



VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
TRANSPORTO INŽINERIJOS FAKULTETAS
MOBILIŲJŲ MAŠINŲ IR GELEŽINKELIŲ TRANSPORTO KATEDRA

Vygantas Verbilis

**KELIŲ TRANSPORTO PRIEMONIŲ SRAUTŲ PROGNOZAVIMO
METODŲ TYRIMAS**
STUDY OF THE FORECASTING METHODS FOR ROAD TRAFFIC

Baigiamasis magistro darbas

Transporto inžinerijos studijų programa, valstybinis kodas **6211EX058**

Transporto inžinerijos vadybos specializacija

Transporto inžinerijos studijų kryptis

Vilnius, 2020

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
TRANSPORTO INŽINERIJOS FAKULTETAS
MOBILIŲJŲ MAŠINŲ IR GELEŽINKELIŲ TRANSPORTO KATEDRA

TVIRTINU
Katedros vedantis

prof. Marijonas Bogdevičius
(Vardas, pavardė)

(Data)

Vygantas Verbilis

KELIŲ TRANSPORTO PRIEMONIŲ SRAUTŲ PROGNOZAVIMO METODŲ
TYRIMAS
STUDY OF THE FORECASTING METHODS FOR ROAD TRAFFIC

Baigiamasis magistro darbas

Transporto inžinerijos studijų programa, valstybinis kodas 6211EX058

Transporto inžinerijos vadybos specializacija

Transporto inžinerijos studijų kryptis

Vadovas

dr. Rimantas Janavičius
(Moksl. laipsnis/pedag. vardas, vardas, pavardė)

(Parašas)

(Data)

Konsultantas

(Moksl. laipsnis/pedag. vardas, vardas, pavardė)

(Parašas)

(Data)

Konsultantas

(Moksl. laipsnis/pedag. vardas, vardas, pavardė)

(Parašas)

(Data)

Lietuvių kalbos konsultantas

(Moksl. laipsnis/pedag. vardas, vardas, pavardė)

(Parašas)

(Data)

Vilnius, 2020

*magistro BD projektas ir jo turinys
pateiktas. Darbas patvirtintas Lietuvos kelių
konsultacijoje. Anelikej Petraitienė, 2020*

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
TRANSPORTO INŽINERIJOS FAKULTETAS
MOBILIŲJŲ MAŠINŲ IR GELEŽINKELIŲ TRANSPORTO KATEDRA

Transporto inžinerijos studijų kryptis
Transporto inžinerija studijų programa, valstybinis kodas
6211EX058
Transporto inžinerijos vadybos specializacija

TVIRTINU

Katedros vedėjas

(Parašas)

Marijonas BOGDEVICIUS

(Vardas, pavardė)

(Data)

BAIGIAMOJO MAGISTRO DARBO
UŽDUOTIS

.....Nr.

Vilnius

Studentui (ei) Vygantui Verbiliui

Baigiamojo darbo tema: Miesto transporto valdymo sistemos su judančiais stacionariais jutikliais tyrimas

patvirtinta 2018 m. lapkričio 12 d. dekanato potvarkiu Nr. 463ti.

Baigiamojo darbo užbaigimo terminas 2020 m. gegužės 25 d.

BAIGIAMOJO DARBO UŽDUOTIS:

Atlikti mokslinių darbų apžvalgą baigiamojo darbo tema. Pagrindu pasirinkti SARIMA modelių grupę ir ištirti jos galimybes eismo srautų parametrų prognozuoti. Modelio mokymui naudoti Vilniaus miesto vienos sankryžos eismo srautų duomenis. Sudaryti modelio svarbiausių kriterijų sąrašą, parinkti rezultatų vertinimo kriterijus ir palyginti gautus rezultatus su eksperimentiniais rezultatais. Pateikti tyrime naudotų modelių rekomendacijas eismo srautams prognozuoti.

Baigiamojo darbo rengimo konsultantai:

.....
.....
.....
(Moksl. laipsnis/pedag. vardas, vardas, pavardė)

Vadovas

(Parašas)

dr. Ramūnas Jurevičius
(Moksl. laipsnis/pedag. vardas, vardas, pavardė)

Užduotį gavau

(Parašas)

Vygantas Verbilis

(Vardas, pavardė)

2018-11-15

(Data)

Vilniaus Gedimino technikos universitetas Transporto inžinerijos fakultetas Mobiliųjų mašinų ir geležinkelių transporto katedra	ISBN Egz. sk. Data-.....-.....
Antrosios pakopos studijų Transporto inžinerijos programos magistro baigiamasis darbas	
Pavadinimas	Kelių transporto priemonių srautų prognozavimo metodų tyrimas
Autorius	Vygantas Verbilis
Vadovas	Raimundas Junevičius
Kalba: lietuvių	
Anotacija Baigiamajame magistro darbe nagrinėjami kelių transporto priemonių eismo srautų prognozavimo metodai. Analitinėje dalyje apibūdinta eismo valdymo specifika, apžvelgti populiariausi naivieji, parametriniai ir neparimetriniai prognozės modeliai. Atlikta mokslinės literatūros analizė. Tiriamojame dalyje nagrinėtas parametrinių bei neparimetrinių modelių taikymas prognozuojant trumpalaikius eismo srautus. Įvertinti reikalavimai pradiniam duomenims. Sudarytos penkiolikos minučių eismo intensyvumo prognozės. Modeliai validuoti prognozes palyginus su realiais duomenimis. Įvertintas sudarytų modelių tikslumas, greitis, geba vertinti skirtingų eismo juostų intensyvumą. Įvertintas modelių tikslumas atliekant skirtingo laikotarpio prognozes. Išbandytas glodinimo metodų taikymas siekiant sumažinti pradinio duomenų triukšmą. Pateiktos rekomendacijos galimam prognozės metodų taikymui. Darbą sudaro 6 dalys: įvadas, analitinė dalis, prognozės modelių tyrimo metodika, eismo srautų prognozės tyrimas, išvados ir siūlymai, literatūros sąrašas. Darbo apimtis – 104 p. teksto be priedų, 44 iliustr., 19 lent., 31 bibliografiniai šaltiniai. Atskirai pridedami darbo priedai.	
Prasminiai žodžiai: Eismo intensyvumas, eismo srautų prognozavimas, eismo srautas, eismo valdymas, prognozės modelis, SARIMA modelis, šviesoforų valdymas, trumpalaikė prognozė.	

Vilnius Gediminas Technical University Faculty of Transport Engineering Department of Mobile Machinery and Railway Transport	ISBN ISSN Copies No. Date-.....-.....
--	---

Master Degree Studies Transport Engineering study programme Master Graduation Thesis	
Title	Study of the forecasting methods for road traffic
Author	Vygantas Verbilis
Academic supervisor	Raimundas Junevičius

Thesis language: Lithuanian

Annotation The master thesis presents study of the forecasting methods for road traffic. Concept of the short-term forecasting methods application for traffic light controlled intersections cycle optimization is addressed in the analytical part. Analytical part explores naive, parametric and nonparametric forecasting methods and identifies these models ability to forecast with real time data. Based on the literature overview, SARIMA model is chosen for further investigation. In the methodology part of this paper, definition of SARIMA method and model construction is described. SARIMA model application with real data is presented in the research part of this paper. Forecast of fifteen minutes traffic flow accuracy is evaluated. For comparison purpose, SARIMA model forecasts are compared with LSTM models. Models ability to forecast different time horizon and application versatility is evaluated. Smoothing methods were used in order to address noisy primary data of the road traffic. As a result, accurate forecast characteristic are described and recommendations for further investigation and application are proposed. Thesis consist of the following parts: introduction, analytical review of the forecasting methods and intersection control, research methodology, short-term forecasting methods research, conclusion and proposals, list of references. Thesis consists of: 104 pages of text without appendixes, 44 illustrations, 19 tables, 31 bibliographical entries. Appendixes included.

Keywords: Forecasting, LSTM, road traffic forecasting, SARIMA, signalized intersections, short-term forecast, traffic flow, traffic prediction.
--

Vilniaus Gedimino technikos universiteto
egzaminų sesijų ir baigiamųjų darbų
rengimo bei gynimo organizavimo tvarkos
aprašo
2 priedas

(Baigiamojo darbo sąžiningumo deklaracijos forma)

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Vygantas Verbilis, 20144889

(Studento vardas ir pavardė, studento pažymėjimo Nr.)

Transporto inžinerijos fakultetas

(Fakultetas)

Transporto inžinerija, TIVfm-18

(Studijų programa, akademinė grupė)

BAIGIAMOJO DARBO (PROJEKTO)

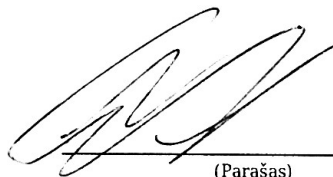
SĄŽININGUMO DEKLARACIJA

2020 m. gegužės 24 d.

Patvirtinu, kad mano baigiamasis darbas tema „Kelių transporto priemonių srautų prognozavimo metodų tyrimas“ patvirtintas 2020 m. balandžio 27 d. dekanų potvarkiu Nr. 139ti, yra savarankiškai parašytas. Šiame darbe pateikta medžiaga nėra plagijuota. Tiesiogiai ar netiesiogiai panaudotos kitų šaltinių citatos pažymėtos literatūros nuorodose.

Mano darbo vadovas docentas daktaras Raimundas Junevičius.

Kitų asmenų indėlio į parengtą baigiamąjį darbą nėra. Jokių įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs (-usi).



(Parašas)

Vygantas Verbilis
(Vardas ir pavardė)

TURINYS

IVADAS.....	14
1. ANALITINĖ DALIS	16
1.1 <i>Eismo srautų specifika ir valdymas</i>	16
1.1.1 Eismo srauto apibūdinimas	16
1.1.2 Eismo valdymas	20
1.2 <i>Prognozės metodų apžvalga ir klasifikavimas.....</i>	22
1.2.1 Prognozės metodų taikymas eismo valdymo sistemoje	22
1.2.2 Naivūs prognozės metodai	26
1.2.3 Parametriniai prognozės metodai	26
1.2.4 Neparametriniai prognozės metodai.....	29
1.2.5 Hibridiniai modeliai	33
1.3 <i>Mokslo darbų apžvalga.....</i>	34
2. PROGNOZĖS MODELIŲ TYRIMO METODIKA	39
2.1 <i>SARIMA modelis</i>	41
2.1.1 ARIMA ir SARIMA modelio matematinis apibrėžimas.....	41
2.1.2 Duomenų parengimas.....	44
2.1.3 Modelio struktūros nustatymas	47
2.1.4 Modelio įvertinimas	48
2.1.5 Modelio diagnostika.....	48
2.1.6 Modelio validacija.....	49
2.2 <i>LSTM modelis</i>	50
2.3 <i>Glodinimo metodai</i>	54
2.3.1 Gauso filtras	54
2.3.2 Slankiosios medianos metodas.....	55
3. EISMO SRAUTŲ PROGNOZĖS TYRIMAS	56
3.1 <i>Prognozės modelio sudarymas</i>	56
3.1.1 Pradiniai duomenys	56
3.1.2 Duomenų stacionarumas	60
3.1.3 Modelio identifikacija	65
3.1.4 Modelio įvertinimas ir diagnostika	66
3.1.5 Modelio validacija.....	69
3.1.6 Dalinės išvados.....	73
3.2 <i>Prognozės modelio universalumo įvertinimas</i>	74
3.2.1 Grafinė analizė	75
3.2.2 Prognozės rezultatai	77
3.2.3 SARIMA modelio įvertinimas su kito jutiklio duomenimis	80
3.2.4 Dalinės išvados.....	82
3.3 <i>Prognozuojamo laikotarpio įvertinimas</i>	83
3.3.1 Prognozės rezultatai	84
3.3.2 Dalinės išvados.....	87

3.4 Modeliui sudaryti reikalingo duomenų kiekio įvertinimas	88
3.4.1 Prognozės rezultatai	89
3.4.2 Dalinės išvados.....	92
3.5 Duomenų glodinimo įvertinimas.....	93
3.5.1 Prognozės rezultatai glodintiems duomenims.....	94
3.5.2 Glodinimo metodų taikymas, prognozuojant trumpesnio laikotarpio intensyvumą	96
3.5.3 Dalinės išvados.....	98
IŠVADOS	99
LITERATŪRA	102
PRIEDAI	105

Paveikslų sąrašas

1.1 pav. Laukiančiųjų eilės formavimas ir gretimos sankryžos blokavimas (Yan <i>et al.</i> 2013)	17
1.2 pav. Fundamentaliosios eismo srautų diagramos (Zaidi <i>et al.</i> 2015)	19
1.3 pav. Prognozės modelio integracija į eismo valdymo sistemą (Zheng, Recker 2013).....	23
1.4 pav. Prognozės modelių klasifikavimas (Hinsbergen <i>et al.</i> 2007).....	25
1.5 pav. Bajeso tinklo grafinis pavyzdys (Sun <i>et al.</i> 2006)	30
1.6 pav. Neuronų ryšiai (Zhang <i>et al.</i> 2016).....	31
1.7 pav. RNN tinklo struktūra (Yuan <i>et al.</i> 2019)	32
2.1 pav. ARIMA modelio sudarymo procedūra (Dong <i>et al.</i> 2009).....	44
2.2 pav. LSTM modelio ląstelės struktūra (Yuan <i>et al.</i> 2019).....	51
2.3 pav. Sudaryto LSTM modelio struktūra	53
3.1 pav. Justiniškių–Virbališkių g. orofoto (<i>Vilniaus miesto vairuotojų informacinė sistema</i> 2020).....	57
3.2 pav. Pradiniai Justiniškių g. (link Laisvės pr.) pirmos eismo juostos jutiklio VD2.1 duomenys	58
3.3 pav. Pradiniai VD2.1 jutiklio duomenys. Savaitės imtis (2018-09-17 – 2018-09-24).....	59
3.4 pav. Diferencijuotos laiko eilutės duomenys.....	59
3.5 pav. Pradinių duomenų autokoreliacijos funkcija.....	61
3.6 pav. Pradinių duomenų dalinės autokoreliacijos funkcija	61
3.7 pav. Sezoniskai diferencijuota laiko eilutė	62
3.8 pav. Sezoniskai diferencijuotų duomenų autokoreliacijos funkcija.....	63
3.9 pav. Sezoniskai diferencijuotų duomenų dalinės autokoreliacijos funkcija	63
3.10 pav. Pirmą eilę ir sezoniskai diferencijuotų duomenų autokoreliacijos funkcija.....	65
3.11 pav. Pirmą eilę ir sezoniskai diferencijuotų duomenų dalinės autokoreliacijos funkcija	65
3.12 pav. SARIMA(1,0,1)(0,1,1) modelio liekamosios paklaidos	67
3.13 pav. SARIMA(1,0,1)(0,1,1) modelio liekamųjų paklaidų histograma	67
3.14 pav. SARIMA(1,0,1)(0,1,1) ₆₇₂ modelio prognozė.....	70
3.15 pav. LSTM modelio prognozė.....	71
3.16 pav. Justiniškių g. (link Laisvės pr.) eismo juostų intensyvumas	75
3.17 pav. Justiniškių g. (nuo Laisvės pr.) eismo juostų intensyvumas	75
3.18 pav. Virbališkių g. (nuo Medėinos g.) eismo juostų intensyvumas	76
3.19 pav. Virbališkių g. (nuo Budiniškių g.) eismo juostų intensyvumas	77
3.20 pav. SARIMA(1,0,1)(0,1,1) modelio prognozė antrai Justiniškių g. (link Laisvės pr.) eismo juostai.....	79
3.21 pav. LSTM modelio prognozė prognozė antrai Justiniškių g. (link Laisvės pr.) eismo juostai (VD2.2 jutiklis)	79
3.22 pav. Sezoniskai diferencijuotų VD2.2 jutiklio duomenų kreivė.....	80
3.23 pav. VD2.2 jutiklio sezoniskai diferencijuotų duomenų autokoreliacijos funkcija	81
3.24 pav. VD2.2 jutiklio sezoniskai diferencijuotų duomenų dalinės autokoreliacijos funkcija	81
3.25 pav. SARIMA modelio liekamųjų paklaidų autokoreliacija su VD2.2 jutiklio duomenimis	82
3.26 pav. 360 s intervalu suformuota laiko eilutė.....	84
3.27 pav. SARIMA(1,0,1)(0,1,1) modelio prognozė, prognozuojant 360 s žingsniu.....	86
3.28 pav. LSTM modelio prognozė, prognozuojant 360 s žingsniu.....	86
3.29 pav. SARIMA(1,0,1)(0,1,1) prognozė, modeliui sudaryti naudojant 2 savaitių duomenis	91
3.30 pav. LSTM prognozė, modeliui sudaryti naudojant 2 savaitių duomenis	91
3.31 pav. Gauso filtru, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant SARIMA(1,0,1)(0,1,1) modelį	95
3.32 pav. Gauso filtru, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant LSTM modelį	95
3.33 pav. Slankiąja mediana, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant SARIMA(1,0,1)(0,1,1)	97
3.34 pav. Slankiąja mediana, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant LSTM modelį	98

