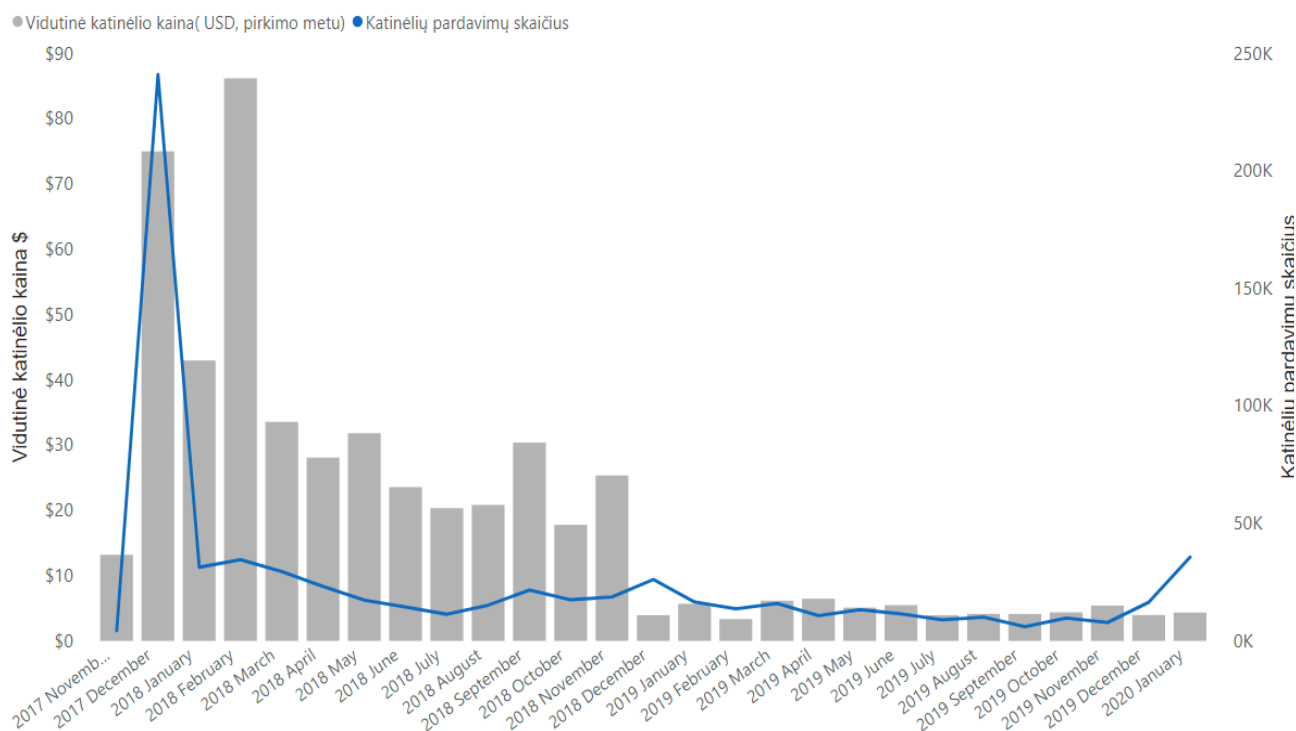
metodu. Tiesinė regresija prognozavo šiek tiek prasčiau, o mažiausiai tikslūs rezultatai pasiekti su dirbtinių neuroninių tinklų algoritmu. Taip pat nekilnojamo turto (butų) kainos buvo vertinamos su Liublianios duomenimis (Marjan Ceh, 2018). Čia buvo lyginami tiesinės regresijos ir atsitiktinio miško modeliai. Daugiau nei du kartus geresnius rezultatus davė atsitiktinio miško algoritmas. Buvo rasta ir daugiau naudojamų algoritmų, pavyzdžiui, viename iš straipsnių nekilnojamo turto kainai vienoje JAV valstijų prognozuoti naudojami *C4.5*, *Ripper*, Naivus Bajeso ir *AdaBoost* algoritmai (Byeonghwa Park, 2015). Tačiau šie algoritmai naudoti klasifikavimui, ar kaina bus mažesnė, ar didesnė nei tam tikra riba.

Taigi matome, kad, norėdami prognozuoti kainą, o ne nustatyti, ar kaina bus aukščiau / žemiau tam tikros ribos, turėsime naudoti regresijos uždaviniams skirtus algoritmus, o ne klasifikacijai skirtus. Vienas geriausiai pasirodžiusių algoritmų apžvelgtuose straipsniuose buvo atsitiktinio miško algoritmas, kurį vertėtų pabandyti panaudoti ir prognozuojant kriptokatinėlių kainą. Taip pat būtų įdomu palyginti rezultatus su kuriuo nors kitu algoritmu.

1.8. Kriptokatinėlių kainos ir pardavimų dinamika laike

Pirmiausiai, tam, kad suprastume, kas gali būti šio žaidimo žaidėjai, pažvelkime į kriptokatinėlių istorinę kainą. Panagrinėkime kainą doleriais ir jos kitimą laike. Pamatysime, kaip keitėsi vidutinė katinėlio kaina ir kokios buvo pardavimų apimtys tiriamuoju laiku.

Vidutinė katinėlio kaina (USD, pirkimo metu) ir katinėlių pardavimų skaičius



9 pav. Vidutinė kriptokatinėlių pardavimo kaina (pardavimo metu, USD) ir parduotų kačiukų skaičius nuo 2017 m. lapkričio iki 2020 m. sausio

Matome, kad vidutinė kaina labiausiai buvo išaugusi 2017 metų gruodį – 2018 metų vasarį. Daugiausiai pardavimų buvo 2017 metų gruodį. Pardavimų skaičius smarkiai nuslūgo jau sekanti mėnesį. O vidutinė kriptokatinėlio kaina smuko į žemumas nuo 2018 metų gruodžio mėnesį.

Tikėtina, kad dideles sumas mokėję asmenys, naudojo žaidimą investavimui. Tačiau, nuslūgus kainų pikui ir pardavimų skaičiui, žaidime galėjo likti dauguma tų, kurie žaidžia žaidimą dėl smagumo, arba bando užsidirbti nedideles sumas. Todėl tikslinė auditorija modeliui šiuo metu būtų pastarieji asmenys, perkantys už mažesnes sumas, nes žaidėjų, mokančių dideles sumas, iki šių dienų išliko nedaug.

1.9. Ateities galimybės virtualioms kolekcinėms vertybėms

Matant, kad kriptokatinėlių populiarumas yra stipriai sumažėjęs, palyginus su 2018 metų pradžia, kyla klausimas, ar 2018 metų pradžios bumai buvo tik vienkartinė sėkmė, ar visgi tai yra tik blokų grandinės technologijos pritaikymo pradžia virtualių kolekcinėms vertybėms srityje, kuri plėsis ir toliau. Apžvelgsime naujoves ir planus šioje srityje.

Pastaruju metu susikūrė daug naujų platformų panašių į *Cryptokitties*, naudojančių blokų grandinės technologiją: *HashPuppies*, *CryptoFighters*, *CryptoCelebrity*, *CryptoAlpaca* ir daugybė kitų. Gana populiarus tapo žaidimas *Gods Unchained*. Šis žaidimas netgi bendradarbiavo su *Cryptokitties*. Turimus kriptokatinėlius buvo galima panaudoti minėtame žaidime. Taip pat buvo sukurtos platformos, kuriose įvairias virtualias vertybes galima pardavinėti ir pirkti vienoje vietoje, pavyzdžiui, jau minėtas puslapis *Opensea*.

Nepakeičiami žetonai gali būti ir skaitmeninis turtas, tačiau taip pat gali būti ir realaus turto nuosavybės įrodymas. Tai gali būti labai brangus meno kūrinys arba tam tikros sporto komandos žaislas. Kai kurie teigia, kad nepakeičiamų žetonų idėja yra sumažinti atskirtį tarp skaitmeninio ir realaus pasaulio (Binance Academy, be datos). Kaip pavyzdį galima pateikti *NBA* komandos *Sacramento Kings* sugalvotą idėją kartu su įmone *CryptoKaiju*, kurios dėka sukūrė 100 vienetų realių *Kaiju* žaislų, kurie susieti su blokų grandine. Buvo planuojama juos nemokamai išdalinti 2019 – 2020 sezono metu. 15 iš šių žaislų leis laimėti tokius prizus, kaip VIP turas arba įvairią atributiką su parašais (NBA, 2019). Kuriant žaislus ir naudojant nepakeičiamus žetonus galima užtikrinti, kad visi žaislai yra unikalūs. Kiekvienas iš jų turi unikalų aprašymą ir bruožus, kuriuos galima atsekti eterio blokų grandinėje. Visi gavę žaislą galės jį įsidėti į savo kriptovaliutos piniginę. Taip pat atsiranda galimybė parduoti nepakeičiamą žetoną atskirai ar kartu su realiu žaislu, kas dar padidina jo vertę.

Kriptokatinėlių kūrėjai – įmonė *Dapper Labs* kuria naują blokų grandinę, kuri vadinama *Flow*. Ši blokų grandinė kuriama daugiausiai būtent žaidimams ir virtualiam turtui. Ji yra kuriama, nes *Ethereum* blokų grandinė tokio tipo platformoms, kaip *Cryptokitties*, turėjo gana nemažų minusų: pasižymėjo transakcijų lėtumu, programavimo sudėtingumu (Gharegozlou, 2019). Be to, kad nauja blokų grandinė bus skirta virtualaus turto kolekcinėms vertybėms, kūrėjai taip pat gana stipriai orientuojasi į menininkus, atlikėjus ir kitus asmenis, kurie įgaus naujus būdus pasiekti savo fanus. Taip pat žaidimus, kur turimas turtas gali būti panaudojamas neribotoje atviroje žaidimų erdvėje. Nepamirštos ir sporto platformos, kurios turės galimybę pardavinėti su sportu susijusių virtualią atributiką. Jau dabar į *Flow* investuoja ir su ja bendradarbiauja tokios įmonės, kaip *Warner Music Group*, *Ubisoft* ir kitos. Yra kuriamos platformos, kurios bus paremtos *Flow* blokų grandine, pavyzdžiui, *NBA Top Shot*, kurios naudotojai galės įsigyti unikalius tikrus krepšinio žaidimų

momentus. Paskelbta, kad *Flow* bloką grandinė taip pat bus įdiegta ir *Cryptokitties* platformai. Šio žaidimo žaidėjai turės galimybę savo kriptokatinėlius iš *Ethereum* bloką grandinės perkelti į *Flow*. Tą padarius, katinėliai įgaus papildomų savybių, pavyzdžiui, pakeis šiuo metu įprastus paprastus katinėlių paveikslėlius į paveikslėlius su animacijomis bei trijų dimensijų galimybėmis.

Kol kas technologijos šioje srityje dar tik po truputį eina į priekį, bet matome, kad yra daug į tai investuojama, didelės įmonės rodo susidomėjimą. Nors ir kriptokatinėlių parduodama mažiau, tai nereiškia, jog kriptovaliutomis paremtos virtualios kolekcinės vertybės lieka praeityje. Puikiai matome, jog atsiranda vis daugiau platformų ir vis daugiau yra investuojama į šią sritį.

1.10. Apžvalgos apibendrinimas

Aptarėme pagrindinius kolekcionavimo ir kolekcinų vertybių aspektus realių bei virtualių vertybių. Taip pat panagrinėjome, kaip bloką grandinės technologija paveikė virtualių kolekcinų vertybių rinką, išsiaiškinome, kas yra *Cryptokitties* ir kaip jis susijęs su kolekcionavimu. Kriptokatinėlių žaidimas susilaukė daug dėmesio, pritraukė nemažai žaidėjų bei investicijų. Šį žaidimą naudojama tiek pramogos, tiek investavimo tikslais. Aptarėme, kad vieni pagrindinių kriterijų, nulemiančių katinėlio kainą, yra retumas, naudingumas, išvaizda ir tuometinė rinkos situacija (eterio ir JAV dolerio kursas). Kadangi kriptokatinėliai turi daug įvairių savybių ir yra unikalūs, tampa sudėtinga nuspėti jų kainą. Todėl tam gali pagelbėti mašininio mokymosi algoritmai. Jų dėka galima būtų sukonstruoti modelį, kuris prognozuotų orientacinę kriptokatinėlio kainą. Tokio modelio sukūrimas būtų naudingas šios platformos žaidėjams, nes šio darbo rašymo metu nebuvo rasta jokių modelių ir įrankių, kurie prognozuotų orientacinę katinėlio vertę. Sukūrus pakankamai tikslų modelį, galima būtų jį panaudoti tolimesnių investavimo strategijų kūrimo.