Lentelių sąrašas

1.1 lentelė. Lewi's skalė prognozės tikslumui įvertinti (Kumar, Vanajakshi 2015)	24
2.1 lentelė. Skaičiavimams naudojama kompiuterinė įranga	40
2.2 lentelė. LSTM tinklo parametrai	53
3.1 lentelė. Justiniškių–Virbeliškių g. sankryžoje naudojami eismo srautų jutikliai	57
3.2 lentelė. Modeliui sudaryti naudojami pradiniai duomenys	58
3.3 lentelė. Pradinių duomenų ADF testo reikšmė	60
3.4 lentelė. Sezoniškai diferencijuotos laiko eilutės ADF testo rezultatai	62
3.5 lentelė. Pirma eile ir sezoniškai diferencijuotos laiko eilutės ADF testo rezultatai	64
3.6 lentelė. Tikėtinos SARIMA modelio eilės	66
3.7 lentelė. Ljung–Box liekamųjų paklaidų testo rezultatai	68
3.8 lentelė. SARIMA(1,0,1)(0,1,1) modelio regresijos koeficientų vertės	69
3.9 lentelė. Prognozės paklaidų ir skaičiavimo trukmės rezultatai.....	71
3.10 lentelė. Skirtingų eismo juostų prognozių paklaidos.....	78
3.11 lentelė. Ljung–Box liekamųjų paklaidų koreliacijos testo rezultatai (VD2.2 jutiklis).....	82
3.12 lentelė. Trumpesnio laikotarpio prognozės rezultatai.....	85
3.13 lentelė. Prognozės rezultatai, modeliui sudaryti taikant skirtingo laikotarpio duomenų imtį	90
3.14 lentelė. SARIMA(1,0,1)(0,1,1) modelio regresijos lygties koeficientai	90
3.15 lentelė. Glodintų duomenų prognozės paklaidos.....	94
3.16 lentelė. Prognozės paklaidos prognozuojant 360 s laikotarpį.....	96

Priedų sąrašas

Literatūros apžvalgos santrauka	105
Pirma eilė ir sezoniškai diferencijuota laiko eilutė.....	106
SARIMA prognozės modelių įvertinimo procedūra	107
SARIMA lyginamųjų modelių liekamosios paklaidos	109
LSTM modelio mokymo kreivė	111
Prognozių grafikai	112
Skaitinės prognozių vertės.....	113
Kitų eismo juostų prognozės	116
Glodintų duomenų prognozės.....	119

Vartojamos sąvokos

• Baltasis triukšmas	Nulinės matematinės vilties, atsitiktinio pobūdžio procesas.
• Glodinimas	Duomenų imties lyginimas, slopinant aukšto dažnio svyravimus.
• Glodinimo plotis	Laiko eilutės nariai, kurie glodinimo metu naudojami vidurkiui ar medianai skaičiuoti
• Laiko eilutė	Duomenų imtis, sudaryta iš fiksuotu dažniu atliktų matavimų.
• Mediana	Variacinės eilutės vidurinioji reikšmė
• Modelio įvertinimas	Regresinių priklausomybių duomenyse nustatymas. Parametrinių modelių atveju, nustatomos regresijos koeficientų vertės. Neparametrinių modelių atveju, atliekamas modelio mokymas
• Modelio įvertinimo trukmė	Laiko trukmė, skirta įvertinti duomenų regresines priklausomybes. Parametrinių modelių atveju, šis laikas apima regresijos koeficientų skaičiavimo trukmę. Neparametrinių modelių atveju – modelio mokymo trukmę.
• Modelio mokymas	Procesas, kurio metu neparametriniai modeliai įvertina duomenyse esančius dėsningumus
• Pasikliautinasis intervalas	Intervalas, kuriame tikėtina matuojamo dydžio reikšmė
• Prognozė	Būsimą sistemos būvį apibūdinančio kintamojo reikšmė
• Prognozės sudarymo trukmė	Laiko trukmė, per kuria modelis sudaro 1 žingsnio prognozę
• Prognozės žingsnis	Laiko intervalas tarp prognozuojamų reikšmių
• Sezonas	Periodas, apimantis dėsningus duomenų pasikartojimus
• Variacinė eilutė	Didėjančia ar mažėjančia tvarka surašytos kintamojo reikšmės
• Vėlavimas	Vėlinimas, kurį sukelia nepriklausomas kintamasis. Literatūroje dar vadinama Laiko perstūma (angl. <i>Lag</i>)

Vartojamos santrumpos

• ACF	Autokoreliacijos funkcija
• ADF	Augmented Dickey–Fuller testas
• AIC	Akaike informacijos kriterijus
• ARIMA	Autoregresinis integruotasis slankiųjų vidurkių modelis (angl. <i>Autoregressive Integrated Moving Average</i>)
• BIC	Bajeso informacijos kriterijus
• BP	Grižtamo ryšio neuroniniai tinklai (angl. <i>Back propogation neural network</i>)
• LSTM	Ilgos trumpalaikės atminties tinklai (angl. <i>Long Short Term Memory networks</i>)
• KNN	k artimiausio kaimyno metodas (angl. <i>K nearest neighbors</i>)
• MAPE	Vidutinė absoliutinė procentinė paklaida (angl. <i>Mean absolute percentage error</i>)
• MAE	Vidutinė absoliutinė paklaida (angl. <i>Mean absolute error</i>)
• NN	Neuroniniai tinklai
• PACF	Dalinės autokoreliacijos funkcija
• RMSE	Vidutinė kvadratinė paklaida (angl. <i>Root Mean Square Error</i>)
• RNN	Rekurentiniai neuroniniai tinklai (angl. <i>Recurrent neaural networks</i>)
• SARIMA	Sezoninis autoregresinis integruotasis slankiųjų vidurkių modelis (angl. <i>Seasonal Autoregressive Integrated Moving average</i>)

IVADAS

Eismo srautai pasižymi netolygiu eismo dalyvių pasiskirstymu laike. Tai sukelia poreikį skirtingai reguliuoti transporto priemonių judėjimą sankryžose. Eismo reguliavimas turi atitikti esamą intensyvumą ir užtikrinti optimalią valdymo strategiją, leidžiančią pasiekti didžiausią įmanomą eismo srautų pralaidumą tam tikromis sąlygomis. Valdant eismo srautus, susiduriama su duomenų inertiškumo problema. Reaguojant į realiuoju laiku fiksuojamus parametrus, valdymo korekcijos atliekamos pavėluotai, kadangi duomenų surinkimas, apdorojimas bei šviesoforų darbo ciklo optimizavimas užtrunka. Dėl šios priežasties dažnai eismo valdymas neatitinka realių eismo dalyvių poreikių. Šią problemą galima sumažinti vertinant būsimus eismo srauto parametrus, atliekant trumpalaikę eismo srautų prognozę. Integravus trumpalaikės prognozės modelius į eismo valdymo sistemą, galima optimizuoti šviesoforų darbo ciklą ir užtikrinti tinkamą reagavimą į besikeičiantį eismo intensyvumą.

Prognozės modelių taikymas eismo srautams tirti plačiau pradėtas nagrinėti XX a. pabaigoje, smarkai patobulėjus kompiuterinėms technologijoms, suteikiančioms galimybę atlikti sudėtingus skaičiavimus realiuoju laiku. Yra daugybė prognozės metodų, kurių taikymo galimybės eismo srautų dėsningumams nustatyti yra tiriamos. Dėl to šiame darbe siekiama atlikti prognozės metodų analizę, įvertinti jų taikymo galimybę sudarant trumpalaikę eismo srautų prognozę bei įvertinti veiksnius, svarbius tiriant eismo srautų specifiką.

Darbo tikslas:

- Įvertinti prognozės metodų taikymą atliekant trumpalaikę eismo srautų prognozę.

Uždaviniai:

- Apibūdinti eismo valdymo specifiką.
- Apžvelgti trumpalaikėms prognozėms taikomus metodus, atlikti mokslo darbų analizę.
- Sudaryti parametrinį prognozės modelį, kurio prognozės paklaida neviršytų 20% vidutinės absoliutinės procentinės paklaidos.
- Įvertinti sudaryto prognozės modelio gebą veikti su skirtingų eismo juostų duomenimis.
- Įvertinti sudaryto prognozės modelio gebą atlikti skirtingo laikotarpio prognozes.
- Nustatyti modeliui sudaryti reikalingų duomenų imties dydį.
- Įvertinti glodinimo metodų taikymą, mažinant svyravimų poveikį duomenyse.

Siekiant įgyvendinti šiuos tikslus, analitinėje darbo dalyje apžvelgiama eismo valdymo specifika ir apibūdinami spūsčių formavimosi principai. Apžvelgiami populiariausi prognozės metodai, taikomi eismo srautams prognozuoti bei atliekama mokslinių darbų apžvalga. Tiriamojoje

dalyje nagrinėjama parametrinio SARIMA modelio sudarymo metodika bei atliekamas uždaviniuose įvardytų veiksnių poveikio tyrimas. SARIMA modelio rezultatai palyginti su neparametrinio LSTM modelio prognoze.

Šio darbo unikalumas atskleidžiamas nagrinėjant prognozės modelių taikymą nedidelio intensyvumo sankryžos srautų prognozei, modelių išbandymu prognozuojant kelių eismo juostų duomenis bei bandytų glodinimo metodų taikymu. Tyrimo rezultatai apibendrinti išvadose bei sudarytos rekomendacijos tolesniam prognozės metodų tyrimui bei galimam taikymui.

1. ANALITINĖ DALIS

1.1 Eismo srautų specifika ir valdymas

1.1.1 Eismo srauto apibūdinimas

Eismo srautas – tam tikroje vietoje kryptingai judančių transporto priemonių visuma. Eismo srauto sąvoka naudojama apibendrinti atskirų eismo dalyvių elgesį ir vertinti kaip visumos, siekiant sudaryti sąlygas modeliuoti bei įvertinti vykstančius procesus. Nagrinėjant atskirų transporto priemonių judėjimą, eismas gali atrodyti chaotiškai, kadangi sunku apibrėžti tiksliai judėjimo trajektorijas. Tiriant apibendrintą eismo dalyvių judėjimą kaip srauto, pastebimi dėsningumai: eismo dalyvių kiekio priklausomybė nuo laiko, aplinkos sąlygų. Žmonės juda įprastais maršrutais į darbus ar namus ir įprastai ši judėjimo specifika mažai kinta. Ir tik įvykus specifiniams įvykiams ar eismo sistemos perpildymui, eismo dalyvių judėjimas sutrinka, sistema išsiderina, iki vėl grįžtama į normalų ritmą.

Sklandus eismo srauto judėjimas priklauso nuo esamo infrastruktūros pralaidumo, eismo dalyvių kiekio konkrečiu laiko momentu, jų elgesio, eismo valdymo signalų. Remiantis Yue *et al.* (2007), išskiriamos specifinės situacijos, sukeliančios staigius eismo srautų pokyčius bei spūstis:

- Eismo įvykiai – įvykus eismo įvykiui susidaro fizinė kliūtis kelyje, dėl kurios sumažėja vidutinis srauto judėjimo greitis, padidėja eismo dalyvių atsargumas ir chaotiškas elgesys, kelionės trukmė.
- Kelio darbai – vykstant kelio darbams yra siaurinama važiuojamoji kelio dalis, naudojami griežtesni greičio ribojimai, dėl to sumažėja eismo sistemos pralaidumas. Taip pat eismo ribojimai sukelia eismo dalyvių nepasitenkinimą, jų elgesys tampa chaotiškas.
- Oro sąlygos – staigūs oro sąlygų pokyčiai sukelia vairavimui nepalankias sąlygas, dėl to padidėja eismo dalyvių atsargumas ir kelionės trukmė, sumažėja važiavimo greitis.
- Eismo valdymo įrangos gedimai – sugedus signalinei įrangai, šviesoforams, eismo judėjimas vyksta pagal nustatytas pirmumo taisykles ir ženklus. Tokiu atveju važiuojantiems iš šalutinio kelio sunkiau kirsti kelių sankirtą, formuojasi spūstis. Netipinė situaciją sukelia papildomą stresą eismo dalyviams.
- Nekorektiškas eismo srautų valdymas – tinkamai neįvertinus eismo dalyvių judėjimo ypatybių, valdymo signalai, kuriais koordinuojamas transporto priemonių judėjimas sankryžose, gali būti neadekvatus, dėl to sumažėja sankryžos pralaidumas.

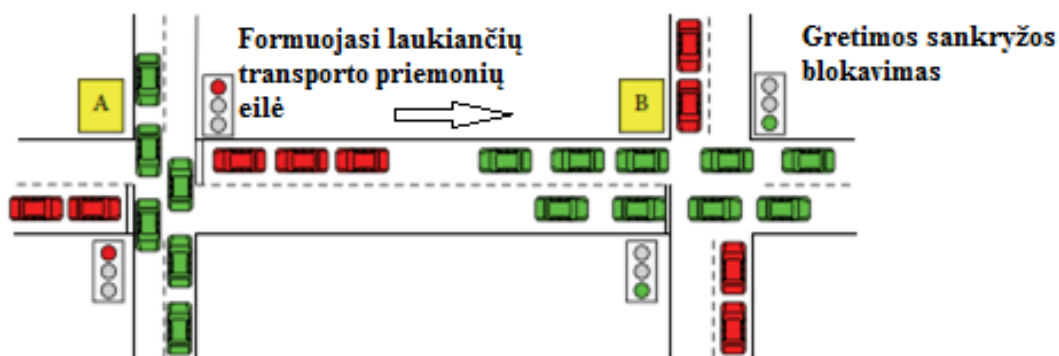
Eismo spūstis – eismo srauto būseną, pasižyminti kelionės trukmės padidėjimu dėl didesnio eisme dalyvaujančių transporto priemonių kiekio, nei infrastruktūros pralaidumas

esamomis sąlygomis (Anokye *et al.* 2013). Spūstis pradeda formuotis išaugus eismo dalyvių skaičiui arba staiga sumažėjus eismo infrastruktūros pralaidumui (įvykus eismo įvykiui ar staiga pasikeitus oro sąlygoms).

Eismo spūsties susidaryme pastebimi bendri eilių formavimosi požymiai: užsidegus raudonam šviesoforo signalui atvykstančios transporto priemonės sustoja prie sankryžos. Eilė didėja tam tikru greičiu, atvažiuojančio srauto kryptimi. Patiriama delsa dėl eismo valdymo signalų. Užsidegus žaliai šviesoforo signalui, eilė išsiformuoja, tačiau dėl srauto inertiškumo šis procesas vyksta palaipsniui. Per eismą leidžiančio signalo fazę dalis transporto priemonių pravažiuoja sankryžą. Jei ši transporto priemonių dalis didesnė nei atvykstančių per ciklą, spūstis nesusidaro, kadangi laukiančiųjų eilė išsisklaido (Anokye *et al.* 2013). Eilei smarkiai išplitus, gali būti blokuojami gretimi keliai (1.1 pav.).

Pasak Nellore, Hancke (2016) spūstys pagal susidarymo priežastis skirstomos į:

- Tipinės – dėsningo pobūdžio spūstys, susidaranti pasikartojančiu paros metu, tam tikrose vietose. Šios spūstys susidaro dėl išaugusio eismo dalyvių kiekio ir įprastai siejasi su tipinėmis kelionėmis, pavyzdžiui į darbą. Dėl dėsningo pobūdžio šių spūsčių eigą bei patiriamą laiko delną galima įvertinti ir naudojant tinkamas eismo organizavimo priemones – sumažinti.
- Netipinės – atsitiktinio pobūdžio, be aiškaus dėsningumo vykstančios spūstys. Dažniausiai jas sukelia eismo įvykiai, trumpalaikiai kelio darbai, orų permainos. Dėl atsitiktinio pobūdžio šias spūstis sunku prognozuoti, nustatyti vykstančius dėsningumus, įvertinti jų vystymosi eigą.



1.1 pav. Laukiančiųjų eilės formavimas ir gretimos sankryžos blokavimas (Yan *et al.* 2013)

Apibūdinti eismo srautus naudojama Boris Kerner (2018) trijų fazių teorija, išskirianti šias eismo būsenas:

1. Laisvas judėjimas – eismo situacija, kai eismo dalyvių skaičius nedidelis, eisme dalyvauja pavienės transporto priemonės. Šis judėjimas pasižymi eismo dalyvių galimybe rinktis norimą judėjimo greitį, patiriama minimali laiko delsa dėl eismo reguliavimo priemonių taikymo.
2. Sinchronizuotas judėjimas – eismo situacija, pasižyminti padidėjusiu transporto priemonių kiekiu, tačiau eismo srauto judėjimas vyksta sklandžiai, nėra didelės delsos ties kelių sankirtomis. Eismo intensyvumas esant šiai būsenai yra artimas kritiniui, kadangi eismo srauto tankis yra padidėjęs.
3. Plintanti spūstis – būsena, kai eisme dalyvauja didesnis transporto priemonių kiekis nei esamas infrastruktūros pralaidumas. Prie sankryžų susidaro laukiančių transporto priemonių eilės, formuojasi spūstis. Kol atvykstančių eismo dalyvių skaičius yra didesnis nei išvykstančių, spūstis plinta, susidaro poveikis ir gretimoms kelių sankirtoms. Ši būsena pasižymi mažu transporto priemonių judėjimo greičiu, sumažėjusiu intensyvumu, kadangi dėl sumažėjusio judėjimo greičio sumažėja ir transporto priemonių, pravažiavusių per nagrinėjamą ruožą, skaičius.

Eismo srautus apibūdinantys parametrai:

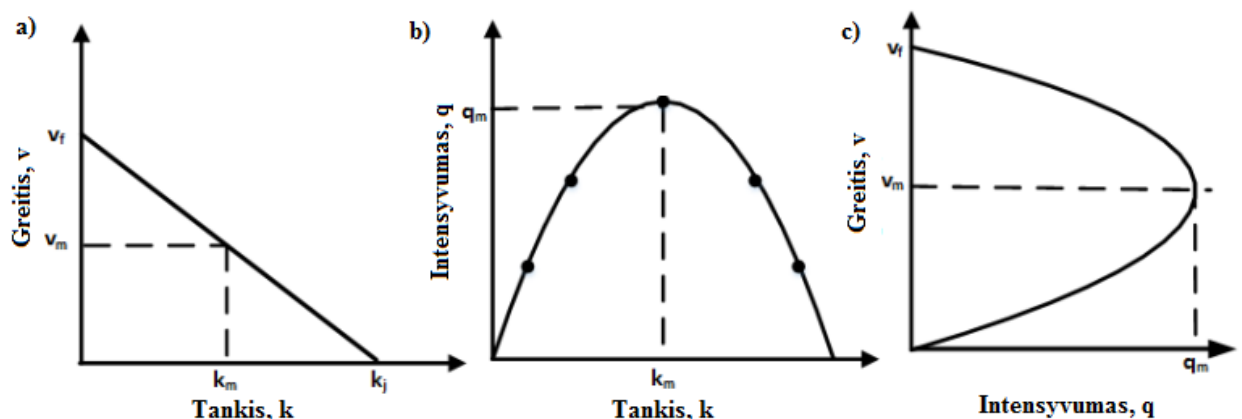
- Eismo srauto greitis – vidutinis transporto priemonių judėjimo greitis nagrinėjamame kelio ruože. Greitis nustatomas remiantis išmatuota kelionės trukme ar naudojant tiesiogines greičio matavimo priemones.
- Intensyvumas – pravažiavusių transporto priemonių skaičius per laiko vienetą.
- Eismo srauto tankis – sąlyginis transporto priemonių kiekis ilgio vienetu, vertinant fizinį transporto priemonės ilgį bei būtiną saugų atstumą tarp transporto priemonių.
- Vidutinė kelionės trukmė – vidutinė vienos transporto priemonės kelionės trukmė nagrinėjamame kelio ruože. Vertinant eismo srautus šiuo parametru nebūtina tirti visų eismo dalyvių, kadangi remiantis statistikos taisyklėmis, kelių eismo dalyvių kelionės trukmė gali apibūdinti visą populiaciją.
- Delsos laikas – vidutinis laukimo laikas ties kelių sankirta. Matuojamas tiesiogiai arba skaičiuojamas remiantis kelionės trukme.

Intensyvumas su transporto priemonių tankiu bei greičiu siejasi tarpusavyje. Jų ryšys apibūdinamas fundamentinėmis eismo srautų diagramomis (1.2 pav.). Intensyvumo vertė nustatoma transporto priemonių tankį k sudauginus su vidutiniu srauto greičiu v (1.1 formulė).

Diagramoje pastebima parabolės formos intensyvumo priklausomybė nuo tankio (1.2 pav. b). Maža intensyvumo reikšmė, esant nedideliam transporto priemonių tankiui, lemia sklandų transporto priemonių judėjimą. Parabolės ekstremumas – kritinė reikšmė, kuriai esant pravažiavusių transporto priemonių skaičius yra didžiausias. Labiau didėjant transporto priemonių skaičiui tankis išauga, tarpai tarp transporto priemonių sumažėja, judėjimo greitis krenta. Todėl vertinant eismo srautą tik pagal intensyvumo reikšmę susiduriama su problema nustatant faktinę intensyvumo vertės vietą parabolėje. Nežinoma, ar dar nepasiekta kritinė reikšmė ir eismo dalyviai juda laisvai, ar jau viršytas kritinis intensyvumas ir dėl sumažėjusio greičio pravažiuoja mažesnis automobilių skaičius.

$$q = k \cdot v \quad (1.1)$$

čia q – intensyvumas;
 k – transporto priemonių tankis;
 v – vidutinis eismo srauto greitis.



1.2 pav. Fundimentaliosios eismo srautų diagramos (Zaidi *et al.* 2015)

1.1.2 Eismo valdymas

Pagrindinis eismo valdymo sistemos tikslas – nustatyti optimalius valdymo parametrus, siekiant sumažinti kelionės trukmę bei delsą, patiriamą spūstyse (Zheng, Recker 2013). Pasak Zheng, Recker (2013), šiam tikslui įgyvendinti sprendžiami šie uždaviniai:

1. Matematinio modelio, apibūdinančio eismo srautus, kūrimas ir integravimas į eismo valdymo sistemą.
2. Valdymo tikslo bei parametro, apibūdinančio esamą sistemos būseną, nustatymas bei įvertinimas.
3. Optimalaus valdymo sprendimo parengimas bei taikymas.

Eismo srautų valdymas atliekamas keičiant eismo reguliavimo prietaisų (šviesoforų, interaktyvių ženklų) parametrus. Pasak Younes, Boukerche (2016), taikant šviesoforinį sankryžos valdymą, reguliuojami šie parametrai:

- Ciklo trukmė – periodas, per kurį atliekama viena eismo valdymo signalų seka.
- Fazės trukmė – eismo valdymo signalų derinio, nustatančio judėjimą sankryžoje, trukmė.
- Intervalo trukmė – vieno konkretaus signalo trukmė (skirto vienai eismo juostai).
- Signalų vėlinimas – laiko tarpas tarp fazių keitimo (geltono šviesoforo signalo trukmė). Skirtas eismo inertiškumo efektui sumažinti.
- Ciklo vėlinimas – laiko pataisa, naudojama koordinuoto valdymo šviesoforų grupėje, siekiant užtikrinti nestabdomo eismo koridoriaus judėjimą (žalios bangos principui įgyvendinti).

Eismas gali būti valdomas taikant fiksuoto laiko strategiją arba adaptyviai reaguojant į realiuoju laiku kintančius eismo srautus.

Taikant fiksuoto laiko strategiją valdymo signalo parametrai (fazė, ciklo bei signalo trukmė) nustatomi remiantis tam tikro paros meto vidutiniais eismo srautais. Sudaromi individualūs planai skirtingoms dienoms, konkrečiam metui. Šis metodas gerai tinka kelių sankirtoms, kai viena kryptimi juda daug didesnis eismo dalyvių kiekis nei kita. Šis metodas nevertina konkrečių aplinkybių, todėl gali netikti netipinei situacijai, nereaguoja į eismo srauto dinamiką realiuoju laiku. Valdymo įgyvendinimo kokybė priklauso nuo to, kaip bus įvertinti vidutiniai paros srautai. Šie planai turi būti nuolatos naujinami, kad atitiktų ilgalaikėje perspektyvoje kintančius eismo srautus.

Remiantis Shrestha *et al.* (2010), adaptyvus eismo valdymas įgyvendinamas taikant intelektualias transporto sistemas. Intelektuali transporto sistema pasižymi reagavimu į realius eismo srautus, parkavimo vietų koordinavimu, eismo dalyvių informavimu, eismo įvykių detekcija

bei reakcija. Lin *et al.* (2017) išskiria 4 pagrindinius sisteminis lygmenis, iš kurių sudaroma adaptyvi eismo valdymo sistema:

1. Fizinis: eismo dalyviai, transporto priemonės, infrastruktūra.
2. Komunikacijos: ryšio sistema tarp eismo dalyvių ir infrastruktūros.
3. Operacinis: valdymo sprendimų priėmimo dalis.
4. Serviso: valdymo įgyvendinimo dalis (šviesoforai, švieslentės, kelio ženklai, eismo dalyvių informavimo sistemos).

Remiantis Lin *et al.* (2017), tolimesnė kryptis kuria turi vystytis intelektualios transporto sistemos yra daiktų interneto integracija (angl. *Internet of Vehicle*, IoV). Tai apima kiekvieno eismo dalyvio komunikacija su aplinka bei valdymo sistema, sujungiant eismo elementus tarpusavyje į interaktyvų duomenų mainų tinklą, vadinamą VANET (angl. *Vehicular ad-hoc network*). Sekiama, kad kiekvienas eismo dalyvis atliktų mobilaus jutiklio funkciją ir taip būtų sudarytas platus bei dinamiškas duomenų tinklas, gebantis įvertinti kiekvieno eismo dalyvio trajektoriją bei nustatyti tipines keliones.

Vienas iš būdų vykdyti daiktų interneto integraciją į eismo sistemą – išmaniųjų įrenginių naudojimas. Pavyzdžiui navigavimas naudojant *Waze* ar *Google Maps* programas. Naudojant šias programas galima sudaryti kelionės maršrutą bei įvertinti preliminarą kelionės trukmę. Programos vykdo duomenų mainus tarp vartotojų bei virtualiame debesyje esančios duomenų bazės. Renkami konkretaus vartotojo duomenys, tokie kaip: identifikacijos žymuo, vidutinis kelionės greitis, kelionės trukmė. Taip pat yra galimybė kaupti duomenis apie eismo ribojimus, maksimalų leistiną greitį, identifikuoti eismo įvykius, kelio darbus. Nustačius vartotojo būvimo vietą, programos patiekia jam aktualią informaciją, padedančią pasirinkti optimalų maršrutą bei įvertinti galimus nesklandumus. Pasak Sumayya, Shefeena (2015), esant dideliame vartotojų kiekiui, šios programos sudaro platų judančių jutiklių tinklą, kurį integravus į eismo valdymo sistemą būtų sudaryta galimybė efektyviau valdyti eismo srautus. Vykdam duomenų kaitą pasiekama abipusė nauda – vartotojas gauna duomenis apie situaciją eisme ir gali pasirinkti greičiausią bei patogiausią maršrutą, o paslaugos tiekėjas kaupia duomenis ir gali juos panaudoti komercinėje veikloje.

1.2 Prognozės metodų apžvalga ir klasifikavimas

1.2.1 Prognozės metodų taikymas eismo valdymo sistemoje

Prognozavimas – būsimos situacijos spėjimas. Prognozės modeliai generuoja prognozę – kintamąjį ar kintamųjų rinkinį, apibūdinantį situaciją ateityje.

Modernios eismo valdymo sistemos pasižymi adaptyviu veikimu, geba realiuoju laiku reaguoti į eismo srautus. Remiantis Boloshinsky, Freidman (2012), adaptyviai veikiant pagal realiuoju laiku gaunamus duomenis, valdymo sprendimo priėmimas bei įgyvendinimas užtrunka, susidaro vėlinimas, priklausantis nuo eismo valdymo sistemos sudėtingumo, skaičiavimo spartos. Kol optimizuojamas eismo valdymas, eismo srautai gali pasikeisti ir optimizuotas eismo valdymo sprendimas tampa nebeaktualus, kadangi buvo sudarytas pagal jau pasenusius duomenis. Siekiant pagerinti eismo valdymą, reikia numatyti, kokia situacija bus ateityje. Eismo srautai pasižymi dėsningumais, kurie kartojasi laike. Tinkamai sumodeliavus šiuos dėsningumus ir nustačius jų priklausomybę nuo laiko ar kitų aplinkybių, galima prognozuoti, kaip keisis eismo srautai netolimoje ateityje.

Taikant prognozės metodus, sukaupti duomenys yra sisteminami ir apdorojami matematiniais modeliais. Paprasčiausi prognozės modeliai geba nustatyti tiesinius duomenų ryšius, priklausomybę laike. Sudėtingesni modeliai geba įvertinti ne tik nustatytus dėsningumus laike, o ir duomenų priklausomybę nuo kitų, išorinių faktorių, tokių kaip oro sąlygos, metų laikas, geba įvertinti eismo dinamiką įvykus eismo įvykiui. Žinant kaip keisis eismo situacija, galima iš anksto optimizuoti šviesoforų darbo ciklą ir pritakyti būsimiems eismo srautams.

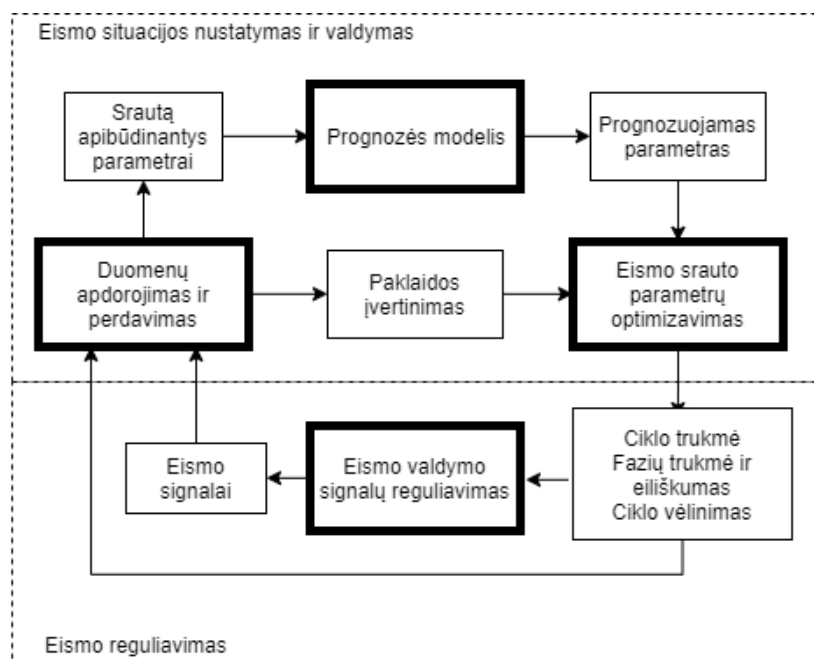
Vykdam adaptyvų eismo valdymą, visų pirma reikia nustatyti esamą situaciją. Tam, kad tiksliai būtų įvertintas eismo sistemos būvis, fiksuojami parametrai, kuriais apibūdinami eismo srautai. Stebint eismo srautus ilgą laiką, nustatomos ribinės stebimų parametrų reikšmės, kurios parodo spūsties dydį, eismo srauto dinamiką. Adaptyviai eismo sistemai reaguojant į šiuos, stebimus parametrus, yra koreguojami eismo valdymo signalai, siekiant padidinti konkretaus kelio ruožo pralaidumą.

Pasak Zheng, Recker (2013), prognozavimo proceso integracija į eismo valdymo sistemą apima šiuos etapus (1.3 pav.):

1. Pradinių duomenų kaupimas: kaupiami pradiniai duomenys, skirti sudaryti prognozės modelį. Reikalavimai pradinių duomenų imčiai nustatomi remiantis specifiniais sudaromo modelio reikalavimais: vertinamų kintamųjų skaičiumi bei pobūdžiu, duomenų fiksavimo dažniu, duomenų kaupimo laikotarpiu.
2. Prognozės modelio sudarymas: nustatomi dėsningumai duomenyse ir sudaromas matematinis modelis, apibūdinantis priklausomų kintamųjų ryšį su nepriklausomais

kintamaisiais, tokiais kaip laikas ar oro sąlygos. Ryšys apibūdinamas regresija. Taikant regresiją atliekamas prognozavimas. Modelio įvertinimo metu prognozuojamos reikšmės yra tikrinamos lyginant su realiais duomenimis ir nustatomos modelio paklaidos. Paklaida minimizuojama, optimizavimas atliekamas keičiant regresijos koeficientų reikšmes. Sudarytas prognozės modelis validuojamas: sudaromos prognozės ir lyginamos su realiais duomenimis. Jei prognozės modelis tenkina nustatytus tikslumo bei kokybės kriterijus, pavyzdžiui greitaveiką, modelis gali būti integruojamas į eismo valdymo sistemą.

3. Sudaromas modulis su parengtu prognozės modeliu. Modulis yra integruojamas į eismo valdymo sistemą, nustatomos pradinės reikšmės, skirtos inicijuoti prognozavimo procesą.
4. Inicijuojamas prognozavimas, kuris vykdomas uždaru ciklu. Pateikus nepriklausomo kintamojo reikšmę, modelis įvertina priklausomo kintamojo vertę. Ši vertė yra perduodama kitiems eismo valdymo sistemos elementams. Prognozavimą vykdant tam tikru periodiškumu, po nustatyto laiko fiksuojama nauja nepriklausomo kintamojo reikšmė, pagal kurią nustatoma sekančio žingsnio prognozė. Priklausomai nuo prognozės modelio struktūros bei greitaveikos, vieni prognozės modeliai geba realiuoju laiku perskaičiuoti modelio paklaidas ir atlikti pakartotinį modelio įvertinimą, kiti atnaujinami rečiau, pagal nustatytą periodiškumą.