2. Medžiagos ir tyrimų metodai

2.1. Modeliai

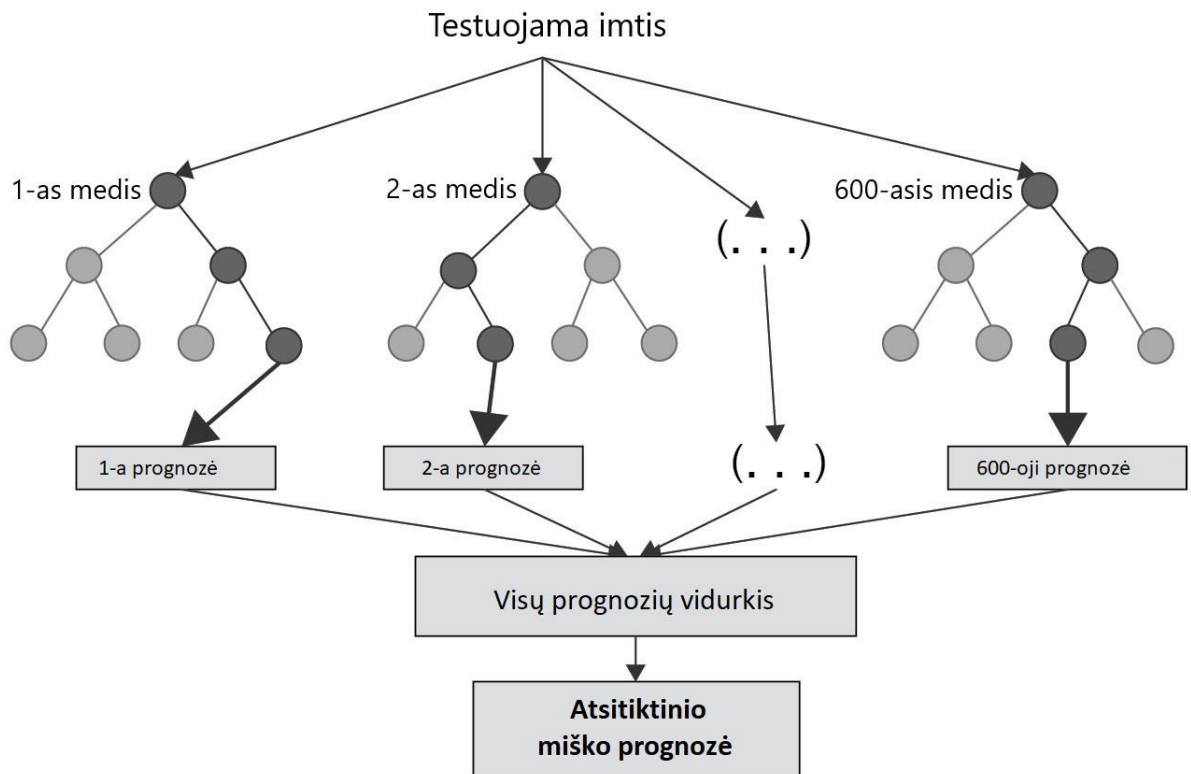
Šiame skyriuje apibrėšime, kaip spręsimė regresijos uždavinius. Kaip ir buvo minėta prieš tai, gerus rezultatus tokio tipo uždaviniams gali duoti atsitiktinio miško algoritmas, todėl jį ir naudosime. Taip pat palyginsime rezultatus su reguliarizuotos regresijos modeliu. Kiekvieną modelį aptarsime atskirai: pažvelgsime į pačius mašininio mokymosi algoritmus ir paminėsime jų svarbiausius aspektus.

2.1.1. Atsitiktinis miškas

Atsitiktinio miško algoritmas gali būti naudojamas tiek klasifikacijos, tiek regresijos uždaviniams. Atsitiktinis miškas yra priskiriamas prie mašininio mokymo algoritmų su mokytoju (angl. supervised). Šis algoritmas pagrįstas tuo, kad sudaro daug atsitiktinių pasirinkimo medžių ir naudoja mašininį ansamblinį mokymąsi. Ansamblinio mokymosi esmė yra sudaryti vieną geriausią modelį iš daug kitų modelių. Pats atsitiktinis miškas naudoja taip vadinamo bagažinės (angl. bagging arba bootstrap aggregating) tipo ansamblinio mokymosi metodą, kai imtys generuojamos iš visų duomenų atsitiktinai paimant tam tikrą kiekį stebėjimų su grąžinimu. Atsitiktinio miško algoritmas kiekvieną medį formuoja atskirai, neatsižvelgdamas į kitus suformuotus medžius.

Knygoje „Statistical Learning from a Regression Perspective“ (Berk, 2008) algoritmo žingsniai apibrėžiami taip:

1. pirmiausiai iš duomenų paimama atsitiktinė stebėjimų imtis su grąžinimu;
2. paimama atsitiktinė imtis kintamųjų be grąžinimo;
3. sukonstruojamas pirmasis klasifikacijos ar regresijos medžių padalijimas duomenims;
4. kartojamas žingsnis nr. 2 kiekvienam padalijimui, kol medis yra norimo dydžio;
5. taip vadinami *out-of-bag* duomenys (duomenys, kurie nebuvo pasirinkti kuriant modelį) išskleidžiami į medį, išsaugoma kiekvienam stebėjimui priskirta reikšmė kartu su prognozuojama reikšme;
6. daug kartų kartojami žingsniai nr. 1-5;
7. suskaičiuojama, kiek kartų stebėjimas priskirtas vienai ar kitai kategorijai stebėjimams, kurie nebuvo naudojami formuojant medžius;
8. kiekvieno stebėjimo reikšmė priskiriama pagal daugumos medžių balsavimą. Regresijos atveju suvidurkinama.



10 pav. Atsitiktinių medžių algoritmas

Atsitiktinių medžių algoritmo schema testuojamai imčiai grafiškai pateiktas aukščiau (Chakure, 2019). Testuojamos imties rezultatai gaunami kiekvieno duomenų įrašo medžių rezultatus suvidurkinant.

Aptarsime šio algoritmo privalumus. Teigiama, kad metodas nėra linkęs priversti modelį persimokyti (Breiman, 2001). Taip pat sakoma, kad tai yra vienas tiksliausių algoritmų, efektyvus dideliame duomenų kiekiui, gali suvaldyti didelį kiekį kintamųjų nepašalinus dalies jų (Chakure, 2019).

2.1.2. Reguliarizuota regresija

Reguliarizacija yra regresijos rūšis, kai įverčių koeficientai yra traukiami link nulio. Ji naudojama tam, kad modeliai sumažintų persimokymo riziką bei tam, kad modelis būtų supaprastintas. Reguliarizacija paremta tuo, kad prie paklaidų kvadratų sumos funkcijos RSS yra pridėjama tam tikra nuobauda. Ir gautą funkciją bandoma minimizuoti. Regresijai be regularizacijos taikome šią nuostolių funkciją:

$$RSS = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2$$

Nagrinėsime trijų tipų regularizacijos būdus.

1. *Ridge* regresija – minimizuoja β koeficientų poveikį modeliui. Parametras λ nurodo reguliarizacijos stiprumą. Verta naudoti, kai tam tikri parametrai daro per didelę įtaką modeliui, versdami jį persimokyti.

$$RSS_{Ridge} = RSS + \lambda \sum_{j=1}^p \beta_j^2$$

2. *Lasso* regresija – ne tik minimizuoja β koeficientų poveikį modeliui, tačiau gali ir visiškai juos pašalinti, jei β tampa lygus 0. Parametras λ nurodo reguliarizacijos stiprumą. Verta naudoti, kai yra naudojama daug kintamųjų, kai dalies kintamųjų svarba modeliui yra nedidelė.

$$RSS_{Lasso} = RSS + \lambda \sum_{j=1}^p |\beta_j|$$

3. *Elastic net* regresija – naudojamos baudos tiek iš *Lasso*, tiek iš *Ridge* regresijos. Koeficientai β yra ir mažinami, ir atsisakoma nesvarbių kintamųjų.

$$RSS_{Elastic\ net} = RSS + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2$$

Elastic net reguliarizacija sujungia tiek *Ridge*, tiek *Lasso* reguliarizacijos funkcionalumus. Šio reguliarizacijos tipo privalumai yra tai, kad šis metodas galimas naudoti net tais atvejais, kai kintamųjų skaičius yra didesnis nei stebėjimų bei šis metodas skatina kintamųjų grupavimo efektą, kur stipriai koreliuojantys kintamieji būna kartu arba naudojami, arba nenaudojami modelyje. (Hastie, 2005).

2.2. Tekstinio lauko vektorizavimas

Yra daug modelių (pavyzdžiui, regresija), kuriems kintamieji turi būti skaitinio tipo. Šiuo atveju, negalėtume į modelį įtraukti tekstinių kintamųjų be papildomų korekcijų. Loginiams kintamiesiems galima sukurti dvireikšmį kintamąjį su reikšmėmis 0 ir 1, kategoriniams kintamiesiems – kiekvienai kategorijai naują kintamąjį, kuris parodytų, ar stebėjimas yra tos kategorijos. Čia susiduriame su iššūkiu, ką daryti, kai kintamasis nėra nei loginis, nei kategorinis, o tekstinis. Tačiau šiai problemai spręsti yra sukurti tam tikri metodai.

Tam, kad tekstinio lauko formatą paverstume skaitiniu, galima panaudoti modelius, vadinamus žodžių ir frazių įterpiniais (angl. embeddings). Šie modeliai skirti tam, kad nagrinėjamus žodžius ir frazes susietų su žodžiais ir frazėmis iš savo žodyno ir atvaizduotų juos skaitiniais vektoriais pagal panašumą.

Šiame darbe tekstinio lauko vektorizavimui bus naudojamas modelis *Sent2Vec*. Tai – paprastas modelis be mokytojo, leidžiantis kurti sakinių įterpinius, naudojant žodžių vektorius kartu su n – gramų įterpiniais. N – gramais vadiname šalia esančių elementų seką iš n elementų. Elementai gali būti tiek žodžiai, tiek žodžių junginiai, tiek raidės ir kiti elementai. Metodas pasižymi efektyvumu, gerais rezultatais, tinkamas dideliems duomenų kiekiams (Matteo Pagliardini, 2018).

Sent2Vec biblioteka yra sukurta *Python* programavimo kalbai. Biblioteka yra paremta *Facebook FastText* biblioteka (GitHub, be datos).

2.3. Modelio patikrinimas

Tam, kad sudarytume patikimą modelį, yra naudojamas modelio patikrinimas. Jei tiesiog sudarytume modelį be jo, negalėtume pasakyti, ar modelis bus toks pat tikslus duomenims, kurie nebuvo įtraukti į modelį. Yra įvairių būdų, kurie naudojami modelio patikrinimui, tačiau aptarsime du populiariausius – *holdout* metodą ir kryžminį patikrinimą.

- *Holdout* metodas. Tai – pati paprasčiausia patikrinimo forma. Duomenų įrašai (atsitiktinai) padalijami į apmokymo ir patikrinimo imtis. Apmokymo imtis naudojama apmokyti modelį, o patikrinimo imtis – įvertinti modelį. Patikrinimo imtis visiškai nebenaudojama modelio kūrimui ir skirta tik patikrinti, kaip modelis reaguoja į naujus duomenis. Įprastai apmokymo imtis yra didesnė nei patikrinimo ir sudaro 60 - 80% visų duomenų, o patikrinimo imtis – likę duomenys. Kartais duomenys yra padalijami ne į 2, o į 3 dalis, įtraukiant patvirtinimo imtį. Šios papildomos imties tikslas – parametrų derinimas.
- Kryžminio patikrinimo metodas. Šis metodas pagrįstas tuo, kad duomenys yra padalijami į k imčių. Viena iš imčių naudojama patikrinimui, o kitos modelio apmokymui. Taip modelis sudaromas su visais skirtingais variantais, kai kiekviena imtis papūna patikrinimo imtimi.

Laiko eilučių modeliams naudojami metodai yra paremti tuo, kad stebėjimų rezultatai gali būti priklausomi nuo laiko momento. Jei modelis bus apmokytas su ateities duomenimis, o patikrintas su praeities, modelis gali tapti ne visai korektiškas. Laiko eilutėms naudojami lizdinio kryžminio patikrinimo metodai (angl. nested cross validation). Aptarsime du būdus, kurie yra gana panašūs į buvusius prieš tai, tačiau atsižvelgta į duomenų atžvilgį laike.

- Įrašai padalijami į apmokymo ir patikrinimo imtis, tačiau tai daroma chronologiškai. Pirmi duomenys naudojami apmokymui, paskutiniai patikrinimui. Šis metodas yra ypač paprastas realizuoti.
- Sekančios dienos grandininis kryžminis patikrinimas. Metodas panašus į paprastą kryžminį patikrinimą. Tačiau paremtas tuo, kad duomenys, suskirstyti į imtis, tačiau atsižvelgiama į laiką ir patikrinimo imtis visada bus paskutiniai duomenys. Pavyzdžiui, jei turėsime 10 dienų duomenis, galėsime pirmų 8 dienų naudoti modelio apmokymui, o 9-ą dieną patikrinimui. Po to panaudosime pirmas 9 dienas apmokymui, o 10-ą dieną paliksime patikrinimui.

Teigiama, kad šie metodai padeda gauti labiau patikimą modelio klaidų įvertinimą ir taip į modelį yra paimama visa galima informacija apmokymui ir patikrinimui, atsižvelgiant į laiko momentą (Christoph Bergmeir, 2012).

2.4. Modelio vertinimas

Apibrėšime paklaidų tipus, kuriuos naudosime vertinant sukurtus modelius.

Pažymėsime N - stebėjimų skaičius. A_t yra lygus tikroms priklausomo kintamojo reikšmėms, o F_t prognozuotoms reikšmėms stebėjimui t , o $t = 1, \dots, N$.