1.3 pav. Prognozės modelio integracija į eismo valdymo sistemą (Zheng, Recker 2013)

Prognozės patikimumas

Prognozės patikimumas priklauso nuo dėsningų įvykių nustatymo bei galimų atsitiktinumų įvertinimo (Yue *et al.* 2007). Prognozuojama reikšmė nuo tikrosios skiriasi determinuota bei atsitiktine paklaida (1.2 formulė). Kuo šios paklaidos mažesnės, tuo prognozuojama reikšmė yra artimesnė tikrajai. Dalis atsitiktinumų gali būti nuspėti ir sumodeliuoti, todėl jų poveikis prognozuojamam kintamajam gali būti įvertintas. Atsitiktinio pobūdžio veiksniai, tokie kaip avarijos ar chaotiškas eismo dalyvių elgesys, yra sunkiai nuspėjami, gali būti nustatyti tik su tam tikra tikimybe. Todėl juos prognozuoti yra sudėtinga, tačiau taikant neparаметrinčius metodus galima įvertinti sistemos reakcija į šiuos įvykius, žinant kaip ji reagavo į panašias aplinkybes praeityje.

$$y = \hat{y} + (e + e') \quad (1.2)$$

čia y – reali reikšmė;
 $\hat{y}$ – prognozuojama reikšmė;
 e – determinuota paklaida;
 e' – atsitiktinė paklaida.

Kumar, Vanajakshi (2015) siūlo prognozės paklaidas vertinti Lewi's skale. Remiantis šia skale, eismo srautų prognozavimui realiuoju laiku galima taikyti prognozės modelius, kurių vidutinė absoliutinė procentinė paklaida (MAPE) yra mažesnė už 20% (1.1 lentelė).

1.1 lentelė. Lewi's skalė prognozės tikslumui įvertinti (Kumar, Vanajakshi 2015)

MAPE paklaida	Prognozės tikslumo įvertinimas
>10%	Labai tiksliai
10-20%	Tiksliai
21-50%	Patenkinama
>50%	Nepatenkinama

Prognozę pagal vertinamo laikotarpio trukmę Smith *et al.* (2002) siūlo skirstyti į:

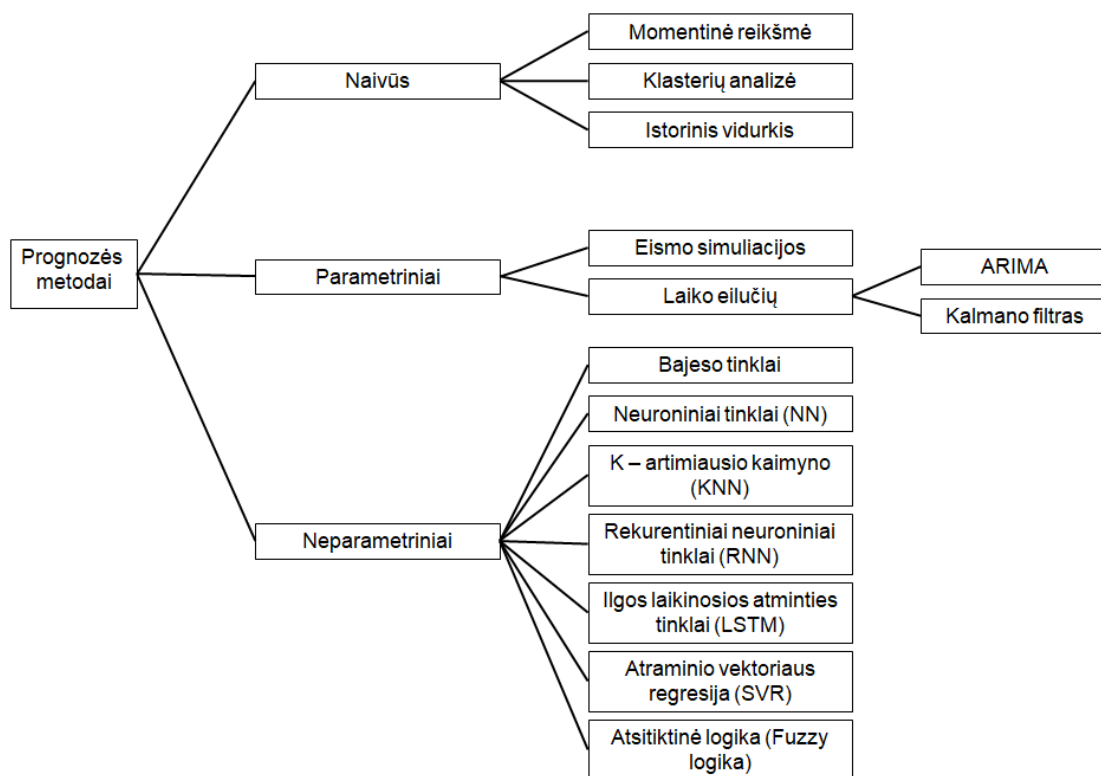
- Ilgalaikė prognozė – apima sąlyginai didelio laikotarpio vertinimą: dienos, savaitės, mėnesio, metų. Šios prognozės vykdomos siekiant nustatyti bendrus eismo srautų dėsningumus, įvertinti esamą ar projektuojamą eismo infrastruktūros pralaidumą.
- Trumpalaikė prognozė – apima trumpo laikotarpio vertinimą, įprastai iki 1 valandos. Šios prognozės tikslas – tiksliai ir greitai įvertinti eismo srauto dinamiką, remiantis praeityje sukauptų duomenų rinkiniu. Šios prognozės naudojamos optimizuoti eismo valdymo sistemą.

Nagy, Simon (2018) išskiria šiuos veiksnius, kurie lemia prognozės metodo pasirinkimą:

- Prognozavimo tikslas (eismo optimizavimui, eismo dalyvių informavimui, moksliniams tyrimams).
- Siekiamas prognozės tikslumas.
- Prognozuojamas laikotarpis.
- Prognozės modelio greitaveika.
- Prognozės modelio sudėtingumas (sudarymo, veikimo ir atnaujinimo atžvilgiu).
- Reikalavimai kompiuteriniams resursams.
- Prognozės modelio universalumas, integracijos į kitas sistemas galimybė.
- Reikalavimai duomenų bazei.
- Vertinamų kintamųjų skaičius.
- Tiriamos vietovės pobūdis (miestas, užmiestis).
- Prognozės mastas (lokalus, ruožas, tinklas).
- Tiriamo kelio intensyvumas.

Modelių klasifikacija

Prognozės modelių klasifikaciją (1.4 pav.) pagal taikomus matematinius modelius siūlo Hinsbergen *et al.* (2007).



1.4 pav. Prognozės modelių klasifikavimas (Hinsbergen *et al.* 2007)

1.2.2 Naivūs prognozės metodai

Naivūs prognozės modeliai pasižymi nesudėtinga struktūra, regresijos koeficientai nėra vertinami. Šie metodai geba atvaizduoti eismo dėsningumus, tačiau prastai reaguoja į eismo srautų dinamiką, staigius pasikeitimus, negali palyginti situacijos su įvykusia praeityje. Nors šių metodų prognozės yra laikomos kaip vienos iš nepatikimiausių, tačiau dėl jų paprastumo, momentinės modelių spartos ir nesudėtingos integracijos, šie modeliai dažnai naudojami eismo valdymo sistemose ar navigacijos programinėje įrangoje (Nagy, Simon 2018).

Momentinė reikšmė (angl. *Instantaneous travel time*)

Sekančio žingsnio prognozuojama reikšmė yra lygi esamai reikšmei (1.3 formulė). Istoriniai duomenys modelio veikimui nėra kaupiami. Modelio veikimas yra momentinis, kadangi papildomi skaičiavimai nėra atliekami, nereikalinga duomenų bazė. Tai yra vienas iš netikslių prognozės modelių ir prognozių, ilgesnių nei vieno žingsnio, nerekomenduojama atlikti.

$$\hat{y}_{t+1} = y_t \quad (1.3)$$

čia y_t – reali reikšmė laiko momentu t ;

$\hat{y}_{t+1}$ – prognozuojama reikšmė.

Istorinis vidurkis (angl. *Historic average*, HA)

Prognozuojamo kintamojo reikšmė apskaičiuojama pagal panašios istorinių duomenų imties reikšmes (1.4 formulė).

$$\hat{y}_{t+1} = \frac{y_t}{y_{t \text{ ist.}}} y_{t+1 \text{ ist.}} \quad (1.4)$$

čia $y_{t \text{ ist.}}$ – istorinė kintamojo reikšmė.

1.2.3 Parametriniai prognozės metodai

Parametrinių metodų taikymas remiasi prielaida, kad išvados apie visą populiaciją (duomenų imtį) sudaromos remiantis parametrais, nustatytais tiriant atsitiktinę populiacijos imtį. Parametriniais metodais tiriami duomenys, kurie atitinka tam tikrą teorinį skirstinį, o parametrai, naudojami skirstiniui aprašyti, yra žinomi. Ryšys tarp priklausomų ir nepriklausomų kintamųjų nustatomas naudojant tiesinės regresijos lygtis. Regresijos lygtimis apibūdinus statistinę dviejų (ar daugiau) kintamųjų priklausomybę ir sudarius regresijos modelį, galima prognozuoti priklausomo kintamojo vertę, žinant nepriklausomo kintamojo reikšmes. Eismo srautų tyrimo atveju atliekama laiko eilučių analizė, kurios metu nustatomas priklausomo kintamojo ryšys laike, su kintamojo vieta duomenų imtyje.

Parametriniai metodai gali veikti tik su stacionariais duomenimis: duomenys turi būti pasiskirstę apie konkrečią vidutinę reikšmę, neturi būti tendencingo kylimo ar leidimosi, sezoniskumo. Šie metodai negali vertinti netiesinių dinaminų procesų (Nagy, Simon 2018).

Parametriniai metodai yra skirstomi į:

- Laiko eilutes (laikines sekas) – duomenų imtis, sudaryta iš nekintančiu dažniu fiksuojamo parametro reikšmių. Modelio sudarymo metu nustatoma duomenų tiesinė priklausomybė, regresijos lygtis, pagal kurią nustatoma būsimo laiko momento reikšmė.
- Eismo simuliacijas – eismo modeliavimas, vykdomas pagal nustatytas pradines parametrų reikšmes. Keičiant pradinis modelio parametrus bei juos siejančias regresijos lygtis yra modeliuojami rezultatai, apibūdinantys proceso kaitą. Taikomi ilgalaikių prognozių metu, siekiant nustatyti bendrus dėsningumus, sankryžų judėjimo schemų, infrastruktūros pokyčių poveikį eismo dalyvių judėjimui (Nagy, Simon 2018).

ARIMA modeliai

Laiko eilutės, kurios yra stacionarios, gali būti apibūdintos taikant autoregresinį slankiųjų vidurkių modelį (angl. *Autoregressive Moving average*, ARMA). Modelis sudaromas sujungus autoregresinį procesą AR, apibūdinantį kintamojo regresinę priklausomybę su kitomis laiko eilutės reikšmėmis bei slankiųjų vidurkių procesą MA, apibūdinantį modelio paklaidų ryšį. ARMA modelį išpopuliarino G. Box, G. Jenkins 1970 m. išleistoje knygoje „Time series analysis: Forecasting and control“. Šioje knygoje rekomenduojama metodika ARIMA grupės modelių sudarymui yra plačiai taikoma prognozės modelių tyrimuose.

ARMA modeliai geba apibūdinti tik stacionarios laiko eilutės duomenis. Jei laiko eilutės stacionarumas yra nepakankamas, naudojamas integruotasis ARIMA modelis. Tai išplėstas ARMA modelis, nustatantis ne tiesioginį ryšį tarp laiko eilutės reikšmių, o ryšį tarp šių reikšmių skirtumų. Naudojant šį modelį, į ARMA modelio struktūrą papildomai įtraukiama diferencijavimo eilė d , nurodanti diferencijavimo eilę, reikalingą duomenų stacionarumui pasiekti.

ARIMA modelių pagrindinis požymis – nėra vienareikšmiškos modelio išraiškos, tiksli struktūra yra sudaroma kiekvieną kartą nustatant konkrečius modelio parametrus, priklausomai nuo tiriamos duomenų imties. Modelio struktūra žymima p, d, q raidėmis, atitinkančiomis koreliuotų vėlavimų bei diferencijavimo eiles: p raide nurodoma autoregresinė priklausomybė su p vėlavimais, žymima $AR(p)$, d – diferencijavimo eilė, q – slankiųjų vidurkių vėlavimo eilė, žymima $MA(q)$. Regresinė modelio dalis $AR(p)$ remiasi ryšiu tarp ankstesnių p reikšmių, todėl viena iš pagrindinių ARIMA modelių sąlygų, keliamų duomenims – buvusios reikšmės turi būti nepriklausomos viena nuo kitos, neturi būti koreliacijos tarp jų. Slankiųjų vidurkių modelis $MA(q)$ remiasi principu, kad

laiko eilutei poveikį sukelia įvykiai, kurių reikšmės yra pasiskirsčiusios pagal žinomą skirstinį, su nuliniu vidurkiu ir nekintančia dispersija. Atsitiktinis įvykis sukelia poveikį reikšmėms, perstumtoms iki q eilės ir tada visiškai išnyksta. MA modeliu bandoma paaiškinti šio poveikio dydį ir sumažinti sukeliamą triukšmą, remiantis nustatytomis modelio paklaidomis. Slankiųjų vidurkių modelis sušvelnina atsitiktinio triukšmo poveikį duomenims, geba gerai įvertinti ir nuslopinti atsitiktinių paklaidų sukeltus šuolius.

ARIMA modelio galimybės yra išplėčiamos modifikuojant modelio struktūrą. Egzistuojant stipriam sezoniniam duomenų ryšiui, ARIMA modelis išplečiamas sezonine dalimi, vertinančia ir per sezoną perstumtų vėlavimų koreliaciją. Tokie modeliai vadinami SARIMA (angl. *Seasonal Autoregressive Integrated Moving Average*).

ARIMA modelis naudojamas nustatyti priklausomo kintamojo regresiją nuo vieno nepriklausomo kintamojo (angl. *univariate*). Jei reikalinga nustatyti ryšį su keliais nepriklausomais kintamaisiais (angl. *multivariate*), naudojamas vektorinį ryšį apibūdinantis VARIMA modelis (angl. *Vector Autoregressive Integrated Moving Average*).

Procesams, kurie pasižymi ilgalaikiu kintamųjų poveikiu, laiko eilutės atmintimi ir lėtai mažėjančia koreliacija, naudojami fraktaliniai ARFIMA modeliai (angl. *Autoregressive Fractionally Integrated Moving Average*), kurie atlieka trupmeninę integraciją.

Kalmano filtras

Metodas, skirtas nustatyti tiesinės dinaminės sistemos būseną. Būsena yra apibūdinama duomenų imtimi, kurioje yra sukaupiti triukšmingi matavimai. Remiantis šia, sukaupytų duomenų imtimi, yra sudaroma būsenos matrica. Taikant šią būsenos matricą ir remiantis buvusios prognozės paklaida, filtras nustato galimas kintamųjų reikšmes, siekiant minimizuoti vidutinę kvadratinę paklaidą (angl. *Root Mean Square Error, RMSE*).

Prognozuojamas sistemos būvis yra nustatomas remiantis esamu sistemos būviu ir nuokrypiu nuo įvertintos paklaidos. Įvedus naują kintamojo reikšmę, yra minimizuojama modelio RMSE paklaida. Minimizavimo procesas vykdomas suteikiant paklaidoms svertinius koeficientus. Kuo nustatyta paklaida mažesnė, tuo didesnę svertinį koeficientą ji įgauna ir ta reikšme yra papildoma sistemos būsenos matrica.

Slankiojo vidurkio metodas (angl. *Moving average*)

Slankiojo vidurkio metodas naudojamas laiko eilutėms, kuriose nėra tendencingo kylimo ar leidimosi, sezoniškumo. Šio metodo esmė – laiko eilutės paskutiniųjų n reikšmių vidurkio skaičiavimas. Prognozuojama reikšmė atitinka nustatyto imties pločio vidurkį. Įvedus naują

reikšmę, ji pakeičia buvusią ir vidurkis perskaičiuojamas, tokiu būdu slankusis vidurkis slenka laiko eilute.

EkspONENTINIS išlyginimas (angl. *Exponential smoothing*)

Prognozuojamo kintamojo reikšmė nustatoma naudojant visų laiko eilutės reikšmių svertinį vidurkį, kuris nekinta laike. Sekančio laiko momento reikšmė apskaičiuojama remiantis prognozuojamo kintamojo reikšme esamu laiko momentu bei šio kintamojo prognozės paklaida. Kartu su šiomis reikšmėmis naudojamas svertinis vidurkis, išreikštas lyginamąja konstanta (1.5 formulė).

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (1.5)$$

čia α – lyginamoji konstanta

Dvigubas eksponentinis lyginimas (angl. *Double exponential smoothing*, DES)

Modelio veikimas remiasi slankiojo vidurkio metodo taikymu su papildomai naudojamais svertiniais koeficientais. Prognozuojama reikšmė nustatoma pagal greta esančių reikšmių vidurkį. Svertiniais koeficientais įvertinamas reikšmių nuotolis, arčiau esančių reikšmių svertiniai koeficientai yra didesni, kadangi jos laikomos kaip labiau aktualios prognozuojamai reikšmei. Tokiu būdu pasenę duomenys yra pašalinami iš prognozės skaičiavimo, kadangi po tam tikro laiko jie praranda aktualumą.