- RMSE (angl. root mean squared error) – šaknis iš vidutinės kvadratinės paklaidos. Šis dydis yra naudojamas gana dažnai vertinant modelių paklaidą, nes yra lengvai interpretuojamas – rodo, kiek vidutiniškai vieno stebėjimo prognozuota reikšmė nukrypusi nuo realios reikšmės. Jis skaičiuojamas taip:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (A_t - F_t)^2}$$

- MdAE (angl. median absolute error) – medianinė absoliutinė paklaida.

$$MdAE = MD_{t=1,N}(|A_t - F_t|)$$

- MAE (angl. mean absolute error) – vidutinė absoliutinė paklaida.

$$MAE = \frac{1}{N} \sum_{t=1}^N |A_t - F_t|$$

Medianinė ir vidutinė absoliutinės paklaidos yra paprastos, lengvai interpretuojamos ir populiaros paklaidos įvertinimui.

Santykiniai dydžiai:

- MAAPE (angl. mean arctangent absolute percentage error) – vidutinė arktangento absoliutinė procentinė paklaida. Tai dydis, kuris parodo absoliučią paklaidą procentais. Jis yra išvestas modifikavus MAPE (angl. mean absolute percentage error), tačiau turi tokius privalumus, kaip kad yra nepriklausomas nuo skalės ir lengvai suskaičiuojamas (Sungil Kim, 2016). MAAPE skaičiuojama pagal formulę:

$$MAAPE = \frac{1}{N} \sum_{t=1}^N (AAPE_t), \quad \text{kai } t = 1, \dots, N$$

$$\text{Kur } AAPE_t = \arctan \left(\left| \frac{A_t - F_t}{A_t} \right| \right)$$

- SMAPE (angl. symmetric mean absolute percentage error) – simetrinė vidutinė absoliutinė procentinė paklaida. Straipsnyje ji pavadinta modifikuota MAPE forma (Makridakis, 1993). Ji apskaičiuojama taip:

$$SMAPE = \frac{1}{N} \sum_{t=1}^N \frac{|A_t - F_t|}{(A_t + F_t)/2}$$

- R² koeficientas, kuris dar vadinamas determinacijos koeficientu. Šis paklaidų vertinimo matas gali būti interpretuojamas, kaip turimo priklausomo kintamojo dispersijos procentas, kuris gali būti paaiškinamas modeliu. Determinacijos koeficientas skaičiuojamas šia formule:

$$R^2 = 1 - \frac{\sum_{t=1}^N (A_t - F_t)^2}{\sum_{t=1}^N (A_t - \bar{A}_t)^2}$$

R^2 reikšmės yra nuo 0 iki 1 ir, todėl, dažnai interpretuojamos kaip procentas. Kuo didesnė reikšmė, tuo geriau modelis prognozuoja priklausomą kintamąjį.

2.5. Kintamųjų svarba

Tam, kad geriau suprastume, kurie kintamieji turi daugiausiai įtakos modeliui, bus skaičiuojama ir grafiškai pavaizduojama kintamųjų svarba.

Pirmiausiai aptarsime kintamųjų svarbos nustatymą atsitiktinio miško modeliui. Atsitiktinio miško modeliui regresijos uždavinio kintamųjų svarba gali būti skaičiuojama keliais būdais.

Vienas iš būdų yra Gini kintamųjų svarba, kitaip vadinama vidutiniu priemaišų sumažinimu (angl. mean decrease in impurity). Jis paremtas tuo, kad skaičiuojama, kiek kartų kintamasis pasirenkamas medžio formavimui, kiek kartų medyje buvo padalijamas, gauti skaičiavimai padauginami ir ši išraiška padalijama iš medžių kiekio. Gini kintamųjų svarbos nustatymo metodas turi minusų. Nustatyta, kad metodas nėra patikimas situacijose, kai kintamųjų matavimų skalė ar kategorinių kintamųjų skaičius skiriasi (Carolin Strobl, 2007).

Dėl minėto Gini kintamųjų svarbos trūkumo, dažniausiai naudojamas kitas būdas – keitimais grįsta kintamųjų svarba (angl. permutation-based variable importance). Šis būdas pirmą kartą pasiūlytas atsitiktiniams miškams L. Breimano (Breiman, 2001) ir yra paremtas tuo, kad kintamųjų reikšmės padalijamos ir matuojamas modelio tikslumo pasikeitimas *out-of-bag* duomenims. Taip yra atsekami kintamieji, kuriuos atskyrus, modelio reikšmių prognozavimas tampa mažiau tikslus.

Regresijos atveju kintamųjų svarbą išmatuoti yra lengviau. Kadangi regresijos modelyje kintamieji yra normalizuojami ir kiekvienam kintamajam regresijos modelyje yra priskiriamas tam tikras koeficientas, šiuos kintamųjų koeficientus galime interpretuoti, kaip kintamojo svarbos matavimo vienetą modelyje.

2.6. Naudojami duomenys

Modeliui sudaryti buvo panaudoti duomenys iš 3 šaltinių.

Pirmas duomenų šaltinis – kriptokatinėlių sandoriai, gauti iš *Kittysales Herokuapp* puslapio (CryptoKitties Sales, 2020). Į šį puslapį yra patalpinti visi buvę kriptokatinėlių pirkimo duomenys. Buvo sugeneruoti visi sandoriai laikotarpio nuo 2017-11-23 iki 2020-01-20. Iš viso buvo gauti 682 548 įrašai. Į šiuos duomenis įeina katinėlių identifikacinis numeris ir sandorio data. Taip pat šiame duomenų šaltinyje pateikta pardavimo kaina trimis skirtingomis išraiškomis: eteriu, JAV doleriais sandorio metu ir JAV doleriais dabartine verte.

Antrasis duomenų šaltinis – informacija apie kriptokatinėlius. Duomenys gauti per oficialią kriptokatinėlių aplikacijų programavimo sąsają (angl. *API*) (Cryptokitties Public API, be datos). Šiuo metu yra išleista tik beta versija. Pateikus prašymą, buvo gauti unikalūs prisijungimo duomenys, su kuriais galima jungtis prie duomenų bazės ir susigeneruoti duomenis pagal tam tikras savybes. Duomenys susideda iš įvairių kriptokatinėlių savybių. Iš viso gauta 100 savybių. Savybių pavyzdžiai duomenyse: identifikacinis numeris, karta, atributai (akių spalva, kailio raštas, burnos

tipas ir panašiai), ar katinėliai yra prabangaus ir išskirtinio tipų, prabangaus ir išskirtinio tipai, pagrindiniai duomenys apie katinėlių tėvus ir kiti duomenys. Yra gana daug neaktualių duomenų modeliui (pavyzdžiui, paveikslėlio puslapio adresas internete) ir savybių, kurių reikšmės yra tuščios ar vienodos visiems katinėliams (pavyzdžiui, pasiūlymo statusas, pasiūlymo kaina). Buvo rinkti duomenys tik tų kriptokatinėlių, kurie buvo bent kartą parduoti laikotarpiu nuo 2017-11-23 iki 2020-01-20. Iš viso buvo gauti 524 270 įrašai.

Dar vienas duomenų šaltinis yra eterio ir JAV dolerio valiutų kursas stebimu laikotarpiu. Duomenys buvo gauti iš puslapio *Yahoo Finance* (Yahoo Finance, 2020).

2.7. Duomenų gavimas ir naudojama programinė įranga

Tam, kad surinktume duomenis iš kriptokatinėlių oficialaus *API* ir iš puslapio, kuriame patalpinti sandoriai, naudojama *Postman* programa. Su *Postman* yra galimybė nurodyti *API* adresą, internetinį adresą su atitinkamais parametrais ir rezultatą tokiu būdu išsaugome *JSON* tipo faile.

Kadangi *API* bei internetinis puslapis naudoja puslapiavimą (taip riboja duomenų paėmimą), reikėjo automatizuoti duomenų gavimą. Tam buvo pasitelkta *Postman* kolekcijų funkcija. Sukūrus kolekciją, joje galima nurodyti kintančius parametrus, šiuo atveju puslapiavimo parametrus. Kiekvieną kartą, kviečiant tą patį adresą, buvo pakeičiamas puslapis į tolimesnį ir gauti duomenys iš visų puslapių jungiami.

Automatiškai išsaugoti kiekvieno adreso rezultatą *Postman* programa neleidžia. Dėl to turėjo būti paleista *Node* programa, kuri sekė *Postman* adresų iškvietimus ir naujus rezultatus pridėdavo į tą patį *JSON* tipo failą. Buvo pasinaudota programa *ResponseToFile-Postman* (*ResponseToFile-Postman*, be datos). Taip buvo sukurti *JSON* duomenų failai, kurie vėliau buvo naudojami mašininio mokymo modelių kūrimui.

Tolimesniems skaičiavimams yra naudojama *R* programinė įranga su *Rstudio* aplinka. *JSON* tipo duomenų failai buvo paversti į *CSV* tipo failus tam, kad būtų sutvarkyta failų struktūra modeliams bei būtų sumažinti duomenų failai. Duomenų apdorojimui bei modelių kūrimui taip pat buvo naudojama *R* programa.

Taip pat bus naudojama programa *Python* tam, kad galėtume vektorizuoti tekstinį lauką.

2.8. Modelių kūrimas su R

2.8.1. Atsitiktinis miškas

Tam, kad surastume optimalius modelio parametrus, naudosime *tuneRanger* funkciją iš *Ranger* paketo programoje *R*. Ši funkcija gali būti naudojama ir detekcijos uždaviniams, tačiau mūsų atveju naudojama regresija. Parametrai, kuriuos reikės suderinti:

Mtry – parametras, rodantis atsitiktinai pasirinktų kintamųjų skaičių, kurie bus panaudojami konstruojant medį.

Min.node.size – parametras, rodantis minimalų mazgų skaičių, kuris bus naudojamas konstruojant medį.

Paskaičiuotus gaunami rezultatai apmokymo imčiai įvairioms parametru kombinacijoms. Pagal tai, kur gaunama mažiausia kvadratinė paklaida, pasirinksiame parametrus, su kuriais ji gauta.

Parametrus įsistatysime į *Ranger* funkciją ir sukursime modelį su apmokymo imtimi. Ši funkcija pasižymi greitumu, todėl ypač tinka dideliems duomenų kiekiams.

2.8.2. Reguliarizuota regresija

Modelio kūrimui ir parametru derinimui buvo naudota funkcija *cva.glmnet* iš paketo *glmnetUtils*. Ši funkcija sukuria regresijos modelį, pasitelkdama *Elastic net* regularizaciją bei kryžminį patikrinimą. Funkcija naudinga tuo, kad gali sukurti modelius iš karto kelioms regularizavimo reikšmėms. Naudojama kryžminė validacija apmokymo imčiai tam, kad suderintume *alpha* ir *lambda* parametrus.

2.9. Kintamieji

Aptarsime, kokio tipo ir pobūdžio duomenis turime. Atskirai apibrėšime skaitinius, kategorinius ir loginius kintamuosius. Taip pat paminėsime vieną tekstinį kintamąjį.

Turime priklausomą kintamąjį katinėlio pardavimo kainą JAV doleriais ir eteriu. Tai – skaitinis kintamasis.

Toliau apibrėšime turimus nepriklausomus kintamuosius.

Pirmoje lentelėje yra pateikti nepriklausomi skaitiniai kintamieji, naudoti modelyje. Palyginus su kitų tipų kintamųjų skaičiumi, skaitinių kintamųjų modelyje turime gana nedaug. Svarbiausi kintamieji yra karta, atvėsimo lygis, valiutos kursas ir kiti. Reikėtų atkreipti dėmesį, kad modelyje yra naudojami katinėlių tėvų identifikaciniai numeriai, nes neturime labai daug duomenų apie tėvus. O taip pat yra tikėtina, kad katinėliai su mažesniu identifikaciniu numeriu turės didesnę vertę, nes atsirado žaidime vieni pirmųjų.

1 lentelė. Skaitiniai kintamieji

generation	Karta
prestige_ranking	Prestižo reitingas
fancy_ranking	Prabangumo reitingas
status.cooldown_index	Atvėsimo indeksas
purrs.count	Kiek kartų katinėlis pamėgtas kitų vartotojų
matron.id	Motinos identifikacinis numeris
matron.generation	Motinos karta
matron.status.cooldown_index	Motinos atvėsimo lygmuo
sire.id	Tėvo identifikacinis numeris
sire.generation	Tėvo karta
sire.status.cooldown_index	Tėvo atvėsimo lygmuo
ethusd	Valiutos ETH ir \$ kursas pirkimo dieną (uždarymo kaina)

Tarp kategorinių kintamųjų vieni svarbiausių yra įvairūs išvaizdos atributai, taip pat katinėlių tipai pagal tam tikrus kriterijus.

2 lentelė. Kategoriniai kintamieji

language	Kalba
kitty_type	Kačiuko tipas
prestige_type	Prestižo tipas
matron.color	Motinos spalva
matron.language	Motinos kalba
matron.fancy_type	Motinos prabangumo tipas
sire.color	Tėvo spalva
sire.language	Tėvo kalba
sire.fancy_type	Tėvo prabangumo tipas
body	Kūno atributas
coloreyes	Akių spalvos atributas
eyes	Akių formos atributas
pattern	Rašto atributas
mouth	Burnos tipo atributas
colorprimary	Pirminės spalvos atributas
colorsecondary	Antrinės spalvos atributas
colortertiary	Tretinės spalvos atributas

Duomenyse yra gana nemažai loginių kintamųjų. Dauguma susiję su katinėlių tipais (prabangaus, išskirtinio) ir įvairiais su veisimu susijusiais ypatumais. Taip pat dalis kintamųjų susiję su katinėlio tėvais.