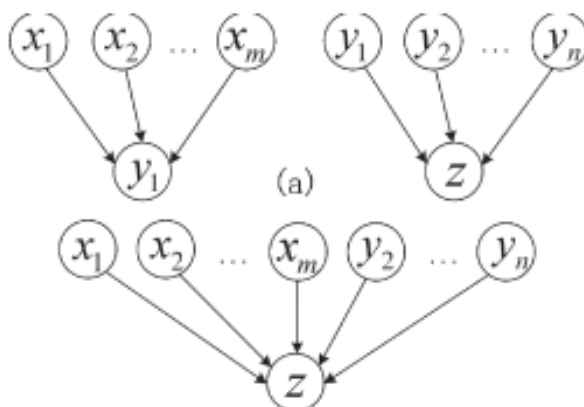
1.2.4 Neparametriniai prognozės metodai

Neparametriniais metodais tiriamos imtys, kurių pasiskirstymas neatitinka teorinio skirstinio arba šis skirstinys yra nežinomas. Šiais metodais nustatomos netiesinės, dinaminės duomenų priklausomybės, kurios vėliau taikomos prognozuojamoms reikšmėms modeliuoti.

Neparametrinių metodų veikimui labai svarbūs pradiniai duomenys, naudojami modeliui sudaryti. Šių modelių sudarymas yra vadinamas modelių mokymu, kadangi modelis susipažįsta su esamais duomenų ryšiais, kintamųjų priklausomybėmis, atlieka duomenų grupavimą ir nustato svertinius koeficientus. Neuroninių tinklų struktūra sudaryti modeliai geba savarankiškai keisti tinklą sudarančių sinapsių svertinius koeficientus, priklausomai nuo įvedamų kintamųjų ir jų sukulto poveikio. Smith *et al.* (2002) teigia, kad neparametrinių modelių paieškos mechanizmas nėra toks svarbus, kaip tinkamai suformuota, įvairiomis situacijomis pasižyminti duomenų imtis. Pasak autorių, modelių paieškos mechanizmas labiau lemia greitaveiką, nei prognozės tikslumą.

Bajeso tinklai (angl. *Bayesian Networks*)

Tai neuroninių tinklų modelis, veikiantis klasifikatoriaus principu. Tinklas mokosi atpažinti duomenis pagal tam tikrus požymius. Prognozavimo metu, prognozės reikšmė parenkamos remiantis didžiausią tikimybę turinčios duomenų grupės pasiskirstymu (1.5 pav.).



1.5 pav. Bajeso tinklo grafinis pavyzdys (Sun *et al.* 2006)

K – artimiausio kaimyno (angl. *K-nearest neighbors*, KNN)

Klasifikatoriaus principu veikiantis metodas. Prognozuojama reikšmė nustatoma pagal panašiausią praeities duomenų rinkinį. Duomenų rinkinys yra apibūdinamas būsenos vektoriumi. Naudojamas regresinis algoritmas, kuris esamos sistemos būsenos vektorių lygina su praeities duomenimis ir pagal nustatytas panašumo taisykles suranda panašiausius duomenų rinkinius. Panašumas yra įvertinamas svertiniais koeficientais. Prognozuojama reikšmė nustatoma su tikimybe, kad panašiausio rinkinio duomenys labiausiai atitiks būsimą laiko momentą. Būsenos vektorių panašumui įvertinti naudojami metodai, tokie kaip Euklido erdvė. Modelio sudarymo metu nustatomi kriterijai ir panašumo riba, kuriai esant duomenys yra laikomi kaip reikšmingi.

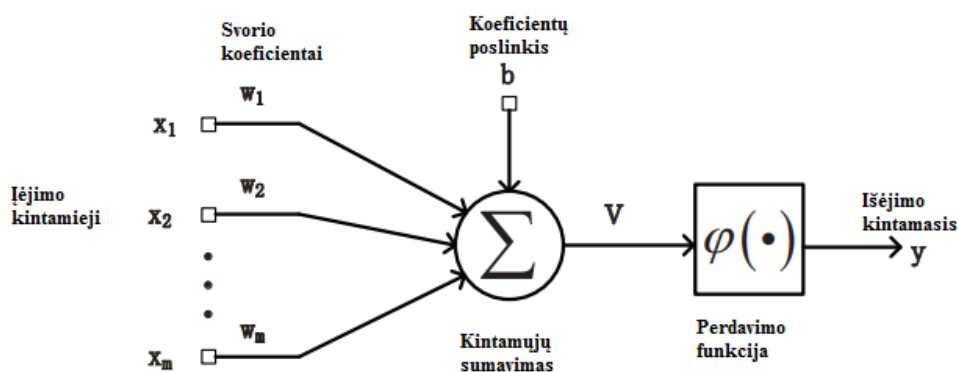
Šis modelis labai priklauso nuo turimų duomenų imties. Kadangi modelis ieško panašiausios sistemos būsenos praeities duomenyse, gali gerai įvertinti ir pasikartojančias netipines situacijas, tokias kaip eismo įvykiai. Vienas iš šio metodo trūkumų – artimų būsenų paieška yra atliekama realiuoju metu. Modelio greitaveika labai priklauso nuo paieškos mechanizmo. Esant dideliai duomenų imčiai, gali būti sudėtinga pritaikyti šį modelį prognozuoti eismo srautus realiuoju metu, jei paieška truks ilgiau nei prognozuojamas laikotarpis. Jei imtyje nėra panašumo kriterijus atitinkančių būsenos vektorių, tiksli prognozė yra neįmanoma. Dėl šios priežasties reikalinga didelė duomenų imtis, pasižyminti įvairių situacijų gausa.

Neuroniniai tinklai (angl. *Neural Networks*, NN)

Neuroninis tinklas susideda iš tarpusavyje sinapsėmis (gijomis) susijungusių neuronų. Neuronų mazgai tarpusavyje išdėstyti skirtingais lygiais, taip sudarant skirtingas sprendimų priėmimo sritis. Šie modeliai geba vertinti netiesines dinamines sistemas, stacionares bei nestacionares laiko eilutes, gali klasifikuoti sistemos būseną pagal nustatytus požymius. Pasak Zhang *et al.* (2016) neuroninių tinklų metodai pagrįsti mokymosi principu, remiantis žmogaus smegenų veikla. Šie tinklai geba atpažinti dėsningumus, juos klasifikuoti, įvertinti priskirtų veiksnių poveikį, išskirti bei vertinti eismo srauto dinamiką.

Neuroną su įvesties duomenimis jungia sinapsės, turinčios svorio koeficientus W_n (1.6 pav.). Sudarant neuroninių tinklų modelį, mokymo procese nustatomos pradinės svorio koeficientų vertės, kurios vėliau koreguojamos modelio optimizacijos metu. Neurone įvertinamas nepriklausomų kintamųjų poveikis. Tą poveikį atitinkanti reikšmė v taikant perdavimo funkciją yra transformuojama į nustatyto formato reikšmę y , kuri perduodama aukštesnio lygio neuronui. Pastarasis, gavęs informaciją iš žemesnio lygio neuronų, atlieka skaičiavimą ir naudojant perdavimo funkciją siunčia nustatytą vertę į aukštesnį lygį. Paskutiniame tinklo sluoksnyje yra regresinis mechanizmas, siejantis tinklo apskaičiuotas reikšmes su pradiniu įvesties duomenų formatu ir pateikia galutinį rezultatą – prognozuojamą priklausomo kintamojo vertę. Tinklo mokymo procesas vyksta mažinant paklaidą, susidariusią tarp realių duomenų bei ankstesnės prognozuojamo kintamojo vertės. Pagal nustatytą paklaidą bei mokymosi parametrus b , ribojančius staigius šuolius, keičiami žemesnių lygių neuronų svertiniai koeficientai, stebimas poveikis galutiniam rezultatui. Dėl galimybės įvertinti didelį kiekį skirtingų nepriklausomų kintamųjų, neuroninių tinklų modeliai gali įvertinti skirtingas aplinkybes, veikiančias eismo srautus.

Neuroniniai tinklai gali būti tiesioginio sklaidimo arba grįžtamojo ryšio. Tiesioginio sklaidimo tinklai remiasi tik nustatytais svertiniais koeficientais. Grįžtamojo ryšio tinklų išvesties reikšmės yra perduodamos sekančio žingsnio iteracijai, todėl grįžtamojo ryšio tinkluose yra įgyvendinta atminties funkcija. Grįžtamąjį ryšį turintys tinklai yra vertinami kaip tinkamesni laiko eilučių analizei atlikti, kadangi gali įvertinti ilgalaikes duomenų priklausomybes.



1.6 pav. Neurono ryšiai (Zhang *et al.* 2016)

Rekurentiniai neuroniniai tinklai (angl. *Recurrent neural network*, RNN)

Tiesioginio sklidimo neuroninių tinklų architektūroje susiduriama su problema sprendžiant laiko eilučių uždavinius, šių tinklų galimybe įvertinti dėsningą duomenų seką, kadangi tinkle kaupiami tik svertiniai koeficientai. Tradiciniai neuroniniai tinklai veikia principu, kad kiekvienu laiko momentu yra įvedamas naujas nepriklausomas duomenų vektorius ir informacija apie kiekvieną naujai įvestą duomenų vektorius nėra kaupiama, o prognozuojama reikšmė nustatoma pagal turimus svertinius koeficientus, apibūdinančius sistemos būvį. Tai yra problema nagrinėjant laiko eilučių uždavinius, kadangi jas sudaro didelės laiko sekos, kuriose gali būti periodiškai pasikartojantys dėsningumai. Siekiant spręsti šią problemą, sudaromi grįžtamojo ryšio rekurentiniai neuroniniai tinklai (RNN).

Kiekvieno tinkle esančio mazgo išvesties reikšmė y_t laiko momentu t yra išsaugoma paslėptame tinklo būsenos sluoksnyje h_t (1.7 pav.). Sekančiame žingsnyje išvesties reikšmė y_{t+1} yra nustatoma įvertinus buvusią sistemos būseną h_t bei įvesto kintamojo reikšmę x_t . Šios reikšmės yra padauginamos iš sudarytų svorio matricių W_{hy} , W_{xh} , o perdavimo funkcija atlieka duomenų transformavimą ir pateikia galutinę reikšmę. Remiantis tinklo struktūra, ši reikšmė yra perduodama aukštesniam tinklo lygiui arba išvedama į išorę, kaip galutinė prognozė.

RNN tinklo išvesties reikšmė (naudojant hiperbolinio tangento perdavimo funkciją) gali būti nustatyta pagal (1.6) formulę:

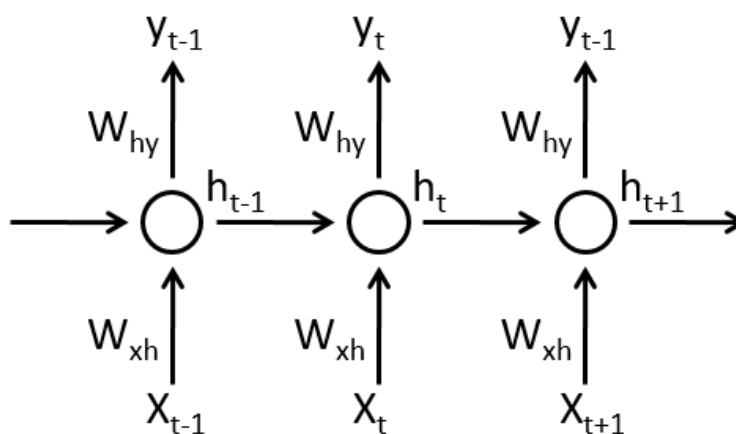
$$y_t = \tanh(W_{hy}y_{t-1} + W_{xh}x_t) \quad (1.6)$$

čia $\tanh$ – hiperbolinio tangento funkcija;

$W_{hy}; W_{xh}$ – įvesties bei išvesties reikšmių svorio matricos;

y_t – išvesties reikšmė laiko momentu t ;

x_t – tinklo įvesties kintamasis.



1.7 pav. RNN tinklo struktūra (Yuan et al. 2019)

Ilgos laikinos atminties tinklas (angl. *Long short term memory*, LSTM)

Taikant RNN modelį susiduriama su ilgalaikės atminties problema – tolstant nuo pradinio mazgo, konkreti tam tikro laiko momento išvesties reikšmė yra pamirštama dėl mažėjančio gradiento. Todėl šių tinklų atmintis nėra ilga ir tiriant ilgalaikėmis priklausomybėmis pasižyminčias duomenų imtis prarandama geba įvertinti senų duomenų reikšmę būsimam laiko momentui. Ilgalaikės atminties problema yra sprendžiama taikant ilgos laikinos atminties tinklus (LSTM), kuriuose RNN mazgai yra pakeičiami į sudėtingesnės struktūros atminties blokus, vadinamus ląstelėmis. LSTM tinklo ląstelių viduje yra paslėptieji sluoksniai, valdomi aktyvacijos funkcijomis. Šie sluoksniai vertina praeities informacijos svarbą, kaupiamą tinklo atmintyje ir ją atnaušina, priklausomai nuo naujai įvestos nepriklausomo kintamojo reikšmės. Prognozė pateikiama įvertinus esamo žingsnio kintamojo reikšmę bei ląstelės atmintyje sukauptą patirtį. Dėl to tinklas pasižymi stabiliu duomenų vertinimu, nėra jautrus atsitiktiniams nuokrypiams. Ląstelės struktūrinė schema ir duomenų srauto valdymas detaliau aptartas 2.2 poskyryje.

Vienas iš pagrindinių RNN ir LSTM modelių skirtumų – skaičiuojant išvesties reikšmę RNN modelis, taikydamas aktyvacijos funkciją, perduoda visus transformuotus duomenis, nevertindamas jų svarbos ir aktualumo. LSTM modelis geba įvertinti kurią informaciją kaupti, o kurią pašalinti. Taip yra įgyvendinama ilgalaikė LSTM modelio atmintis, kuria jis naudojami skaičiuodamas tolimesnes prognozes. Todėl šis modelis geba tiksliai atskirti atsitiktinius nuokrypius nuo egzistuojančių laiko eilutės dėsnų.

1.2.5 Hibridiniai modeliai

Siekiant pagerinti prognozavimo galimybes, bandoma jungti skirtingus prognozės metodus, siekiant geriau išnaudoti jų stipriasias puses. Galimi įvairūs jungimo būdai, pavyzdžiui parametrinių modelių jungimas tarpusavyje (ARIMA derinama kartu su Kalmano filtru) ar jungiami parametriniai modeliai su neparametriniais (Bolshinsky, Freidman, 2012). Įprastai tokiais atvejais vienas metodas naudojamas duomenų paruošimui ar grupavimui, kitas atlieka prognozę. Taip pat galimas atvejis, kai vienas metodas atlieka pirminę prognozę, o kitas atlieka pirminės prognozės optimizavimą.

Šių metodų tikslumas dažnai būna panašus arba geresnis nei atskirų modelių, tačiau jų struktūra sudėtinga, o taikant kelis sujungtus modelius skaičiavimo trukmė smarkiai padidėja. Pasak Nagy, Simons (2018), neretai tyrimai įrodo, kad reikšmingo skirtumo tarp nemodifikuotų metodų ir hibridinių modelių nėra.

1.3 Mokslo darbų apžvalga

Prognozės modelio pasirinkimą eismo srautams tirti lemia jų pobūdis. Parametriniai prognozės metodai dėl statistinių ryšių vertinimo geba gerai apibrėžti stabilius eismo srautus, tačiau negali tinkamai įvertinti staigių pokyčių bei tiksliai prognozuoti esant netipinei situacijai. Neparametriniai metodai priešingai, geba įvertinti sistemos reakciją esant netipinėms aplinkybėms, tačiau šie metodai pasižymi kompleksiskumu ir sunkiau vertina sezoninius dėsningumus. Dėl šių metodų skirtumų daugybė tyrėjų stengiasi nustatyti prognozės modelių taikymo ribas bei sudaryti prognozės modelių naudojimo rekomendacijas optimizuojant šviesoforų darbo ciklą.

Vienas iš pavyzdinių darbų, dažnai cituojamų kituose literatūros šaltiniuose, yra Smith *et al.* (2002) atliktas parametrinių ir neparametrinių modelių palyginimas. Autorių darbe buvo tiriami parametriniai ARIMA, SARIMA bei neparametrinis KNN modelis. Šių modelių palyginimas buvo atliktas taikant naivųjį istorinio vidurkio metodą. SARIMA modelio periodas sudarė 1 savaitę. Modelių tikslumas vertintas vidutine procentine absoliutine paklaida (MAPE). Autoriai siūlo taikyti vieną MAPE paklaidos vertinimą tiriant eismo srautus, kadangi ši paklaida mažiau priklauso nuo tiriamojo parametro matavimo skalės ir galima palyginti skirtingų tyrimų rezultatus, prognozuojant skirtingus kintamuosius. Tyrimas buvo atliekamas didelio intensyvumo greitkelio ruože, prognozuojant 15 min laikotarpio intensyvumą. Visų modelių prognozės buvo įvertintos kaip tikslios. Netiksliausias buvo istorinio vidurkio metodas, kurio MAPE siekė 9,9%. Geriausia KNN modelio versija siekė 9,4% paklaidą. Mažiausia paklaida buvo fiksuota taikant SARIMA(1,0,1)(0,1,1) modelį, 8,8%. Išvadose autoriai išryškino parametrinių modelių gebą gerai nustatyti dėsningumus savaitės laikotarpiu ir nesant išskirtinių situacijų, šių modelių veikimas yra geresnis už tirtų neparametrinių modelių.