3 lentelė. Loginiai kintamieji

is_fancy	Ar prabangus
is_exclusive	Ar išskirtinis
hatched	Ar išsiritęs
is_prestige	Ar prestižinis
is_special_edition	Ar specialaus leidimo
status.is_ready	Ar pasiruošęs daugintis
status.is_gestating	Ar laukiasi palikuonių
matron.is_fancy	Ar motina prabangi
matron.is_exclusive	Ar motina išskirtinė
matron.status.is_ready	Ar motina pasiruošusi daugintis
matron.status.is_gestating	Ar motina laukiasi palikuonių
sire.is_fancy	Ar tėvas prabangus
sire.is_exclusive	Ar tėvas išskirtinis
sire.status.is_ready	Ar tėvas pasiruošęs daugintis
sire.status.is_gestating	Ar tėvas laukiasi palikuonių
sire.owner.hasDapper	Ar tėvo savininkas turi <i>Dapper</i> piniginę

owner.hasDapper	Ar kačiuko savininkas turi <i>Dapper</i> pinigę
hatcher.hasDapper	Ar kačiuką išveisęs asmuo turi <i>Dapper</i> pinigę

Dar vienas kintamasis, kuris bus naudojamas ir nebuvo įtrauktas į prieš tai buvusias kategorijas – tekstinis kintamasis. Turėsime vieną tokio tipo kintamąjį, kuris apibūdina kriptokatinėlio biografiją. Tai yra laisvos teksto formos laukas.

Konstruojant modelius, turėsime atsižvelgti į tai, kad turime skirtingų tipų kintamuosius. Todėl duomenys turės būti atitinkamai paruošti prieš modelių kūrimą.

3. Tyrimų rezultatai ir jų aptarimas

3.1. Duomenų paruošimas

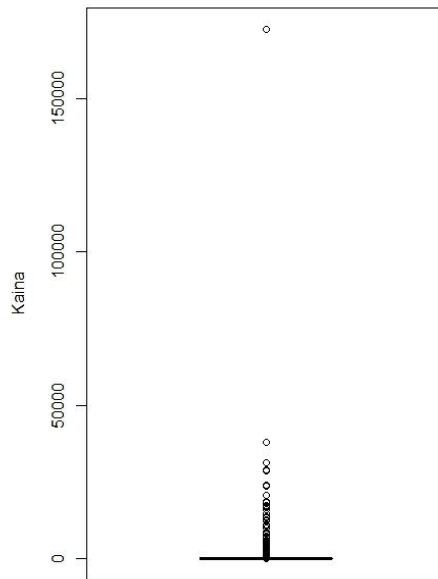
Pirmiausiai duomenys yra paruošiami modeliui. Sandorių duomenys sujungiami su informacijos apie kriptokatinėlius duomenimis pagal identifikacinį numerį. Prie gauto duomenų rinkinio prijungiama eterio ir dolerio valiutų kurso informacija pagal datą. Pašalinami neaktualūs kintamieji.

Duomenyse buvo gana daug pasikartojančių sandorių (tie patys katinėliai pirkti už tą pačią kainą). Taip pat kai kurie katinėliai buvo pirkti daugiau nei 1 kartą su skirtingomis kainomis, tačiau kitos katinėlio savybės nebuvo pakitusios. Atsitiktinio miško algoritmui tai gali trukdyti sukurti tikslų modelį, todėl buvo nuspręsta, kad, jei katinėlis buvo pirktas daugiau nei 2 kartus, duomenyse palikti tik pirmą ir paskutinę transakciją. Atsitiktinio miško algoritmas tokių pačių reikšmių priklausomąjį kintamąjį suvidurkina. Todėl toms reikšmėms, kurios buvo vienodos, katinėlio kaina išliks ta pati. Skirtingoms reikšmėms bus paimtas vidurkis.

Dar vienas duomenų paruošimo iššūkis – sutvarkyti tuščias reikšmes, kurių buvo gana nemažai. Didžioji dalis kintamųjų buvo neskaitinio tipo ir jiems tuščios reikšmės buvo priskirtos todėl, kad kintamieji neturėjo tos savybės. Pavyzdžiui, katinėliams, kurie yra priskiriami prie išskirtinių tipo, atributų kintamieji yra tušti, nes tokie katinėliai neturi atributų. Todėl šios tuščios reikšmės buvo paliktos ir priskirtos naujam faktoriaus lygiui. Tokie patys pakeitimai buvo padaryti ir loginiams kintamiesiems – po pakeitimo į kategorinius kintamuosius, buvo pridėtas naujas faktoriaus lygis, kuris vadinasi NA. Taip pat buvo ir skaitinių kintamųjų su tuščiomis reikšmėmis. Tačiau šios tuščios reikšmės taip pat dažnai turėjo prasmę. Pavyzdžiui, nulinės kartos katinėlio tėvų kartos. Nulinės kartos katinėliai buvo pirmieji pridėti į žaidimą ir buvo sukurti žaidimo kūrėjų, nebuvo išveisti iš kitų katinėlių ir neturėjo tėvų. Todėl tokioms tuščioms reikšmėms buvo priskirta nulinė reikšmė, nes, pavyzdžiui, įvedus vidurkį tuščioms reikšmėms, gali būti neigiamai paveiktas modelio tikslumas.

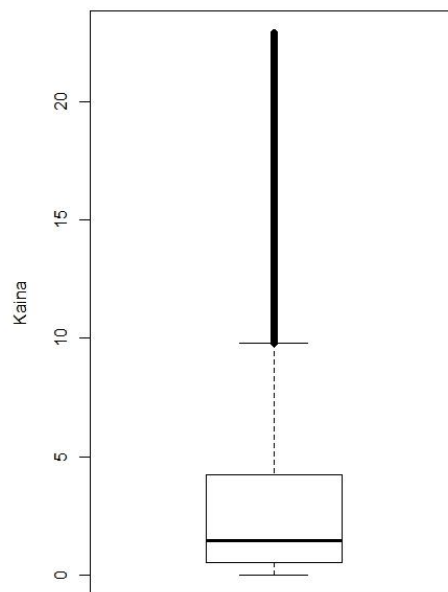
Regresijos atveju galima naudoti tik skaitinius kintamuosius. Taigi, po prieš tai minėto sutvarkymo, kaip atsitiktinio miško algoritmo atveju, visiems kategoriniams kintamiesiems buvo sukurti fiktyvūs kintamieji atskirai kiekvienam kategoriniam lygiui. Tam, kad kintamųjų skalių įtaka neturėtų didelės reikšmės, skaitiniai kintamieji buvo normalizuoti, naudojant R funkciją *preProcess*, metodus *center* ir *scale*.

Dar vienas duomenų tvarkymo žingsnis – pašalintos priklausomojo kintamojo (katinėlio kainos) išskirtys iš duomenų. Išskirtys gali pabloginti modelio prognozavimą, todėl tai padaryta tam, kad patikslintume modelį. Pažvelkime, kaip atrodo katinėlio kainos duomenų stačiakampė diagrama prieš išskirčių pašalinimą.



11 pav. Kainos stačiakampė diagrama prieš išskirčių pašalinimą

Išskirtys pašalintos apskaičiuojant tarpkvartilinį plotą tarp 0,25 ir 0,75 kvartilų ir jį padauginus iš 1,5. Gautas skaičius pridedamas prie 0,75 kvartilio, atimamas iš 0,25 kvartilio. Stebėjimai, nepatenkantys į gautą intervalą, yra pašalinami. Pažvelkime, kaip kainos stačiakampė diagrama atrodo dabar.



12 pav. Kainos stačiakampė diagrama po išskirčių pašalinimo

Dabar duomenys atrodo žymiai tinkamesni modeliui sudaryti.

Tolesniuose skyreliuose aptarsime, kad dalis modelių bus sudaryti su logaritmuotu priklausomu kintamuoju. Reikėtų atkreipti dėmesį, kad šiuo atveju išskirtys pašalintos po logaritmovimo. Todėl modeliuose, kai priklausomas kintamasis yra logaritmuotas, yra skirtingas kiekis duomenų įrašų negu modelyje, kur priklausomas kintamasis nelogaritmuotas.

3.2. Išvestiniai kintamieji

Amerikos dolerio ir eterio valiutų kursui buvo išvesta daugiau kintamųjų. Buvo paskaičiuotas trisdešimtys dienų slenkantis vidutinis valiutų kursas. Taip pat kurso pokyčiai dienos ir trisdešimtys dienų tikraja verte ir procentais.

Pagal tai, kada įvykdyta transakcija, išvesti kintamieji, nusakantys savaitės dieną ir mėnesio pavadinimą, kada katinėlis buvo nupirktas. Šie kintamieji gali padėti modeliui atpažinti sezoninę įtaką, jei tokia yra.

Kai kurie katinėliai turėjo papildomus atributus. Pavyzdžiui, du skirtingus tipus akių spalvos arba du tipus kūno formos. Tai yra programinė klaida, tačiau katinėliams prideda unikalumo. Kadangi išskirtinumas gali pridėti vertės katinėliui, buvo sukurtas naujas kintamasis, kuris nusako, ar katinėlis turi papildomų atributų.

Kai kuriems kategoriniams kintamiesiems (atributams, katinėlio tipui, prabangumo tipui ir kitiems) buvo sukurti nauji kintamieji, kurie skaičiuoja kiekvieno kategorinio kintamojo lygio dažnį. Taip modelis gali turėti pranašumą atpažįstant, ar katinėlio savybės yra retos, ar dažnos. Mažesnio dažnumo atributai katinėliui gali pridėti papildomos vertės.

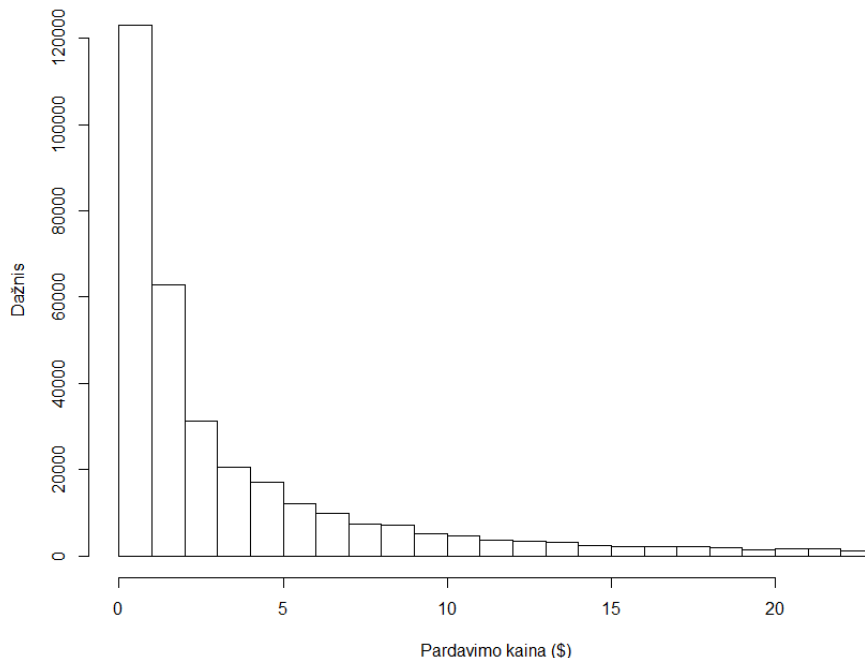
Taip pat paskaičiuotas dažnis visų atributų porų dažnių kategoriniams kintamiesiems. Selektyvus kačiukų veisimas dažnai reikalauja, kad katinėlis turėtų iš karto kelis atributų požymius tam, kad galima būtų išveisti tam tikro tipo katinėlį. Pavyzdžiui, žaidėjas sužino, kad tam tikro prabangaus tipo katinėlį galima išveisti su vienu iš tėvų, kuris turi *Munchkin* tipo kūną ir akių spalvos tipą *Crazy*. Tokiu atveju žaidėjas ieškos būtent tokio tipo katinėlio ir bus pasiryžęs už jį mokėti daugiau. Kadangi minėjome, kad pirkėjui svarbus yra funkcionalumas veisiant, šie kintamieji, gali padėti atskirti, kuriuos katinėlius yra vertinga pirkti dėl veisimo. Taip pat iš dalies šie kintamieji gali atspindėti ir estetinę katinėlio vertę, galbūt tam tikros atributų poros gali atrodyti patrauklios žaidėjui.

Modelyje taip pat panaudosime kiekvienam katinėliui, jo tėvui ir motinai priskirtą paskutinę kainą. Kiekvienam įrašui paskutinė kaina bus surasta individualiai, žiūrint į prieš tai buvusius pardavimus. Jei katinėlis, jo tėvas ar motina nebuvo anksčiau parduotas, reikšmė bus priskiriama nuliui. Taip pat priskirsime reikšmę nuliui, jei katinėlio tėvai neegzistuoja.

Modeliui, su kuriuo bus gauti geriausi rezultatai, bus įvesti papildomi kintamieji. Patikrinsime, ar galima dar pagerinti modelį, įtraukus katinėlio tekstinio lauko (biografijos) kintamuosius. Tam tekstinis laukas bus vektorizuojamas, naudojama *Python* funkcija *sent2vec*. Gautas vektorius kiekvienam duomenų įrašui bus pridedamas prie turimų kintamųjų ir skaičiuojamas su *R* programa.

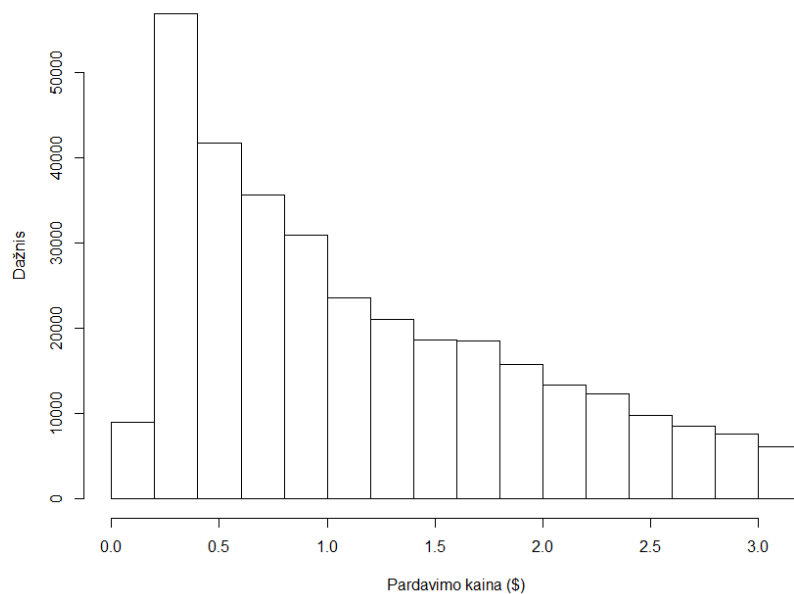
3.3. Priklausomo kintamojo transformacija

Kadangi priklausomas kintamojo Y pasiskirstymas nėra panašus į normaliojo dydžio, jį daliai modelių logaritmuosime. Ypač svarbu tą daryti regresijos modeliams. Pateikiu histogramos pavyzdį, kai kriptokatinėlio kaina yra prognozuojama doleriais.



13 pav. Pardavimo kainos histograma prieš logaritmvimą

Naudojamas logaritmvimas $Y = \ln(Y + 1)$. Gavus prognozių rezultatus, pardavimo kaina bus atstatoma į pradinę. Duomenys po logaritmvimo atrodo taip:



14 pav. Pardavimo kainos histograma po logaritmvimo

Daliai modelių pardavimų kaina nebus transformuota. Palyginsime, ar modeliams ši transformacija padeda tiksliau prognozuoti katinėlių kainą.

3.4. Modeliai

Šiame darbe mūsų tikslas yra suformuoti modelį, kuris nustatytų orientacinę katinėlio rinkos kainą. Todėl turime priklausomą kintamąjį Y – katinėlio kainą. Kainą bus bandoma nustatyti dviem valiutomis – eteriu ir doleriais. Naudosime du modelius – reguliarizuotą regresiją ir atsitiktinius medžius. Taip pat bus žiūrima, ar modeliams turės įtakos logaritminė kintamojo Y transformacija. Palyginsime, kuris modelis, kuria valiuta ir, ar logaritmuojus, Y prognozės duoda tiksliausius rezultatus. Bus sukurti 8 modeliai:

1. Atsitiktinio miško modelis, valiuta USD
2. Atsitiktinio miško modelis, valiuta USD, Y logaritmuotas
3. Atsitiktinio miško modelis, valiuta ETH
4. Atsitiktinio miško modelis, valiuta ETH, Y logaritmuotas
5. Reguliarizuotos regresijos modelis, valiuta USD
6. Reguliarizuotos regresijos modelis, valiuta USD, Y logaritmuotas
7. Reguliarizuotos regresijos modelis, valiuta ETH
8. Reguliarizuotos regresijos modelis, valiuta ETH, Y logaritmuotas

Papildomai, geriausiam modeliui, bus sukurtas dar vienas modelis. Bus išbandyta tekstinio lauko – kačiuko biografinės vektorizacija ir palyginta, ar tai gali pagerinti rezultatus.

Duomenys išskirstyti į apmokymo ir patikrinimo imtis. Pasirinkta 90% duomenų skirti apmokymui, o 10% duomenų patikrinimui, nes duomenų įrašų turime gana daug. Atsižvelgta į tai, kad duomenų struktūra priklauso nuo laiko atžvilgio. Todėl apmokymui paimti pirmi 90% duomenų, o patikrinimui – paskutiniai 10%.

3.5. Tyrimų rezultatai

3.5.1. Paklaidų rezultatai

Pirmiausiai apžiūrėsime atsitiktinių medžių modelių paklaidų rezultatus gautus patikrinimo duomenims. Gautų modelių vertinimui naudosime šias metrikas: determinacijos koeficientą R^2 , RMSE, MdAE, MAE, MAAPE, SMAPE.

4 lentelė. Modelių rezultatai.

	R ²	RMSE	MdAE	MAE	MAAPE	SMAPE
Atsitiktinis miškas (USD)	0,5085	2,6373	1,0152	1,6838	82,1711	81,9741
Atsitiktinis miškas (USD), log(Y+1)	0,3900	7,5829	0,7081	2,6209	74,4463	75,1806
Atsitiktinis miškas (ETH)	0,3817	0,0092	0,0048	0,0066	77,4901	76,1207
Atsitiktinis miškas (ETH), log(Y+1)	0,3857	0,0093	0,0048	0,0067	77,5694	76,2350
Reguliarizuota regresija (USD)	0,3807	3,0409	1,5166	2,0554	93,2102	100,6589
Reguliarizuota regresija (USD), log(Y+1)	0,2392	8,8692	1,0813	3,0800	83,4204	89,7479
Reguliarizuota regresija (ETH)	0,2519	0,0103	0,0059	0,0076	83,5887	85,5298
Reguliarizuota regresija (ETH), log(Y+1)	0,2583	0,0103	0,0058	0,0076	83,3502	85,1087

Matome, kad atsitiktinių medžių algoritmas davė geresnius rezultatus nei reguliarizuota regresija. Taip pat tiek atsitiktinių medžių, tiek reguliarizuotos regresijos atveju geresnius rezultatus gavome prognozuodami kriptokatinėlių kainą doleriais, negu eteriu.

Didžiausias gautas R² yra lygus apytiksliai 0,51. Tai reiškia, kad modelis paaiškina 51% katinėlio kainos sklaidos kitų kintamųjų atžvilgiu. Vidutiniškai modelis nukrypsta nuo realios reikšmės 2,64 \$. Kadangi atmetus išskirtis nagrinėjame transakcijas, kurių vertė ne didesnė nei 30\$, ši paklaida gali būti per didelė katinėliams su maža kaina, tačiau santykinai nedidelė didesnės kainos kriptokatinėliams.

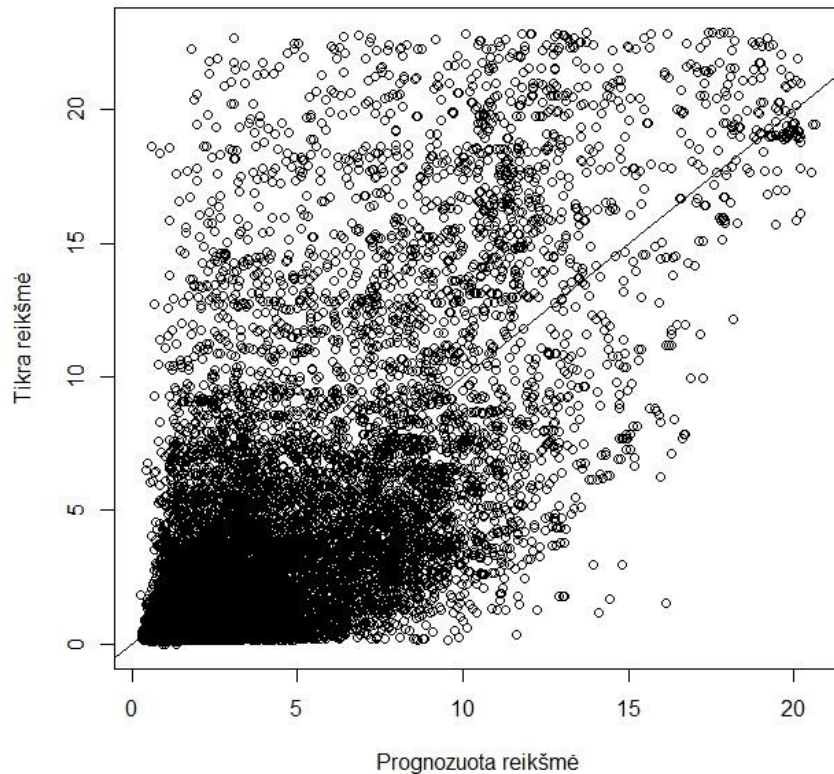
MAAPE ir SMAAPE rezultatai nėra geri – vidutinė procentinė paklaida visais atvejais yra didesnė daugiau nei 74 %. Šiuo atveju negalime lyginti RMSE tarp modelių, nes skiriasi Y kintamojo skalė (doleriais ir eteriu) bei todėl, kad išskirtys duomenims buvo pašalintos skirtingais momentais. Tais atvejais, kai buvo naudojama Y transformacija logaritmuojant, išskirtys pašalintos po logaritmvimo. Todėl šie modeliai sudaryti su platesne kačiukų kainos skale nei modeliuose, kur Y nebuvo logaritmuotas.

Kintamojo Y transformacija logaritmuojant pasiteisino tik tais atvejais, kai modeliai buvo kuriami su eterio valiuta. Tačiau modelio pagerinimas išties minimalus, pavyzdžiui atsitiktinio miško atveju R² pakilo nuo 0,3817 iki 0,3857.

3.5.2. Prognozuotos ir tikros reikšmės grafikai

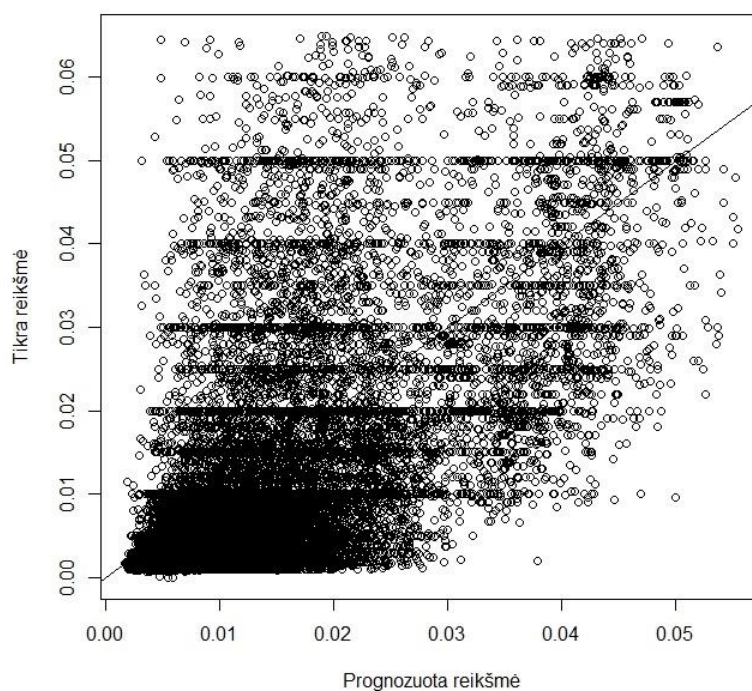
Toliau pažvelgsime, kaip atrodo grafikai tarp prognozuotų ir tikrų reikšmių skirtinguose modeliuose. Tačiau apžvelgsime ne visus modelius, o 4, kurie geriausiai atspindi skirtingų modelių rezultatus.

Atsitiktinių medžių doleriais grafikas pateikiamas žemiau (**15 pav.** Atsitiktinio miško modelis (USD), patikrinimo imties tikros ir prognozuotos reikšmės. Didelė dalis reikšmių yra nedidelės – iki 5\$. Modeliui sunkiai sekėsi prognozuoti dideles reikšmes. Tačiau matome, kad nemaža dalis mažesnių reikšmių buvo pervertinta, pavyzdžiui, kai katinėlio reali kaina yra tarp 0 – 1\$, modelis prognozuoja, kad jis yra vertas 3 – 6\$.