Eismo srautų sezoniškumas įrodytas Williams, Hoel (2003) straipsnyje. Autoriai nustatė 1 savaitės periodinius duomenų pasikartojimus ir išbandė SARIMA modelio taikymą trumpalaikėms eismo srautų prognozėms. Tirtas 15 min eismo intensyvumas greitkelyje. Rezultatai palyginti su atsitiktinio klaidžiojimo (angl. *Random Walk*, RW) ir HA metodais. Tiksliausiai veikė SARIMA(1,0,1)(0,1,1) modelis, kurio paklaida siekė 9% MAPE. HA metodo paklaida siekė 11%, RW 12%.

Parametrinių ir neparametrinių modelių tikslumo vertinimą atliko ir Zhang *et al.* (2016). Autorių straipsnyje palyginti 5 metodai: Dvigubas eksponentinis lyginimas (DES), SARIMA, atraminio vektoriaus regresija (angl. *Support Vector Regression*, SVR), KNN, grįžtamo ryšio neuroniniai tinklai (angl. *Back propogation neural networks*, BP). Autoriai siekė nustatyti prognozės modelių tikslumą skirtingomis savaitės dienomis. Buvo tiriamas eismo intensyvumas greitkelyje. Pradiniai duomenys, skirti modeliui sudaryti, fiksuoti 1 mėn., prognozuojamas

laikotarpis nenurodytas. Rezultatai palyginti MAPE paklaidomis. Tiksliausią prognozę darbo dienomis bei savaitgalio metu sudarė SARIMA modelis, MAPE vidutiniškai siekė iki 6%. BP neuroninio tinklo prognozės rezultatas buvo artimas SARIMA modeliui, paklaida siekė 6–7%. Kitų modelių paklaidos buvo didesnės 2–10%, tačiau neviršijo 15%, todėl visi modeliai buvo įvertinti kaip tikslūs. Tyrimo metu autoriai pastebėjo, kad prognozės modeliai prasčiau prognozavo eismą, kai srauto intensyvumas buvo nedidelis, visų tirtų modelių tikslumas buvo mažesnis prognozuojant intensyvumą savaitgalio metu. Autoriai tai susiejo su mažesniu automobilių skaičiumi ir mažiau tendencingais eismo srautais.

Xu *et al.* (2013) atliko klasifikuoti gebančių prognozės modelių tyrimą. Lygintas vektorinės regresijos modelis CART, KNN bei Kalmano filtras. Duomenys buvo fiksuojami 1 mėn., penkiuose matavimo punktuose, didelio intensyvumo greitkelio ruože. Iš pradinių duomenų buvo suformuoti būsenos vektoriai, skirti 15 min laikotarpiui prognozuoti. Autorių darbe pastebėtas didelis duomenų sezoniškumas, atitinkantis 1 paros periodą. Mažiausia MAPE paklaida buvo CART modelio – 8–9%. KNN modelio paklaidos siekė 8–12%, Kalmano filtro 10–15%.

Prognozės modelių tikslumą skirtingu savaitės bei paros metu tyrė Dong *et al.* (2009). Savaitė buvo išskaidyta į: pirmadienio, antradienio–ketvirtadienio, penktadienio, savaitgalio duomenų imtis. Antru atveju para buvo išskaidyta į piką atitinkančius laiko intervalus bei vidurdienį. Taip pat autoriai įvertino skirtingo laikotarpio prognozių tikslumą, prognozuojant 5 ir 10 min žingsniu. Paklaidų įvertinimui naudota MAPE paklaida. Tyrimo metu buvo taikomi ARIMA modeliai. Siekiant palyginti rezultatus, pagal pradinių duomenų imtį buvo sudarytas bazinis ARIMA modelis. Modeliai, kurie buvo taikomi specifiniam paros metui prognozuoti, pasižymėjo didesniu tikslumu. Mažiausios paklaidos buvo vidurdienio bei savaitgalio metu, kai eismo srautai buvo stabiliausi, MAPE paklaida siekė 10%. Piko metu šių modelių paklaidos padidėjo iki 14%. Bazinio ARIMA modelio paklaidos siekė 18%. Dėl šių rezultatų autoriai rekomenduoja sudaryti ARIMA modelius, atsižvelgiant į nagrinėjamą laikotarpį ir jei įmanoma – išsikirti charakteringus laiko intervalus. Taip pat nustatyta, kad prognozuojant 5 min laikotarpį paklaidos buvo 2% mažesnės, nei prognozuojant 10 min laikotarpiu.

Prognozės modelių taikymą tik tam tikro paros meto intensyvumui prognozuoti tyrė ir Qiao *et al.* (2017). Autoriai darbe lygino ARIMA, LSTM, SVR ir Kalmano filtro prognozės modelius, prognozuojant 7–10 ir 17–19 valandų intensyvumą. Prognozuojamas 3 min laikotarpis. Vertintos MAPE ir RMSE paklaidos. Tiksliausias buvo LSTM modelis, kurio MAPE paklaida siekė 6,7%. ARIMA modelio paklaida siekė 7,6%, SVR 7,8%. Mažiausiu tikslumu pasižymėjo Kalmano filtras – 9,46%.

Eismo srautai gali būti apibūdinami taikant skirtingus parametrus. Mohamed *et al.* (2015) tyrė prognozės modelių efektyvumą prognozuojant vidutinę kelionės trukmę. Autoriai vidutinę

kelionės trukmę apskaičiavo naudodami vidutinio greičio duomenis, gautus iš *Google Maps* svetainės. Tyrimo metu susidurta su dideliu duomenų nestabilumu, todėl siūloma papildomai taikyti glodinimo metodus. Sudarytas hibridinis prognozės modelis: Holt–Winter‘*s* metodas buvo naudojamas eksponentiniam duomenų lyginimui, o ARIMA modeliu buvo atliekamas prognozavimas. Modelio tikslumas vertintas determinacijos koeficientu R^2 . Hibridinio modelio tikslumas palygintas su paprasto ARIMA modelio tikslumu. Hibridinis modelis tiksliau prognozavo piko metu, kitu laikotarpiu hibridinio modelio paklaidos nesiskyrė nuo ARIMA modelio. Autoriai apibendrino, kad piko metu duomenys pasižymėjo didesniu nestabilumu, todėl tik tuo metu išryškėjo hibridinio modelio pranašumas.

Prognozuojamo parametro poveikis tirtas ir Dong *et al.* (2014) darbe. Tyrime palygintas ARIMA ir Elman modelių veikimas prognozuojant intensyvumą bei vidutinį greitį. Remiantis autorių rezultatais, vidutinė absoliutinė paklaida (MAE) ir MAPE buvo mažesnė prognozuojant vidutinį kelionės greitį, MAPE siekė 10%. Prognozuojant eismo intensyvumą, MAPE padidėjo iki 15–20%.

Prognozuoti kelionės trukmę bandė Billings, Yang (2006). Kelionės trukmė buvo nustatoma atliekant tiesioginius laiko matavimus, važiuojant nustatytu maršrutu. Siekiant užtikrinti tyrimo tikslumą, maršrutas išskaidytas į tam tikrus ruožus ir buvo fiksuojamos laiko žymos, naudojant GPS siūstuvą. Autorių darbe nagrinėta ARIMA modelio sudarymo specifika vertinant vidutinę kelionės trukmę, detaliau analizuotas duomenų chaotiškas pobūdis bei galimas nestacionarumas. Išvados apie modelio tikslumą atliktos remiantis nedidelėmis modelio liekamosiomis paklaidomis. Rezultatai pateikti tik grafiškai, modelių validacija lyginant sudarytas prognozes su realiais duomenimis nebuvo nagrinėta.

Prognozuojamo laikotarpio tyrimą bei parametrinių ir neparametrinių modelių palyginimą atliko Zheng *et al.* (2016). Autoriai tyrė LSTM, ARIMA, SVR, RNN modelių veikimą. Tyrimo metu buvo naudojami 3 jutiklių duomenys. Intensyvumas miesto gatvėje buvo fiksuojamas 5 min dažniu ir duomenys buvo agraguojami į 15, 30, 45 ir 60 min intervalus, siekiant palyginti skirtingo prognozuojamo laikotarpio paklaidas. Vertinta vidutinė santykinė paklaida (angl. *Mean relative error*, MRE). Autoriai nustatė stiprų ryšį tarp prognozės paklaidų ir prognozuojamo laikotarpio trukmės. Mažiausiomis paklaidomis pasižymėjo LSTM modelis. Esant 15 min laikotarpiui, modelio paklaidos siekė 6% MRE. RNN modelio tikslumas buvo artimas LSTM modeliui, o ARIMA mažesnis, siekė 8–15%. Padidinus prognozuojamą laikotarpį iki 30 min, LSTM modelio paklaida buvo mažiausia, siekė 8,5–10%. Šiuo atveju išryškėjo skirtumas tarp LSTM ir RNN modelių, RNN modelio paklaidos padidėjo labiau, iki 12–14%. ARIMA modelio paklaidos siekė 12–15%. Padidinus prognozuojamo laikotarpio trukmę iki 45 min, LSTM modelio paklaidos siekė 12–14%, RNN 15–18%, ARIMA 14–18%. Prognozuojant 60 min laikotarpį, LSTM modelio paklaidos

padidėjo iki 16–17%, RNN 20–30%, ARIMA 25–26%. Tyrimo metu autoriai įrodė LSTM modelio ilgalaikės atminties poveikį didesniai prognozės tikslumui, palyginę su RNN modelio paklaidomis. RNN modelio paklaidos labiau priklausė nuo prognozuojamo laikotarpio trukmės, modelio tikslumas reikšmingai sumažėjo prognozuojant 60 min laikotarpį. Autoriai apibendrina, kad nagrinėjamomis sąlygomis LSTM modelis veikė tiksliausiai ir mažiausiai priklausė nuo prognozuojamo laikotarpio trukmės.

Autoriai Abbas *et al.* (2018) išbandė prognozės modelių tikslumą prognozuojant skirtingo laikotarpio eismo srauto tankį. Buvo tiriamas greitkelio ruožas, sudarytos 10, 20, 30 min laikotarpio prognozės. Naudoti LSTM, AR, SVR, RNN, ARIMA modeliai. Paklaidos vertintos RMSE ir MAE. Tyrimo metu autoriai nustatė paklaidų padidėjimą didinant prognozuojamo laikotarpio trukmę. Tiksliausios prognozės sudarytos 10 min laikotarpiui, naudojant LSTM modelį – 5,5 RMSE, 2,4 MAE. Tyrimo metu autoregresinio proceso (AR) paklaidos buvo mažesnės nei ARIMA modelio, siekė 6,9 RMSE, 5,9 MAE. Didžiausios paklaidos buvo ARIMA modelio. Esant 10 min prognozuojamo laikotarpio trukmei, ARIMA modelio paklaidos siekė 7,7 RMSE ir 6,7 MAE. Padidinus prognozuojamo laikotarpio trukmę iki 30 min, LSTM modelio paklaidos siekė 7,6 RMSE ir 3,5 MAE. ARIMA modelio paklaidos padidėjo iki 10,9 RMSE, 9,8 MAE. Išvadose autoriai apibendrina, kad prognozės modelių tikslumas labai priklauso nuo prognozuojamo laikotarpio trukmės. Didėjant šio laikotarpio trukmei, prognozės paklaidos taip pat didėja. Tyrimo metu tiksliausiai prognozavo LSTM modelis.

Siekiant pagerinti prognozavimo procesą gali būti sujungiami keli metodai ir sudaromi hibridiniai modeliai. Hibridinių modelių ypatumai nagrinėti Xu *et al.* (2017) darbe. Autoriai sujungė du neparimetrinius, ARIMA ir Kalmano filtro, metodus. Sudaryto modelio koncepsija rėmėsi pirminiu ARIMA modelio įvertinimu bei regresijos koeficientų optimizavimu taikant Kalmano filtrą. Galutinė prognozė buvo nustatoma taikant optimizuotą ARIMA modelį. Tyrimo metu autoriai palygino vidutinio greičio bei intensyvumo prognozių tikslumą prognozuojant eismo srautus mieste. Gauti rezultatai buvo palyginti su nemodifikuotais ARIMA ir Kalmano filtro modeliais, palyginimui naudota MRE paklaida. Hibridinis ARIMA–Kalmano filtro modelis buvo tikslesnis už atskirus modelius, vidutinio greičio prognozės buvo tikslesnės nei prognozuojant intensyvumo vertes. Remiantis pateiktomis rezultatų kreivėmis, hibridinis ARIMA–Kalmano filtro modelis gerai atkartojė realių duomenų kreivę, atskirų ARIMA bei Kalmano filtro modelių prognozės pasižymėjo nemažais nuokrypiais nuo realių duomenų, ypač piko metu.

Vienas iš prognozės metodų trūkumų – siekiant sudaryti tikslų modelį, reikalinga didelė duomenų imtis. Duomenų imties klausimas nagrinėtas Kumar, Vanajakshi (2015) darbe. SARIMA modeliui sudaryti buvo naudojami skirtingą laikotarpį fiksuoti duomenys, vertintos duomenų imtys, sudarytos iš 3 ir 9 dienų laikotarpiu fiksuotų duomenų. SARIMA modelio sezoną atitiko 1 para.

Modelių tikslumas palygintas MAPE paklaida. Esant 3 dienų laikotarpio duomenims, MAPE paklaida siekė 10%. Naudojant 9 dienas fiksuotus duomenis, modelio paklaidos sumažėjo 1,5–2%. Skirtumas nėra ženklus, 3 dienų laikotarpio duomenys buvo pakankami sudaryti adekvatų modelį. Todėl autoriai apibendrina, esant poreikiui SARIMA modeliui sudaryti galima naudoti ir trumpą laikotarpį fiksuotus duomenis. Tačiau akcentavo, kad esant nedideliai duomenų imčiai gali būti netinkamai įvertinti savaitės dėsningumai, ypač savaitės pradžioje.

Literatūros apžvalgos santrauka (lentelės formatu) pateikta 1 Priede.

2. PROGNOZĖS MODELIŲ TYRIMO METODIKA

Literatūros apžvalgoje pastebėtas geras parametrinių SARIMA modelių įvertinimas dėl jų tikslumo bei nedidelės duomenų imties, reikalingos modeliui sudaryti. Tiriamojoje dalyje pasirinkta nagrinėti SARIMA modelio sudarymo procedūrą bei įvertinti šio modelio prognozės tikslumą lemiančius veiksnius. Siekiant palyginti parametrinių bei neparimetrinių modelių savybes, bus naudojamas neparimetrinis LSTM modelis.

Tyrimo metu bus nagrinėjami šie veiksniai, svarbūs eismo srautų specifikai:

- Modelių universalumas.
- Modelių greitaveika.
- Prognozuojamo laikotarpio trukmė.
- Modeliui sudaryti reikalingas duomenų kiekis.
- Duomenų glodinimo poveikis prognozės tikslumui.

Tyrimo eiga

1. Sudaromi prognozės modeliai ir įvertinamas jų tikslumas su pradiniais duomenimis.
2. Sudaryti prognozės modeliai yra išbandomi su skirtingų eismo juostų duomenimis.
3. Sudaryti prognozės modeliai yra išbandomi atliekant skirtingo laikotarpio prognozes.
4. Įvertinamas modeliui sudaryti reikalingas duomenų kiekis.
5. Įvertinamas duomenų glodinimo metodų taikymas, siekiant padidinti prognozavimo tikslumą.

Validacija

Prognozuojamas laikotarpis atitinka vieną laiko intervalą tarp duomenų matavimų, t. y. prognozuojamas vienas laiko žingsnis. Prognozuojamos validacijai skirtų duomenų imties reikšmės. Kiekviena būsimo laiko prognozė $\hat{y}_{t+1}$ yra sudaroma remiantis realiojo laiko momento y_t duomenimis, o ne buvusiomis prognozėmis. Pakartotinis modelio įvertinimas (regresijos koeficientų perskaičiavimas) nėra atliekamas. Prognozuojamos vertės yra palyginamos su realiais duomenimis ir nustatomos suminės paklaidos. Paklaidų skaičiavimo metodika aptarta 2.1.6 poskyryje.

Greitaveikos vertinimas

Esminiuose tyrimo etapuose bus vertinama modelių greitaveika. Nustatomas modelio įvertinimo bei prognozavimo laikas. Modelio greitaveika labai priklauso nuo naudojamos kompiuterinės įrangos, todėl tyrimo metu bus labiau kreipiamas dėmesys į susidariusius skirtumus tarp modelių nei į bendras skaičiavimo trukmes. Laiko fiksavimas atliktas MATLAB programos funkcija *Run&Time*.