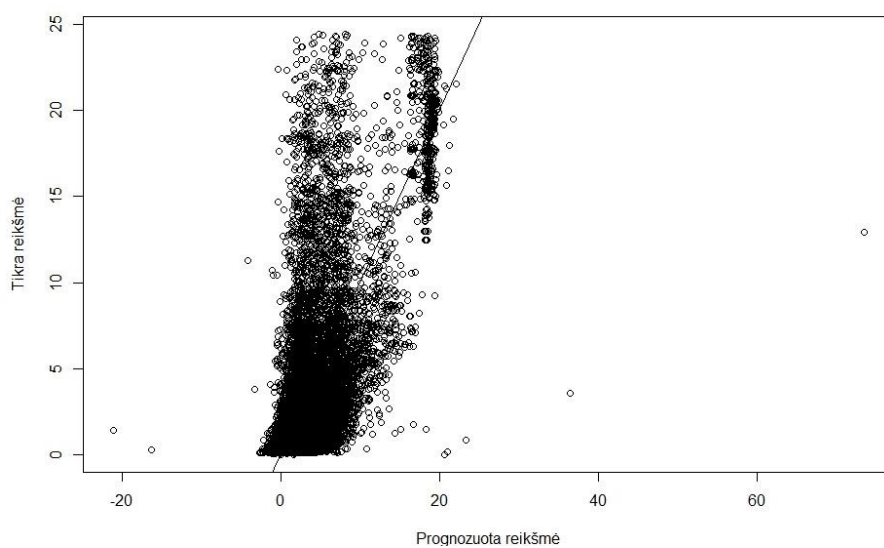
15 pav. Atsitiktinio miško modelis (USD), patikrinimo imties tikros ir prognozuotos reikšmės

Pažvelkime į kitą modelį – atsitiktinis miškas, prognozė eteriu, su logaritmine transformacija (**16 pav.** Atsitiktinio miško modelis (ETH) su logaritmine transformacija, testavimo imties realios ir prognozuotos reikšmės). Matomos panašios tendencijos, kaip ir prieš tai nagrinėtame grafike. Šiek tiek didesnis skirtumas kainos skalės struktūroje. Kadangi valiuta eteriu ta pačia skaitine išraiška yra vertingesnė nei doleriais, skalė yra siauresnė. Puikiai matosi, kad iš tiesų žmonės linkę mokėti apvalias sumas už kriptokatinėlius, todėl didelė dalis kainų yra lygi, pavyzdžiui, 0,01 ar 0,02 ETH. Tačiau čia puikiai pavaizduota, kad modelis tos pačios vertės katinėlių yra linkęs įvertinti labai skirtingai. Pavyzdžiui, kai reali vertė yra lygi 0,05 ETH, prognozuojama jos vertė buvo nuo netolimos nuliui reikšmės iki 0,05 ETH. Šią tendenciją matome ir ties kitomis reikšmėmis.



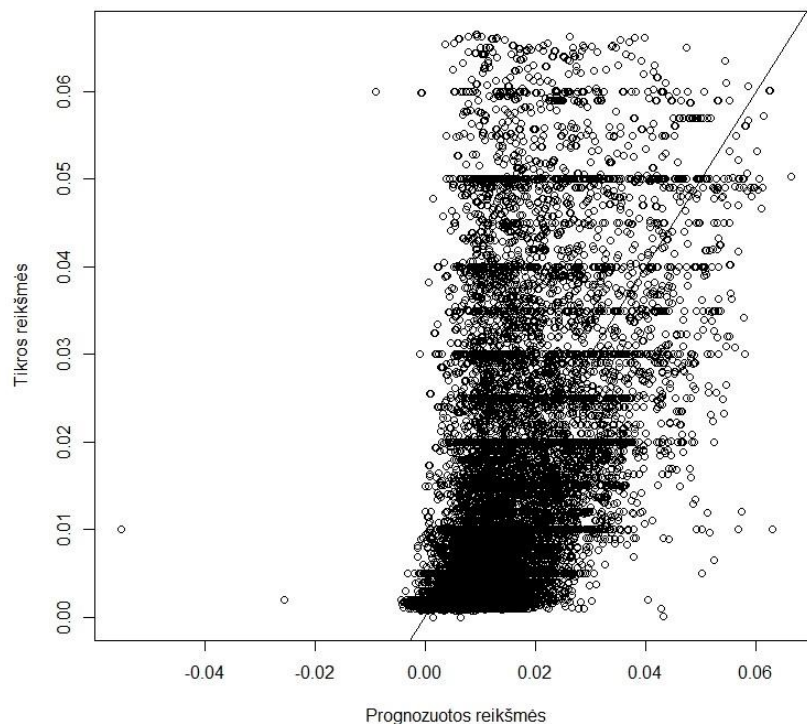
16 pav. Atsitiktinio miško modelis (ETH) su logaritmine transformacija, testavimo imties realios ir prognozuotos reikšmės

Pereikime prie reguliarizuotos regresijos modelių. Pažvelkime į reguliarizuotos regresijos modelį, kai prognozuota doleriais (**17 pav.** Regresijos modelis (USD), testavimo imties realios ir prognozuotos reikšmės. Šis modelis gana prastai įvertina didesnes reikšmes. Didelė dalis reikšmių, kurios buvo smarkiai didesnės už 0, buvo įvertintos kainai artimai nuliui. Taip pat grafike pavaizduotos kelios gana grubios klaidos, kai prognozuotos reikšmės buvo tarp 30 ir 100 \$, tačiau iš tikro buvo lygios 3 – 15 \$. Taip pat atsirado šiek tiek prognozuotų neigiamų kainų.



17 pav. Regresijos modelis (USD), testavimo imties realios ir prognozuotos reikšmės

Reguliarizuotos regresijos modelio grafikas eteriu su logaritmine transformacija (**18 pav.** Regresijos modelis (ETH) su logaritmine transformacija, testavimo imties realios ir prognozuotos reikšmės atrodo gana panašiai, kaip ir prieš tai nagrinėto regresijos modelio doleriais grafikas, tačiau nėra tokių didelių išskirčių dėl per didelės prognozuotos kainos. Tačiau šiame modelyje taip pat buvo prognozuotų neigiamų pirkimo kainų.

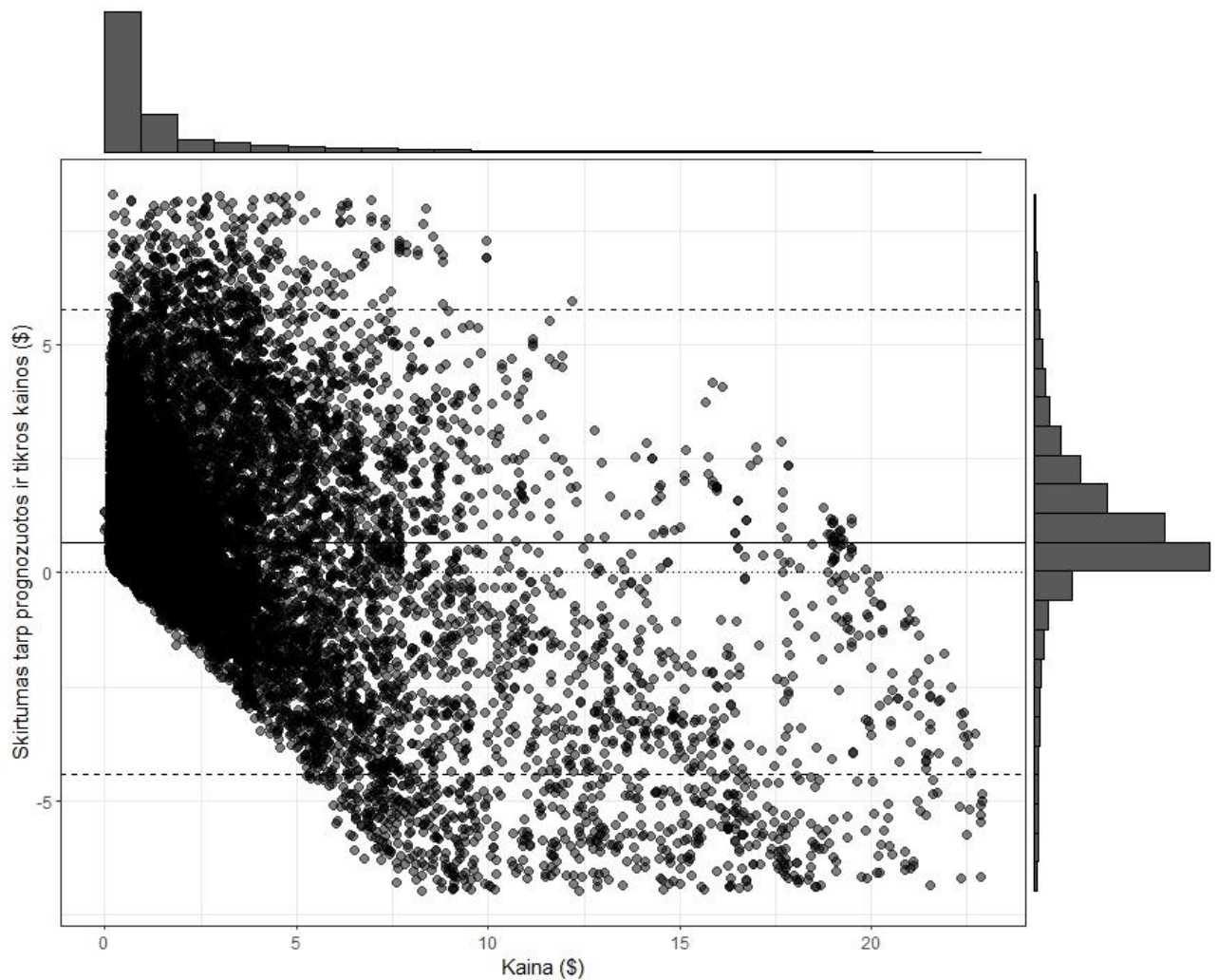


18 pav. Regresijos modelis (ETH) su logaritmine transformacija, testavimo imties realios ir prognozuotos reikšmės

3.6. *Bland Altman* grafikas su paraštinėmis histogramomis

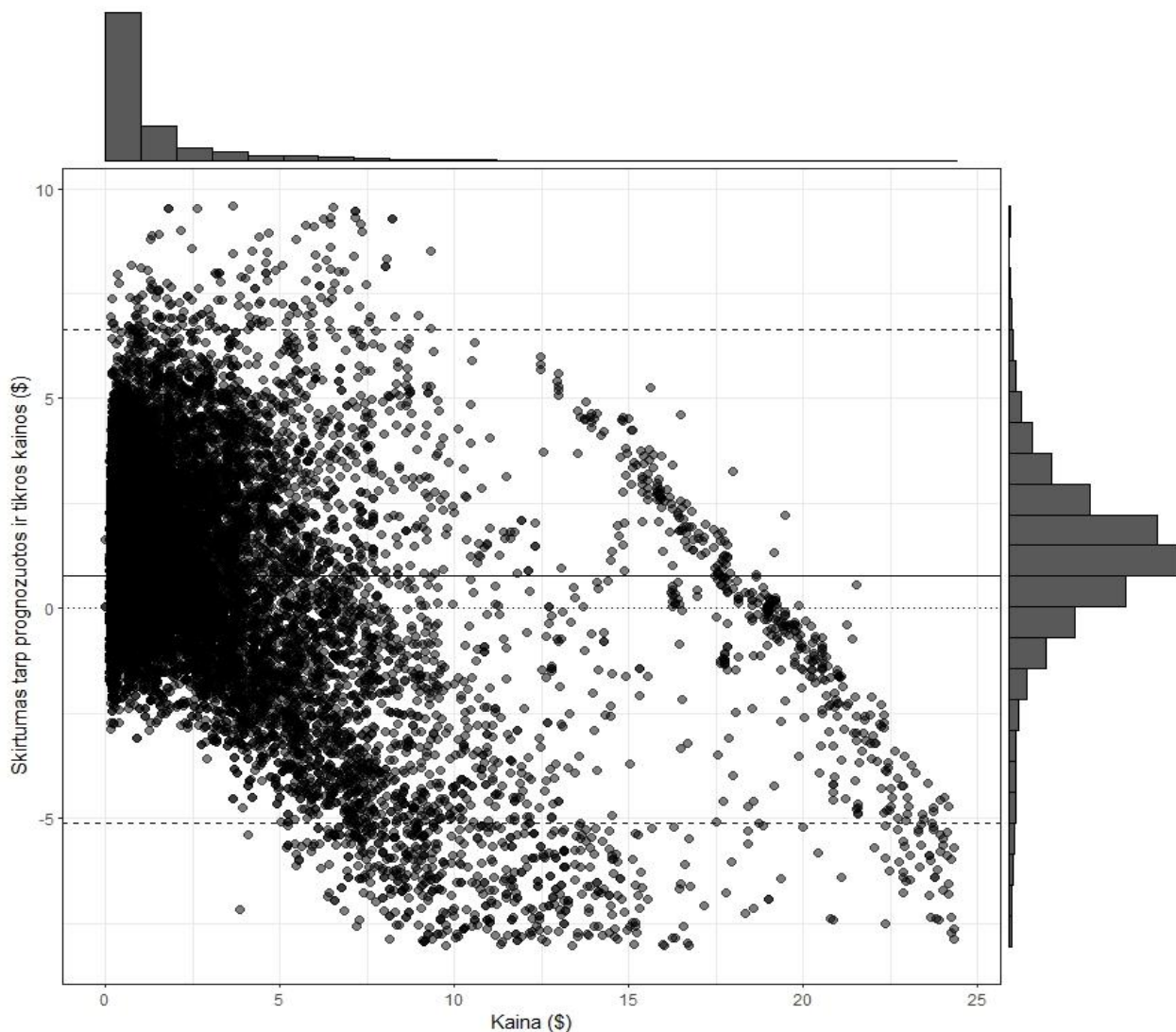
Toliau nagrinėsime atsitiktinio miško (USD) rezultatus grafiškai, pasitelkdami kitą būdą.