Vertintas laikas:

- Modelio įvertinimo laikas – laikas, per kurį įvertinamos regresinės duomenų priklausomybės. Parametrinių modelių atveju tai laikas, per kurį nustatomos arba perskaičiuojamos regresijos koeficientų vertės. Neparametrinių modelių atveju tai laikas, per kurį nustatomi duomenyse vyraujantys ryšiai. Tiriant neuroninių tinklų pagrindu sudarytus modelius (LSTM modelis), šis laikas įprastai vadinamas modelio mokymosi laiku, kadangi modelis savarankiškai nustato regresines duomenų priklausomybes atlikdamas optimizavimo procesą. Modelio įvertinimo laikas svarbus pirmą kartą vertinant pradinį duomenis ir atliekant pakartotiną modelio atnaujinimą. Jei regresinės priklausomybės nebus periodiškai perskaičiuojamos, modelis gali neatitikti ilgalaikių eismo srautų pokyčių. Dėl šios priežasties būtina tam tikru periodiškumu atlikti pakartotiną modelio įvertinimą į duomenų imtį įtraukus naujai sukauptus duomenis. Šis laikas MATLAB programoje matuojamas tiesiogiai (angl. *Estimation time*).
- Prognozavimo laikas – laikas, per kurį atliekama vieno žingsnio prognozė. Šis laikas nustatytas iš visos skaičiavimo trukmės, išmatuotos MATLAB programa, atėmus modelio įvertinimo laiką ir padalinus iš visų prognozuotų reikšmių skaičiaus.

2.1 lentelė. Skaičiavimams naudojama kompiuterinė įranga

Procesorius	Intel® Core™ i5-4200U CPU @ 1.60GHz 2.30GHz (normalus taktinis dažnis 1.60 GHz, spartinimas iki 2.30 GHz).
Operatyvioji atmintis (RAM)	8,00 Gb
Operacinė sistema	Windows 10, 64-bit
Naudojama programinė įranga	MATLAB R2020a (9.8.0.1323502), 64-bit (win64), February 25, 2020 versija

2.1 SARIMA modelis

ARIMA modelio sudarymas vyksta pagal Box–Jenkins metodiką, kuria identifikuojamas ir įvertinamas modelis. Patikrinamas duomenų imties stacionarumas, nustatomas sezoniškumas, duomenų regresija, modelio parametrai. Kadangi eismo srautai pasižymi sezoniškumu, bus tiriamas sezoninę dalį vertinantis SARIMA modelis (3.1.2 poskyryje detaliau argumentuotas SARIMA modelio pasirinkimas). Taikoma išplėsta ARIMA modelio sudarymo procedūra, papildomai įvertinant sezoninę duomenų regresiją.

2.1.1 ARIMA ir SARIMA modelio matematinis apibrėžimas

Modelio sudarymo procedūra aprašyta remiantis Brockwell, Davis (2002).

Modelio sudarymo metu naudojami operatoriai:

- Poslinkio operatorius – skirtas apibūdinti reikšmių vietą laiko eilutėje, vėlavimo eilę (angl. *Lag*):

$$L^p y_t = y_{t-p} \quad (2.1)$$

čia L – poslinkio operatorius;
 p – poslinkio operatoriaus (vėlavimo) eilė;
 y_t – laiko eilutės reikšmė laiko momentu t .

- Skirtumų operatorius – nurodo skirtumą tarp išmatuotų reikšmių. Skirtumo operatorius esant τ eilės skirtumui:

$$\nabla y_t = y_t - y_{t-\tau} \quad (2.2)$$

Poslinkio ir skirtumo operatorius sieja ryšys, sudarant regresijos lygtį apibūdinantis vėlavimo eilę:

$$\nabla y_t = y_t - y_{t-\tau} = y_t - L y_t = (1 - L) y_t \quad (2.3)$$

Autoregresinis procesas AR(p)

Autoregresinis procesas AR(p) nusako esamos reikšmės y_t priklausomybę nuo p eilės praeities reikšmių: $y_{t-1}, y_{t-2}, \dots, y_{t-p}$.

AR(p) procesas aprašomas lygtimi:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t \quad (2.4)$$

čia y_t – reikšmė momentu t ;

p – autoregresinio proceso vėlavimo eilė;

ϕ_p – p eilės vėlavimo regresijos koeficientas;

e_t – atsitiktinė paklaida.

Atsitiktinė paklaida sudaro baltąjį triukšmą. Tai pagal Gauso skirstinį pasiskirstęs atsitiktinis procesas, turintis nulinį vidurkį ir dispersiją:

$$e_t \sim WN(0, \sigma^2) \quad (2.5)$$

Kai duomenų imties vidurkis μ pradiniu laiko momentu nėra lygus nuliui, autoregresinio proceso lygtis papildoma α reikšme, kuri nusako vidurkio ryšį su AR proceso regresijos parametrais:

$$y_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t \quad (2.6)$$

$$\alpha = \mu(1 - \phi_1 - \dots - \phi_p) \quad (2.7)$$

Reikšmės susiejamos taikant poslinkio operatorių:

$$y_t(1 - \phi_1 L - \dots - \phi_p L^p) = e_t \quad (2.8)$$

$$\phi(L)y_t = e_t \quad (2.9)$$

Slankiųjų vidurkių procesas MA(q)

Slankiųjų vidurkių procesas MA(q) susieja laiko eilutės reikšmes su modelio paklaidomis $(e_{t-1}, e_{t-2}, \dots, e_{t-q})$ ankstesniais q eilės vėlavimais ir atsitiktinėmis paklaidomis e_t :

$$y_t = \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t \quad (2.10)$$

Reikšmės susiejamos taikant poslinkio operatorių:

$$\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q \quad (2.11)$$

$$y_t = \theta(L)e_t \quad (2.12)$$

ARMA(p, q) modelis

Sujungus AR(p) ir MA(q) modelius sudaromas ARMA(p, q) modelis, kurio lygtis:

$$y_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (2.13)$$

Taikant poslinkio operatorių:

$$\phi(L)y_t = \theta(L)e_t \quad (2.14)$$

ARIMA(p,d,q)

ARMA modelis gali apibūdinti tik stacionarę laiko eilutę. Kai laiko eilutės stacionarumas nepakankamas, ji gali būti diferencijuojama. Tokiu atveju prognozuojamos ne tiesiogiai išmatuotos reikšmės y_{t+h} , o jų d eilės skirtumai Δy_{t+h} . Absoliutinės reikšmės (galutinė prognozė) nustatomos skirtumus pridėjus prie esamo laiko momento reikšmės:

$$y_{t+h} = y_t + \Delta y_{t+h} \quad (2.15)$$

Taikant duomenų diferencijavimą, ARMA modelis papildomas integruojančia dalimi I, kuria nurodoma duomenų diferencijavimo eilė d , reikalinga duomenų stacionarumui užtikrinti.

Sezoniškumo įvertinimas

Kai tiriama laiko eilutė pasižymi periodiškai laike pasikartojančiais dėsningumais, papildomai vertinama duomenų sezoninė koreliacija ir sudaromas SARIMA modelis, žymimas:

$$\text{SARIMA}(p,d,q)(P,D,Q)_s$$

čia s – skaitiklis, apibūdinantis sezono trukmę (reikšmėmis);

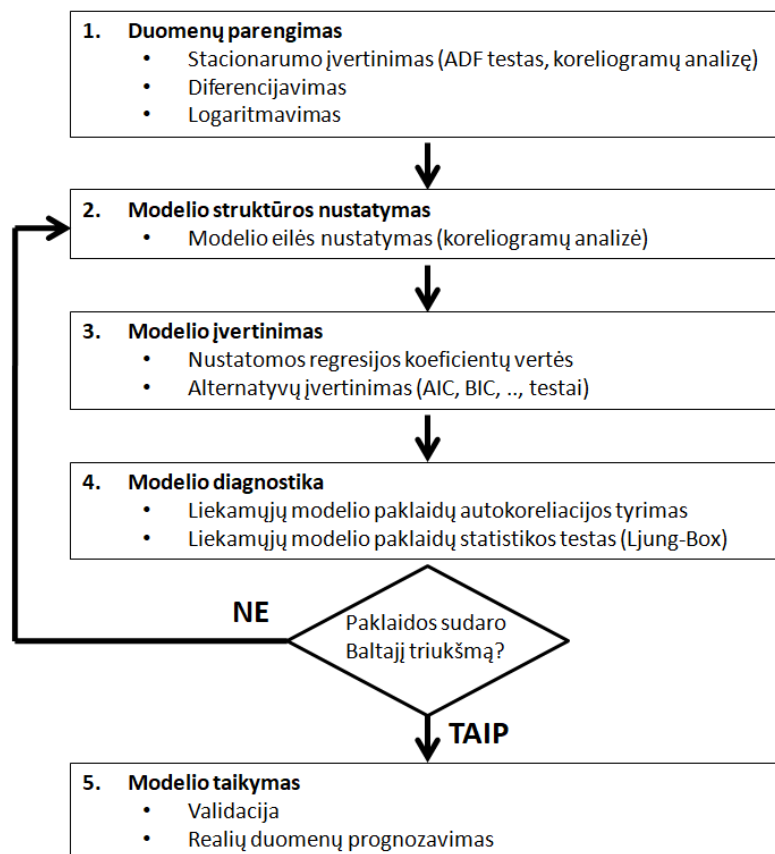
p,d,q – nesezoninės AR(p), MA(q) modelio eilės ir nesezoninio diferencijavimo eilė d ;

P,D,Q – sezoninės SAR(P), SMA(Q) modelio eilės ir sezoninio diferencijavimo eilė D .

Bendroji SARIMA modelio lygtis užrašoma:

$$\begin{aligned} & \underbrace{(1 - \phi_1 L - \dots - \phi_p L^p)}_{\text{Nesezoninė AR}(p) \text{ dalis}} \underbrace{(1 - \phi_1 L^s - \dots - \phi_p L^{ps})}_{\text{Sezoninė AR}(P) \text{ dalis}} \underbrace{(1 - L)^d}_{\substack{d \text{ eilės} \\ \text{(nesezoninis)} \\ \text{diferencijavimas}}} \underbrace{(1 - L^s)^D}_{\substack{D \text{ eilės} \\ \text{sezoninis} \\ \text{diferencijavimas}}} (x_t - \mu) = \quad (2.16) \\ & = \underbrace{(1 - \theta_1 L - \dots - \theta_q L^q)}_{\text{Nesezoninė MA}(q) \text{ dalis}} \underbrace{(1 - \theta_1 L^s - \dots - \theta_p L^{qs})}_{\text{Sezoninė MA}(Q) \text{ dalis}} e_t \end{aligned}$$

SARIMA modelio sudarymo procedūra



2.1 pav. ARIMA modelio sudarymo procedūra (Dong *et al.* 2009)

2.1.2 Duomenų parengimas

Reikalavimai duomenims:

- Duomenys privalo sudaryti laiko eilutę – fiksuoti pastoviu dažniu.
- Laiko eilutė turi būti vientisa, be trūkių.
- Laiko eilutė turi būti stacionari – pastovus vidurkis ir dispersija.

Esant trūkiams duomenyse, trūkstamos reikšmės keičiamos buvusio laiko momento reikšme, gretimų reikšmių vidurkiu ar taikoma interpoliacija, priklausomai nuo trūkių dydžio.

ARIMA modelis gali būti sudarytas tik stacionariai laiko eilutei. Stacionarios laiko eilutės duomenų imties vidurkis ir dispersija yra pastovūs, o paklaidos – atsitiktinės. Duomenų nestacionarumą sukelia: tendencingas kylimas ar leidimasis, sezoniškumas, cikliškumas, didelė dispersija.

Modelis laikomas stacionariu, kai:

- Laiko eilutės vidurkis pastovus:

$$E(y_t) = \mu_t = \text{const.} \quad (2.17)$$

čia $E(y_t)$ – matematinė viltis.

Laiko eilutę padalinus į atskirus intervalus, jų vidurkis privalo būti lygus.

- Laiko eilutės dispersija pastovi:

$$E(x_t - \mu_y)^2 = \sigma_y^2 = \gamma_0 = \text{const.} \quad (2.18)$$

čia γ_0 – kovarijacija neesant poslinkio

- Laiko eilutės stebėjimų kovarijacija nepriklauso nuo laiko, priklauso tik nuo poslinkio:

$$E[(y_t - \mu_y)(y_{t-k} - \mu_y)] = \gamma_k = \text{const.} \quad (2.19)$$

Laiko eilutės stacionarumo nustatymas atliekamas:

- Vaizdiniu būdu, vertinant pradinių duomenų grafiką. Įvertinama ar nėra tendencijų duomenyse, cikliškų svyravimų.
- Vienetinės šaknies testais.
- Koreliogramų analize (autokoreliacijos (ACF) ir dalinės autokoreliacijos (PACF) grafikų vizualinis įvertinimas).
- Vertinant sudaryto modelio liekamųjų paklaidų dispersiją. Stacionaraus modelio liekamosios paklaidos atitinka normalųjį skirstinį. Naudojami statistiniai testai, tokie kaip Ljung–Box.

Vienetinės šaknies testai

Duomenų imtis laikoma nestacionare jei egzistuoja vienetinė šaknis – laiko eilutės reikšmės visiškai priklauso nuo ankstesnių vėlavimų, nėra atsitiktinės paklaidos. Duomenų stacionarumo patikrai naudojami vienetinės šaknies testai, pavyzdžiui Augmented Dickey–Fuller (ADF).

Taikant ADF testą formuojama nulinė hipotezė, teigianti apie laiko eilutėje egzistuojančią vienetinę šaknį. Alternatyvi hipotezė – vienetinė šaknis neegzistuoja. Modelis laikomas stacionariu, jei nulinė hipotezė yra atmetama:

$$H_0 : b = 0;$$

$$H_1 : b < 0.$$

Testo metu vertinama priklausomybė:

$$\Delta y_t = (\rho_1 - 1)y_{t-1} + \sum_{i=1}^p \rho_i \Delta y_{t-i+1} + e_t \quad (2.20)$$

čia ρ_i – skirtumo operatorius;
 e_t – modelio paklaida.

Koreliogramų analizė

Sudaromos autokoreliacijos ir dalinės autokoreliacijos funkcijos.

- **Autokoreliacijos funkcija (ACF):** stacionaraus proceso autokoreliacijos funkcijos reikšmės gėsta greitai ir yra artimos nuliui, išskyrus pirmąjį vėlavimą, kuris yra lygus vienetui. Kai procesas nestacionarus, autokoreliacijos funkcijos reikšmės mažėja lėtai ir tiesiškai. Autokoreliacijos reikšmės nustatomos pagal (2.21) formulę:

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^T (y_t - \bar{y})^2} \quad (2.21)$$

čia y_t – kintamojo reikšmė laiko momentu t ;
 k – k -ojo vėlavimo autokoreliacijos koeficientas;
 $\bar{y}$ – kintamojo vidutinė reikšmė.

- **Dalinės autokoreliacijos funkcija (PACF):** funkcija nustato tiesioginį konkretaus vėlavimo y_{t-k} poveikį esamam laiko momentui y_t , atmetus kitus vėlavimus. Dalinės autokoreliacijos analizė atliekama pagal (2.22) formulę:

$$r_{kk} = \begin{cases} r_1, & \text{kai } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \end{cases} \quad (2.22)$$

čia k – k -ojo vėlavimo eilė;
 j – vertinamų vėlavimų skaičius;
 r_k – autokoreliacijos koeficientas.

Visos stacionaraus proceso dalinės autokoreliacijos reikšmės yra artimos nuliui (nekertamas reikšmingumo lygmuo). Reikšmingumo lygmuo atitinka baltojo triukšmo autokoreliacijos ($\gamma = 0,95$) pasikliautinąjį intervalą:

$$\pm 1,95/\sqrt{n} \quad (2.23)$$

Laiko eilutės diferencijavimas

Laiko eilutės diferencijavimas, dar vadinamas skirtumine transformacija, atliekamas pradinių duomenų laiko eilutę transformuojant į skirtumų laiko eilutę. Atliekant skirtuminę transformaciją nustatoma laiko eilutės skirtumų sklaida apie 0 reikšmę:

$$\Delta^d y_t \sim I(0) \quad (2.24)$$

Diferencijavimo eilė nustatoma:

$$\Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1} \quad (2.25)$$

Jei diferencijuoti duomenys atitinka stacionarumo sąlygas, naudojamas integruojantis ARIMA modelis.

2.1.3 Modelio struktūros nustatymas

ARMA modelio $AR(p)$ ir $MA(q)$ parametrų eilės parenkamos pagal nustatytus koreliuotus vėlavimus autokoreliacijos bei dalinės autokoreliacijos funkcijose. Nustačius koreliuotų vėlavimų skaičių, parametrais p ir q nurodoma iki kurio vėlavimo yra koreliacija.

Koreliogramos vertinamos pagal šiuos principus:

- Jei autokoreliacijos funkcija gęsta lėtai, o dalinės autokoreliacijos funkcija nutrūksta po p eilės vėlavimo (nebekerta pasikliautinojo intervalo ribos), naudojamas autoregresinis $AR(p)$ modelis. Taikomas $ARIMA(p,d,0)$ modelis.
- Jei dalinės autokoreliacijos funkcija gęsta lėtai, o autokoreliacijos funkcija nutrūksta po q eilės vėlavimo, naudojamas slankiųjų vidurkių $MA(q)$ modelis. Taikomas $ARIMA(0,d,q)$ modelis.
- Jei autokoreliacijos ir dalinės autokoreliacijos funkcijos nutrūksta greitai, po panašios vėlavimų eilės, pagal dalinės autokoreliacijos vėlavimo eilę p nustatoma $AR(p)$ modelio eilė, o pagal autokoreliacijos vėlavimo eilę q nustatoma $MA(q)$ modelio eilė. Taikomas $ARIMA(p,d,q)$ modelis.