Žemiau pavaizduotas *Bland Altman* grafikas su paraštinėmis histogramomis (**19 pav.** Skirtumas tarp prognozuojamos ir realios kainos kartu su paraštinėmis histogramomis – atsitiktinis miškas (USD)). Ašyje x pažymėjome katinėlio kainą doleriais, o y ašyje – skirtumą tarp prognozuojamos ir realios reikšmės. Iš grafiko galime teigti, kad didelės dalies prognozių ir realių reikšmių skirtumas buvo aplink 0. Įdomus dalykas, kad prognozuota reikšmė dažniau buvo didesnė nei reali vertė, negu prognozuota reikšmė mažesnė nei reali vertė. Tai galėjo nutikti dėl to, kad didžioji pirkimų yra nedidelės vertės – iki 2 \$. Prognozės paklaida nebuvo didesnė nei 9 \$.



19 pav. Skirtumas tarp prognozuojamos ir realios kainos kartu su paraštinėmis histogramomis – atsitiktinis miškas (USD)

Tačiau verta palyginti šį grafiką su analogišku grafiku regresijos modelio doleriais (**20 pav.** Skirtumas tarp prognozuojamos ir realios kainos kartu su paraštinėmis histogramomis – reguliarizuota regresija (USD)). Grafikai yra gana panašūs. Čia taip pat prognozuota reikšmė dažniau buvo didesnė nei reali vertė, negu prognozuota reikšmė mažesnė nei reali vertė. Tačiau šia matomas išryškėjęs didesnių verčių kainų trendas. Kačiuko kainą, kai ji yra apie 18\$, modelis gana tiksliai prognozuoja. Tačiau reikšmes tarp 15 ir 18\$ pervertina, o nuo 18 iki 25\$ nuvertina. Panašu, kad visos šios reikšmės buvo įvertintos modelio labai panašiai.



20 pav. Skirtumas tarp prognozuojamos ir realios kainos kartu su paraštinėmis histogramomis – reguliarizuota regresija (USD)

3.7. Vektorizuoto tekstinio lauko įtaka paklaidoms

Panagrinėkime plačiau geriausią modelį – atsitiktinis miškas, kaina doleriais. Pažvelgsime, kokie papildomas žingsnis gali jį kažkiek pagerinti. Pirmiausiai, pažiūrėkime, kaip pasikeičia modelio rezultatai, kai įtraukiamas vektorizuotas tekstinis laukas su kačiuko biografija. Gauti tokie patikrinimo imties rezultatai:

5 lentelė. Tekstinio lauko vektorizavimo rezultatai

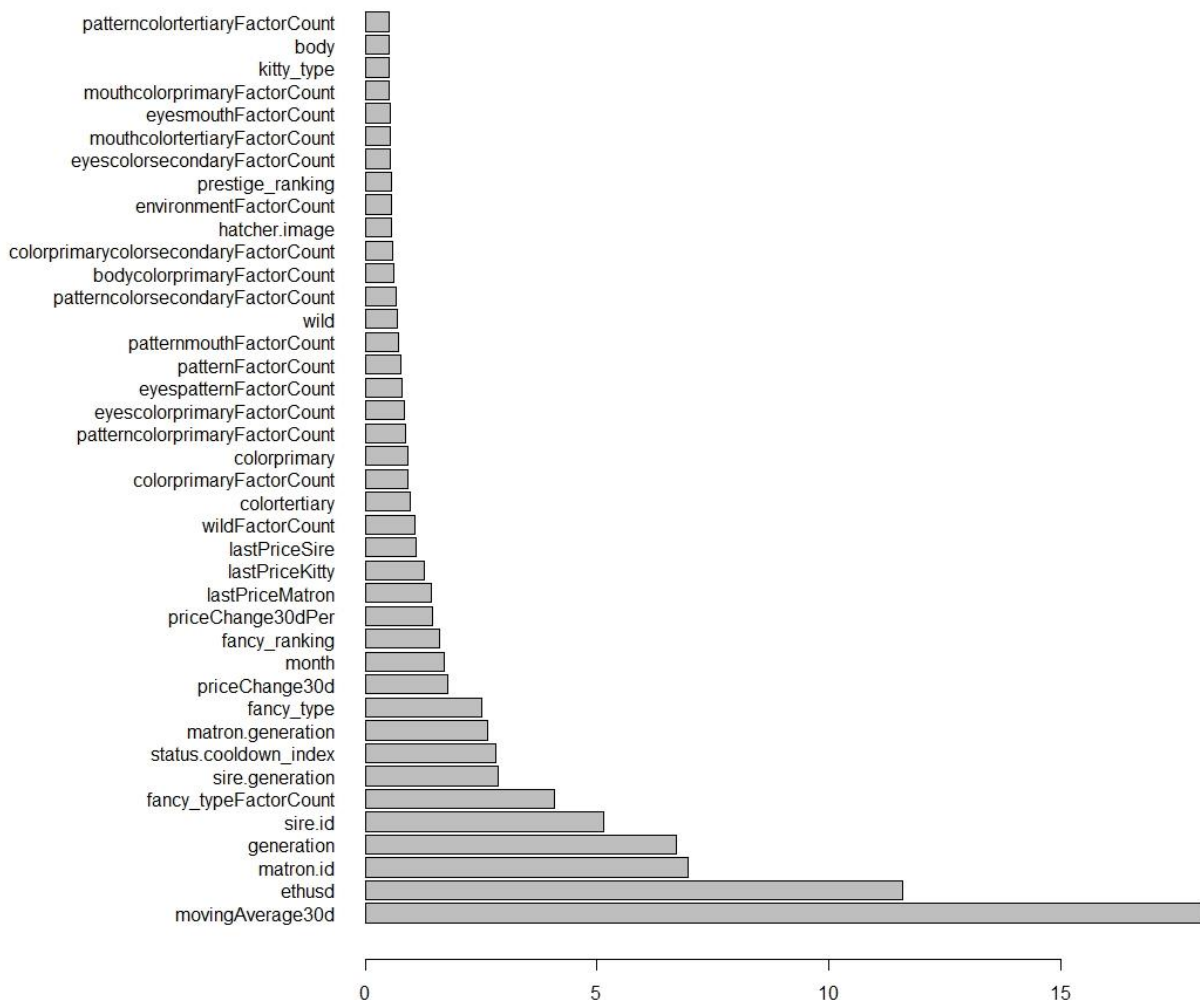
	R ²	RMSE	MdAE	MAE	MAAPE	SMAPE
Atsitiktinis miškas (USD)	0,5085	2,6373	1,0152	1,6838	82,1711	81,9741
Atsitiktiniai miškas (USD) su tekstinio lauko vektorizavimu	0,5169	2,6147	0,9941	1,6567	81,2514	81,1403

Matome, kad rezultatai pagerėjo, tačiau labai nedaug. R² padidėjo šiek tiek mažiau nei procentu – nuo 50,85 % iki 51,69 %. MAAPE ir SMAPE sumažėjo po mažiau nei 1%. Nors ir matome, kad

rezultatai šiek tiek pagerinti, tačiau galime teigti, kad tekstinis laukas nedaro labai didelės įtakos katinėlio kainai.

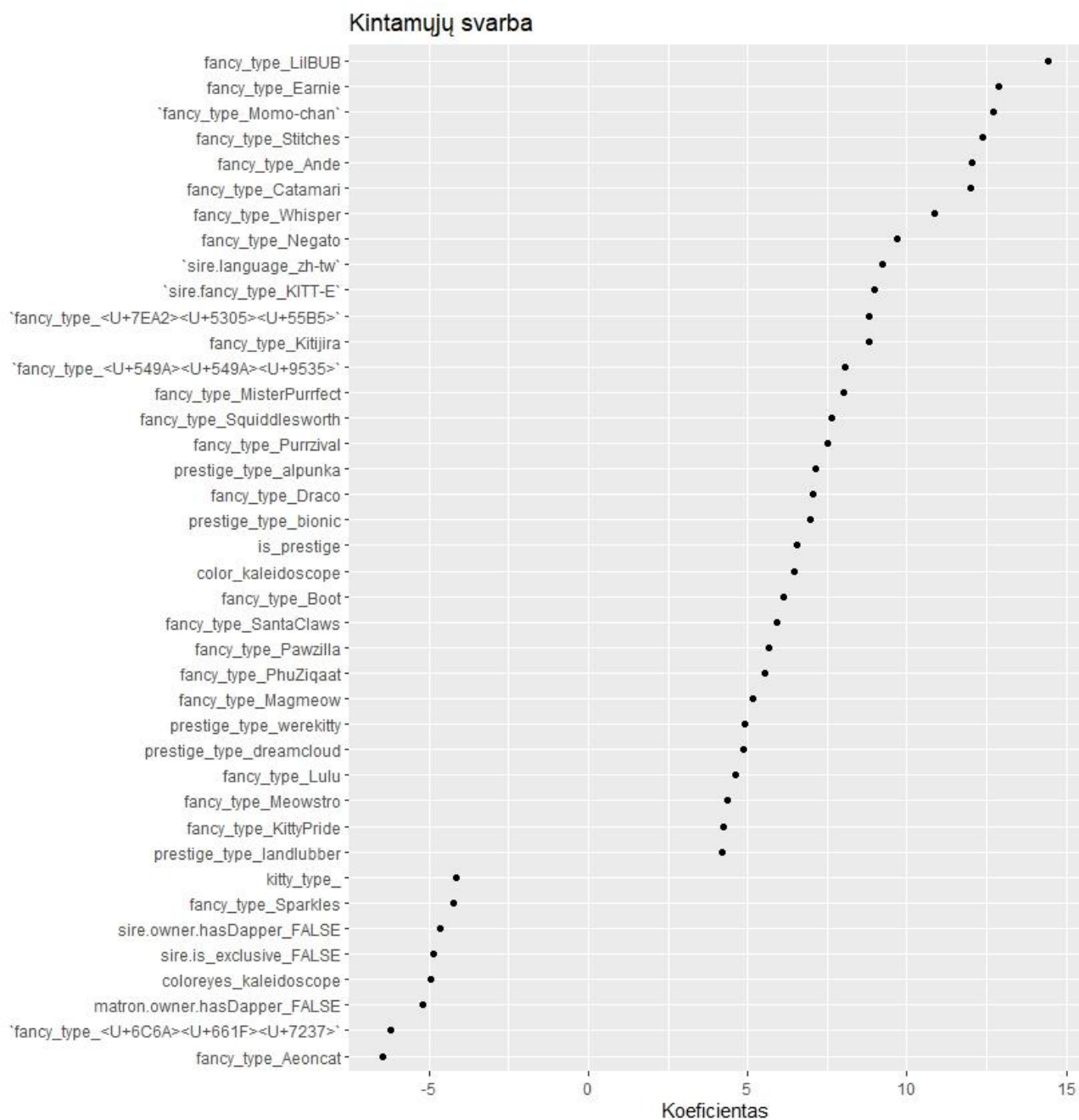
3.8. Kintamųjų svarba modelyje

Toliau pažvelkime, kurie kintamieji turėjo daugiausiai įtakos atsitiktinio miško doleriais modeliui. Grafike yra pateikta 40 kintamųjų, turėjusių didžiausią svarbą pagal keitimais grįstos kintamųjų svarbos nustatymo metodą. Matome, kad labiausiai išsiskyrė ir didžiausią svarbą turėjo kintamasis, kuris parodo eterio ir dolerio valiutos slenkamąjį 30-ies dienų vidurkį. Kiti svarbūs kintamieji: eterio ir dolerio valiutų kurso santykis pirkimo dieną, tėvų identifikaciniai numeriai, katinėlio karta, katinėlio prabangumo tipo dažnis bei tėvų identifikaciniai numeriai. Matome, kad pasiteisino sukurti naujus kintamuosius, kurie parodo tam tikrų savybių bei savybių porų dažnį. Didelė dalis svarbiausių kintamųjų yra būtent šio tipo. Galime spėti, kad pirkėjai dažnai moka didesnes sumas kačiukams, kurių savybių dažnis yra mažesnis, tai yra retesniems kačiukams.



21 pav. Kintamųjų svarba – atsitiktinis miškas (USD)

Pažvelkime, ar skirsis pagrindiniai svarbiausi kintamieji reguliarizuotos regresijos atveju. Kadangi kategoriniai regresijos kintamieji buvo išskaidyti kintamaisiais, šiame modelyje turime žymiai daugiau kintamųjų. Matome, kad čia daugybė didžiausią svarbą turinčių kintamųjų yra susiję su prabangumo tipu. Gali pasirodyti, kad prabangumo tipui modelis skiria per daug svarbos, tačiau vienas katinėlis gali turėti daugiausiai vieną prabangumo tipą, o dažniausiai jo iš viso neturi. Keista, kad tarp svarbiausių keturiasdešimties kintamųjų nepateko ir eterio – dolerio valiutų kursas bei išvestiniai su tuo susiję kintamieji. Tačiau šiame modelyje yra netoli 800 kintamųjų ir tai, kad nepateko tarp svarbiausių, dar nereiškia, kad modelyje neturėjo didelės svarbos. Pavyzdžiui, eterio ir dolerio kursas šiame modelyje pagal kintamųjų svarbą yra 76-as.



22 pav. Kintamųjų svarba – reguliarizuota regresija (USD)

3.9. Prekybos strategija

Kadangi modelio prognozavimas nėra pakankamai tikslus, nebūtų korektiška jį naudoti prekybos strategijoje kaip pagrindinį įrankį. Galima jį panaudoti nebent orientaciniais tikslais, tačiau ir tam jis gali būti nepakankamai tikslus.

Sukonstravus tikslesnį modelį, kuris prognozuoja katinėlio kainą, galima būtų pasitelkti tam tikras investavimo strategijas, kurios leistų investuotojui užsidirbti. Pats investuotojas, pamatęs parduodamą katinėlį ir jo savybes, aprašymus ir kitus duomenis gali pasimesti, nes duomenų yra išties nemažai. Kiekvienas kačiukas yra individualus, o tai reiškia, kad kiekvieną kačiuką reiktų vertinti iš naujo. Todėl pateiksiu kelias galimas strategijas, kuriomis galima būtų pasinaudoti, norint užsidirbti.

Pirmoji ir turbūt pati akivaizdžiausia – prognozavus parduodamų katinėlių kainą spręsti, ar juos įsigyti. Jei prognozuojama vertė didesnė už realią kainą, vertėtų kačiuką įsigyti. Jei prognozuojama vertė mažesnė už realią kainą, katinėlio pirkti nereikėtų. Jei prognozuojama kaina lygi realiai kainai, investuoti į tokį katinėlį taip pat greičiausiai nevertėtų, nes pardavus katinėlį būtų mažai tikėtina užsidirbti.

Dar viena strategija – ieškoti katinėlių, kurių kaina yra kintanti laike (katinėlio pardavėjas gali pasirinkti, ar nustatyti fiksuotą kainą, ar kintančią laike tarp tam tikrų rėžių). Modeliui prognozavus kačiuko kainą, tą kainą reikėtų palyginti su realios kačiuko kainos rėžiais laike. Jei prognozuota kaina neįeina į rėžius, kačiuko pirkti nevertėtų. Jei prognozuota kaina yra rėžiuose, reikėtų palaukti kol kaina nukris žemiau prognozuotos kainos ir tuomet pirkti kačiuką (žinoma, jei per tą laiką kačiuko nenuperka kitas pirkėjas).

Pasirinkus strategiją, taip pat būtina atsižvelgti į eterio dujų kainas konkrečiu pirkimo momentu, nes ši kaina turi būti įtraukta į išlaidas. Kaina priklauso ne tik nuo laiko momento, bet ir nuo to, kaip greitai norėsime, kad transakcija būtų įvykdyta. Mokant nedidelę kainą už eterio dujų kainas, norimas pirkti objektas gali būti per tą laiką nupirktas kito asmens.

Tačiau reikėtų dar kartą pabrėžti, kad gautų modelių rezultatai nėra pakankamai tikslūs, todėl paminėjome tik svarbiausius aspektus, į ką reikėtų atsižvelgti, jei turėtume tiksliau prognozuojantį modelį. Kadangi šio modelio pritaikymas prekybos strategijoje nebūtų korektiškas, prekybos strategijų toliau nebenagrinėsime.

3.10. Tyrimų ribotumas ir tyrimai ateityje

Pašalinus išskirtis, modeliai išmokyti prognozuoti tik sąlyginai nedideles kainas. Prognozuojant vertingų kačiukų kainą, tikėtina, kad gausime labai smarkiai nukrypusius nuo realybės rezultatus. Todėl investavimo strategijas modeliu grįsti aukštos vertės kačiukais nerekomenduotina. Tačiau apsibrėžėme pradžioje, kad didžioji dalis paskutinių pardavimų yra neaukštos kainos ir, todėl modelis yra orientuotas į asmenis, neinvestuojančius didelių sumų į katinėlius.

Jei katinėlio kaina būtų lengviau įvertinama, galėtume formuoti atskirus modelius kainų klasteriams, pavyzdžiui, didelės ir mažos kainos. Tačiau čia susidurtume su problema, kad nebūtų aišku, su kuriuo modeliu reikėtų naudoti naujus duomenis - katinėlius, kurių kainos rėžių net nenutuokiame. Taip pat reikėtų atsižvelgti į tai, kad nėra daug didelės kainos transakcijų duomenų, todėl, net jei sukurtume modelius jie gali būti mažiau tikslūs. Bet kokių atveju, kainos

prognozavimo modelius pagal kainos klasterį vertėtų naudoti tik tuo atveju, jei mūsų sukurtas modelis būtų tikslesnis. Jei, kaip šiuo atveju, modeliui sunku yra atspėti katinėlio kainą, tai būtų taip pat sunku atspėti ir kainos klasterį.

Taip pat vertėtų aptarti tai, kas galėjo lemti netikslius modelio rezultatus. Viena iš priežasčių gali būti ta, kad modeliuose galbūt nėra iki galo atsižvelgta į tai, kokią estetinę reikšmę pirkėjui turi kačiukas. Nors ir dalinai įtraukėme kačiukų išvaizdos ypatybes per atributus, per juos negalime įvertinti, kiek gražus katinėlis yra pirkėjui. Šis dalykas yra subjektyvus ir, todėl, sunkiai įtraukiamas į modelį per kintamuosius. Tai, kad modelio prognozės nėra labai tikslios, gali būti nulemta to, kad pirkėjui vienas iš svarbiausių pirkimo kriterijų yra kačiuko išvaizda. Tačiau, atlikus šį tyrimą, negalime daryti išvadų, ar taip iš tikrųjų yra. Tam reikėtų tolimesnių tyrimų. Ateityje galima būtų surinkti dar vieną duomenų šaltinį – katinėlių paveikslėlius ir panaudoti konvoliucinio formato paveikslėlius sudarant modelį, pavyzdžiui, dirbtinių neuroninių tinklų pagalba.

Išvados

1. Buvo aptarta, kas gali skatinti žmones investuoti į kriptokatinėlius. Tai gali lemti tos pačios psichologinės priežastys, kaip ir kolekcionavimo proceso: dėl aistros, investicijų, patiriamo malonumo, socialinių aspektų. Sužinojome, kad kriptokatinėlių platforma yra naudojama tiek dėl smagumo, tiek norint užsidirbti. Kriptokatinėliams taip pat vertės prideda ir juos pirkti skatina jų retumas, naudingumas ir išvaizda. Tikėtina, kad norą investuoti taip pat lemia tokia išorės kintamieji, kaip eterio kriptovaliutos kursas.
2. Buvo sukurti dviejų skirtingų algoritmų kriptokatinėlių kainos prognozavimo modeliai: atsitiktinio miško ir reguliarizuotos regresijos. Pastebėta, kad atsitiktinio miško algoritmas geriau įvertina kriptokatinėlio kainą negu reguliarizuotos regresijos algoritmas.
3. Empiriškai ištirta, kad abu mašininio mokymo algoritmai geriau kainą prognozavo JAV doleriais, o eterio kaina prognozavimo tikslumas mažesnis.
4. Panaudojus tekstinę lauką – kačiuko biografiją ir šį kintamąjį vektorizavus buvo nustatyta, kad geriausią modelį įmanoma dar pagerinti, tačiau labai mažai. R^2 koeficientas padidėjo šiek tiek mažiau nei procentu – nuo 50,85 % iki 51,69 %. MAAPE ir SMAPE sumažėjo po mažiau nei 1%.
5. Nustatyta, kad daugiausiai įtakos atsitiktinio miško modeliui turėjo kintamieji, susiję su eterio ir dolerio valiutų kursu, atributų bei jų porų dažniai ir kiti tokie kintamieji, kaip karta ir atvėsimo lygmuo. Kadangi reguliarizuotos regresijos modelyje kategoriniai kintamieji buvo išskirstyti į atskirus kintamuosius, o pačių kintamųjų taip buvo sukonstruota žymiai daugiau nei atsitiktinių miškų atveju, buvo sunku palyginti, ar regresijos atveju didžiausią svarbą turėjo tie patys kintamieji. Tačiau buvo pastebėta, kad didelę svarbą turėjo katinėlio prabangumo tipas.
6. Kriptokatinėlio kaina buvo sunkiai prognozuojama, pasitelkiant turimus duomenis. Matėme, kad netgi pats geriausias modelis nebuvo labai tikslus, todėl sunku būtų jį taikyti praktikoje. Todėl kurti prekybos strategiją, paremtą šiuo modeliu, nebūtų verta.

Literatūros sąrašas

1. Alanfalcon & Poopie.cat. (2019). *kittycalc*. Nuskaityta iš [kittycalc.co: https://kittycalc.co/read/?k1=1053420&k2=461679](https://kittycalc.co/read/?k1=1053420&k2=461679)
2. Berk, R. A. (2008). Statistical Learning from a Regression Perspective.
3. Byeonghwa Park, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*.
4. *Binance Academy*. (be datos). Nuskaityta iš <https://www.binance.vision/blockchain/a-guide-to-crypto-collectibles-and-non-fungible-tokens-nfts>
5. Bitcoin. (be datos). Nuskaityta iš <https://bitcoin.org/en/faq>
6. Breiman, L. (2001). Random Forests. *Machine Learning*.
7. Carolin Strobl, A.-L. B. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*.
8. Chakure, A. (2019). *Towards Data Science*. Nuskaityta iš <https://towardsdatascience.com/random-forest-and-its-implementation-71824ced454f>
9. Cheng, E. (2017 m. 12 6 d.). *Meet CryptoKitties, the \$100,000 digital beanie babies epitomizing the cryptocurrency mania*. Nuskaityta iš CNBC: <https://www.cnbc.com/2017/12/06/meet-cryptokitties-the-new-digital-beanie-babies-selling-for-100k.html>
10. Christoph Bergmeir, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*.
11. *CoinDesk*. (2018). Nuskaityta iš <https://www.coindesk.com/ebay-cryptokitties-raises-2-million-star-vcs>
12. *Coinmarketcap*. (2020). Nuskaityta iš <https://coinmarketcap.com/>
13. *Cryptokitties Public API*. (be datos). Nuskaityta iš <https://docs.api.cryptokitties.co/?version=latest>
14. *CryptoKitties Sales*. (2020). Nuskaityta iš <https://kittysales.herokuapp.com/>
15. Dimson, E., & Spaenjers, C. (2014). Investing in Emotional Assets. *Financial Analysts Journal*, 70, 20-25.
16. Dong, F. X. (2018). Bubbly Bitcoin. *SSRN Electronic Journal*.
17. Gharegozlou, R. (2019). *Medium*. Nuskaityta iš <https://medium.com/dapperlabs/introducing-flow-a-new-blockchain-from-the-creators-of-cryptokitties-d291282732f5>
18. *GitHub*. (be datos). Nuskaityta iš <https://github.com/epfml/sent2vec>
19. Hastie, H. Z. (2005). Regularization and variable selection via the elastic net. *Royal Statistical Society*.
20. <https://www.cryptokitties.co/>. (2019). Nuskaityta iš <https://www.cryptokitties.co/>
21. *Yahoo Finance*. (2020). Nuskaityta iš <https://finance.yahoo.com/quote/ETH-USD/>
22. Jaehwan Lee, B. Y. (2018). Is a Blockchain-Based Game a Game for Fun, or Is It a Tool for Speculation? An Empirical Analysis of Player Behavior in Crypokitties. *The Ecosystem of e-Business: Technologies, Stakeholders, and Connections*.

23. Jung, Y., & Pawlowski, S. D. (2014). Understanding consumption in social virtual worlds: A sensemaking perspective on the consumption of virtual goods. *Journal of Business Research*, 2231-2238.
24. Kampakis, S. (2019). Is Blockchain Part of the Future of Art? *The JBBA*.
25. Lafferty, B. A., Matulich, E., & Liu, M. X. (2013). Exploring worldwide collecting consumption behaviors. *Journal of International Business and Cultural Studies*.
26. Lehdonvirta, V. (2009). Virtual item sales as a revenue model: identifying. *Electron Commer Res*, 97-113.
27. Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting* 9.
28. Marjan Ceh, M. K. (2018). Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments. *Geo-Information*.
29. Matteo Pagliardini, P. G. (2018). Unsupervised Learning of Sentence Embeddings. *NAACL*.
30. McIntosh, W. D., & Schmeichel, B. (2004). Collectors and Collecting: A Social Psychological. *Leisure Sciences: An Interdisciplinary Journal*, 85-97.
31. Mearian, L. (2019). *Computerworld*. Nuskaityta iš <https://www.computerworld.com/article/3191077/what-is-blockchain-the-complete-guide.html>
32. medium.com. (2018 m. 04 23 d.). *A primer on Kitty cooldown speeds*. Nuskaityta iš medium.com: <https://medium.com/cryptokitties/a-primer-on-kitty-cooldown-speeds-379980fe7f2b>
33. NBA. (2019). Nuskaityta iš <https://www.nba.com/kings/blog/sacramento-kings-partner-cryptokaiju-launch-first-physical-crypto-collectible-professional>
34. *OpenSea*. (2020). Nuskaityta iš <https://opensea.io/blog/guides/welcome-to-opensea/>
35. *ResponseToFile-Postman*. (be datos). Nuskaityta iš <https://github.com/sivcan/ResponseToFile-Postman>
36. Sungil Kim, H. K. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*.
37. *The many ways to value a kitty*. (2018). Nuskaityta iš medium: <https://medium.com/cryptokitties/the-many-ways-to-value-a-kitty-e286f9e061d1>
38. V.H. Masías, M. V. (2016). Property Valuation using Machine Learning Algorithms: A Study in a Metropolitan-Area of Chile. *AMSE*.
39. Wolfgang Karl Hardle, C. R. (2020). Editorial: Understanding Cryptocurrencies. *Journal of Financial Econometrics*.
40. Zolfagharian, M. A., & Cortes, A. (2011). Motives For Purchasing Artwork, Collectibles And Antiques. *Journal of Business & Economics Research*.