2.1.4 Modelio įvertinimas

Modelio įvertinimo etape pagal sudarytą regresijos lygtį ir remiantis didžiausia tikimybe, yra nustatomi regresijos koeficientų įverčiai. Patariama sudaryti ne vieną, o keletą artimų modelių ir atlikti jų palyginimą. Palyginimo metu apie modelių kokybę sprendžiama informacinių kriterijų testų rezultatais. Šie informaciniai kriterijai įvertina modelio bei jo struktūros tinkamumą duomenims. Tinkamiausiu laikomas tas modelis, kurio paklaidos yra mažiausios ir sudaryta paprasčiausia struktūra.

Naudojami informaciniai kriterijai:

- **Akaike informacinis kriterijus**

$$AIC(p, q) = \ln \hat{\sigma}_e^2 + \frac{2}{T} (p + q) \quad (2.26)$$

čia $\hat{\sigma}_e^2$ – modelio paklaidų dispersiją apibūdinanti funkcija;
 T – reikšmių skaičius imtyje;
 p, q – sudaryto modelio (AR(p), MA(q)) eilės.

- **Bajeso informacinis kriterijus**

$$BIC(p, q) = \ln \hat{\sigma}_e^2 + \frac{\ln T}{T} (p + q) \quad (2.27)$$

2.1.5 Modelio diagnostika

Tinkamai sudaryto modelio paklaidos turi būti atsitiktinio pobūdžio, atitikti baltąjį triukšmą. Tokių paklaidų vidurkis turi svyruoti apie nulio reikšmę ir bet kuriuo laiko momentu negali būti fiksuojama liekamųjų paklaidų koreliacija. Nustačius liekamųjų paklaidų koreliaciją, reikia keisti modelio struktūrą.

Paklaidos turi tenkinti šias baltojo triukšmo savybes:

- **Matematinė viltis** $\mu_e = E(e_t) = 0$;
- **Dispersija** $\gamma_0 = E(e_t^2) = \sigma_e^2$;
- **Kovariacija** $\gamma_k = 0$.

Modelio liekamosios paklaidos tikrinamos vizualiai vertinant autokoreliacijos funkciją bei Ljung–Box testu.

- Liekamųjų paklaidų autokoreliacijos koeficientai apskaičiuojami pagal formulę:

$$\rho_i = \frac{\text{cov}(e_t \cdot e_{t-1})}{\sqrt{\text{var } e_t} \sqrt{\text{var } e_{t-1}}}, i = 1, 2, \dots, p \quad (2.28)$$

čia e_t – modelio liekamosios paklaidos laiko momentu t .

- Ljung–Box testo metu tikrinama H_0 (nulinė) hipotezė, teigianti apie nereikšmingą liekamųjų paklaidų autokoreliaciją:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$$

Atmetus nulinę hipotezę, liekamųjų modelio paklaidų autokoreliacija yra reikšminga:

$$H_a: \rho_j \neq 0, j \in 1, \dots, m$$

Tikrinama autokoreliacijos koeficientų statistika:

$$Q = n(n+2) \sum_{i=1}^p \frac{\rho_i^2}{n-1} \quad (2.29)$$

čia ρ_i – liekamųjų paklaidų autokoreliacijos koeficientas;

n – konkretaus vėlavimo numeris;

p – tiriamų vėlavimų kiekis.

Kai Q statistikos reikšmė yra didesnė už kritinę skirstinio χ^2 reikšmę χ_a^2 , tai su α reikšmingumo lygmeniu nulinė hipotezė atmetama:

$$Q = \sim^{H_0} \chi^2(p) \quad (2.30)$$

2.1.6 Modelio validacija

Modelio validacija atliekama tikrinant sudaryto prognozės modelio tikslumą. Prognozavimo metu sudaromos vieno žingsnio, atitinkančio laiko intervalą tarp duomenų matavimų, prognozės. Nustatomas prognozuojamų reikšmių skirtumas nuo realių duomenų. Remiantis atlikta literatūros apžvalga, Smith *et. al* (2002), Kumar, Vanajakshi (2015) paklaidų palyginimui siūlo vertinti MAPE paklaidą. Naudojant Lewi's skalę (1.1 lentelė) prognozės, kurių MAPE reikšmė mažesnė už 10%, gali būti vertinamos kaip labai tikslios, 11–20% – tikslios (Kumar, Vanajakshi 2015). Siekiama sudaryti modelį, kurio MAPE paklaida būtų mažesnė nei 20%. Taip pat skaičiuojamos MAE ir RMSE paklaidos.

Vertinamos paklaidos:

- Vidutinė absoliutinė paklaida (angl. *Mean Absolute Error*, MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i^r| \quad (2.31)$$

čia $\hat{y}_i$ – prognozuota reikšmė;
 y_i^r – realūs duomenys;
 n – duomenų imties dydis.

- Vidutinė procentinė absoliutinė paklaida (angl. *Mean Absolute Percentage Error*, MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i^r|}{y_i^r} \cdot 100\% \quad (2.32)$$

- Vidutinė kvadratinė paklaida (angl. *Root Mean Square Error*, RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i^r)^2} \quad (2.33)$$

2.2 LSTM modelis

LSTM modelio tinklą sudaro į grandinę sujungtos ląstelės (celės), kurios atlieka duomenų saugojimo bei perdavimo funkciją (2.2 pav.). Visas ląsteles jungia bendras ryšio kanalas C_t , kuriuo vykdomi būsenos duomenų mainai. Šis kanalas atlieka tinklo atminties funkciją, kadangi ląstelės, skaičiuodamos naują prognozės reikšmę, remiasi ne tik svorių matricomis ir naujai įvedamomis kintamojo reikšmėmis, o ir ankstesne tinklo būsena, perduodama šiuo kanalu.

Ląstelės viduje yra ryšio kanalai, kurie valdomi vartais. Šie vartai – tai ląstelės viduje esantys paslėptieji neuroninių tinklų sluoksniai. Jais įvertinami duomenys, atliekami duomenų svorio skaičiavimai. Informacijos valdymas yra atliekamas taikant aktyvacijos funkcijas. Šiuo atveju nagrinėjamas bazinis tinklo atvejis, kuriame naudojama sigmoidinė bei hiperbolinio tangento aktyvacijos funkcijos.

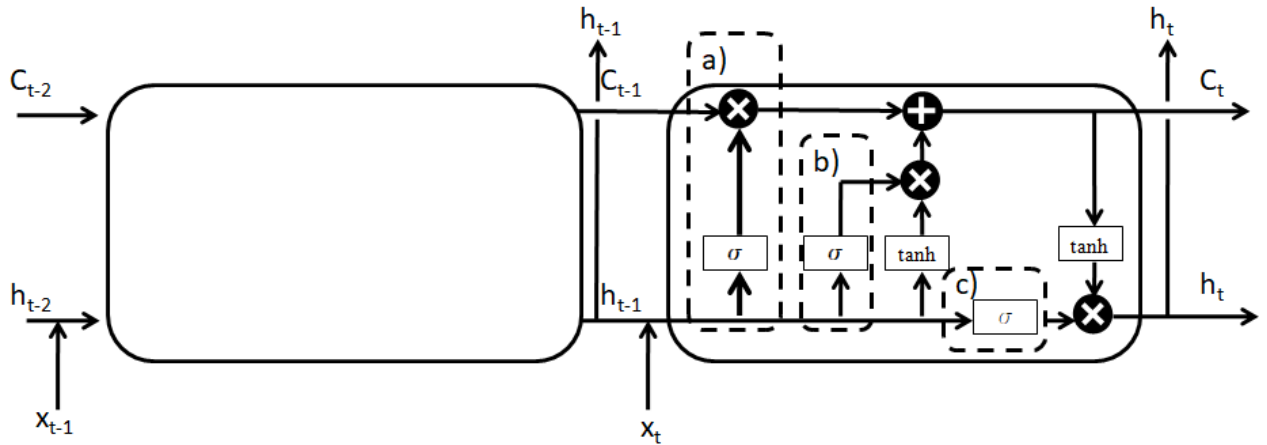
- Sigmoidinės aktyvacijos funkcijos (2.34 formulė) reikšmės gali kisti $[0;1]$ intervale. Jei norima perduoti visą informaciją, aktyvacijos funkcijos reikšmė yra artima arba lygi 1, jei norima duomenis pamiršti, išvedama 0 reikšmė:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.34)$$

čia x – įvesties kintamasis;
 e – eksponentė.

- Hiperbolinio tangento funkcija gali įvertinti ir neigiamą poveikį (2.35 formulė). Šia funkcija duomenys yra apribojami $[-1; 1]$ reikšmių intervale:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.35)$$



2.2 pav. LSTM modelio ląstelės struktūra (Yuan *et al.* 2019): a) užmaršties vartai; b) įvesties vartai; c) išvesties vartai

Informacijos kelias ląstelėje (Yuan *et al.* 2019):

1. Užmaršties vartai įvertina, kokia informacijos dalis, saugoma tinklo atmintyje, turi būti pamiršta. Tai yra paslėptasis tinklo sluoksnis, kuris vertina praeities duomenų C_{t-1} aktualumą, naudodamas buvusios ląstelės išvesties reikšmę h_{t-1} bei esamo laiko momento kintamojo reikšmę x_t . Pagal šių reikšmių vertes, aktyvacijos funkcija įvertina, kokia praeities duomenų dalis bus pamiršta. Taikant sigmoidinę aktyvacijos funkciją (2.34 formulė) yra nustatomas daugiklis, iš kurio bus padauginta praeities tinklo būseną. Jei daugiklis yra artimas nuliui, didžioji praeities duomenų dalis bus pamiršta. Jei daugiklis artimas vienetui, praeities duomenys bus toliau saugomi atmintyje. Šio etapo rezultate užmaršties vartai sudaro funkciją (2.36 formulė):

$$f_t = \sigma(x)(W_f[h_{t-1}, x_t] + b_f) \quad (2.36)$$

čia W_f – įvesties duomenų svorio matrica;
 h_{t-1} – ankstesnė ląstelės išvesties reikšmė;
 x_t – įvesties reikšmė;
 b_f – svorio koeficientų poslinkis.

2. Įvesties vartai įvertina, kokia naujai įvedamos informacijos dalis turi būti saugoma atmintyje. Naudojama sigmoidinė funkcija nustato daugiklį, įvertinantį naujai įvestos informacijos svarbą (2.37 formulė), o hiperbolinio tangento sluoksnis atlieka įvesties duomenų transformaciją, parengdamas informaciją saugojimui (2.38 formulė).

$$i_t = \sigma(x)(W_i[h_{t-1}, x_t] + b_i) \quad (2.37)$$

čia W_i – įvesties duomenų svorio matrica;
 b_i – svorio koeficientų poslinkis.

$$C_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (2.38)$$

čia W_c – atmintyje saugojamų duomenų svorio matrica;
 b_c – svorio koeficientų poslinkis.

Tinklo būseną (atmintis) yra atnaujinama įvertinant praeities duomenis (sudauginant užmaršties vartų reikšmę su buvusią sistemos būseną) bei papildant naujo kintamojo reikšmės sukeltu poveikiu (įvesties vartų reikšmė sudauginama su transformuotais įvesties duomenimis) (2.39 formulė).

$$C_t = f_t C_{t-1} + i_t C_t \quad (2.39)$$

3. Išvesties vartai nustato išvedamo kintamojo vertę. Taikant sigmoidinę funkciją, nustatoma kokia įvesto kintamojo dalis turi būti išvedama (2.40 formulė), o hiperbolinio tangento sluoksnis nustato vertę, apibūdinančią esamą tinklo būseną (remiantis ilgalaikėmis duomenų tendencijomis). Galutinė reikšmė, kurią išveda šis sluoksnis yra nustatoma sudauginus išvesties vartų nustatytą vertę, su tinklo atmintyje saugoma būseną (2.41 formulė).

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2.40)$$

$$h_t = O_t \tanh C_t \quad (2.41)$$

Tinklo sudarymas ir mokymas

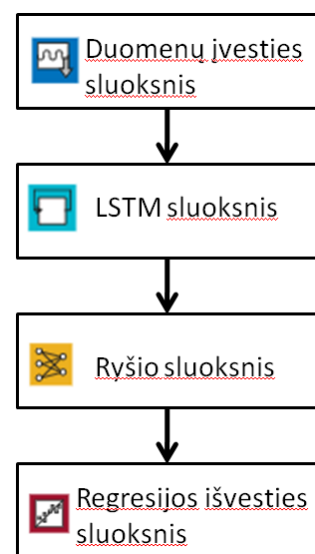
LSTM tinklas buvo sudaromas lyginant pradinę duomenų imtį su viena eile perstumtomis duomenų reikšmėmis. Tinklas, atlikdamas šių reikšmių palyginimą, nustato įvedamo kintamojo poveikį rezultatui. Vėliau šis poveikis optimizuojamas. Optimizavimas atliekamas mažinant RMSE paklaidą. Siekiant išvengti duomenų neįvertinimo ar pervertinimo dėl duomenų sklaidos, turimos laiko eilutės reikšmės yra normalizuojamos, t. y. perskaičiuojamos į [-1; 1] intervalą. Prognozė gaunama atlikus atvirkštinę duomenų transformaciją.

Duomenų įvesties sluoksnis esamo laiko momento reikšmės y_t perduoda į LSTM sluoksnį (2.3 pav.). LSTM sluoksnyje atliekamas duomenų regresijos įvertinimas ir šis sluoksnis išveda h_t reikšmę. Ryšio sluoksniu ši reikšmė perduodama į regresijos išvesties sluoksnį, kuriame h_t reikšmė yra konvertuojama į prognozuojamą $\hat{y}_{t+1}$. Atliekama vieno žingsnio prognozė. Po kiekvieno žingsnio prognozės, tinklo būseną yra atnaujinama realia reikšme (įvesties duomenimis).

LSTM tinklas sudaromas naudojant MATLAB *Machine learning* programinę įrangą bei MATLAB dokumentacijoje pateiktus tinklo sudarymo pavyzdžius. Naudotos vertės pateiktos (2.2 lentelėje).

2.2 lentelė. LSTM tinklo parametrai

Paslėptų sluoksnių skaičius	200
Naudojami sluoksniai	• Įvesties • LSTM • Išvesties • Regresijos
Optimizacijos algoritmas (angl. <i>solver</i>)	adam
Mokymosi trukmė	250 epochų
Ląstelės sužadinimo slenkstis (angl. <i>Gradient threshold</i>)	1
Mokymosi santykis (angl. <i>initial learning rate</i>)	0,005 (po 125 epochų didinamas į 0,2)



2.3 pav. Sudaryto LSTM modelio struktūra

2.3 Glodinimo metodai

Glodinimas apima statistinių filtrų taikymą, siekiant sumažinti aukšto dažnio svyravimus duomenyse. Taikant glodinimą, kiekviena laiko eilutės reikšmė yra pakeičiama greta esančių reikšmių vidurkiu arba mediana. Glodinimui naudojamų reikšmių skaičius nurodomas glodinimo pločiu.

Šiame darbe detaliau nagrinėjams Gauso filtras bei slankiosios medianos glodinimo metodai.

2.3.1 Gauso filtras

Gauso filtras remiasi slankiojo vidurkio metodo taikymu, papildomai naudojant Gauso funkcija nustatytus svertinius koeficientus. Glodinamą plotį atitinka tam tikras skaičius reikšmių, esančių šalia glodinamos reikšmės. Slankiojo vidurkio metodas nustato glodinamo pločio reikšmių vidurkį ir judėdamas duomenų eilute pakeičia kiekvieną imties narį perskaičiuota reikšme. Siekiant įvertinti kiekvienos reikšmės svarbą, naudojami svertiniai koeficientai. Glodinama reikšmė apskaičiuojama pagal (2.42) formulę:

$$y_g = \sum_{j=-k}^k w_j y_{i+j} \quad (2.42)$$

čia y_g – glodinta reikšmė;
 k – glodinimo plotis;
 w – svertinis koeficientas;
 y_{i+j} – į glodinimo plotį patenkanti reikšmė.

Gauso filtro atveju, slankiojo vidurkio metodo svertiniai koeficientai nustatomi remiantis reikšmių nuokrypiais, įvertintais Gauso funkcija (2.44 formulė):

$$f(y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2\right)} \quad (2.43)$$

čia σ – vidutinis nuokrypis;
 e – eksponentinė funkcija.

Vidutinis kvadratinis nuokrypis σ apriboja laiko eilutės nuokrypius nagrinėjamame glodinimo ploteje ir laiko eilutės vertės, kurios viršija nustatytą ribą, įgauna svertinius koeficientus, lygius 0. Taip atsitiktinius nuokrypius turinčios reikšmės pašalinamos iš slankiojo vidurkio imties.

2.3.2 Slankiosios medianos metodas

Taikant šį metodą nustatoma glodinamo pločio imties mediana. Glodinimo pločio imčiai slenkant laiko eilute, į šią imtį patenkančios reikšmės yra išdėstomos variacine eilute. Glodinama reikšmė pakeičiama variacinės eilutės viduriniu nariu (esant nelyginiam glodinamo pločio narių kiekiui) arba viduriniųjų reikšmių vidurkiu (kai glodinamą plotį apima lyginis narių skaičius).

3. EISMO SRAUTŲ PROGNOZĖS TYRIMAS

3.1 Prognozės modelio sudarymas

Uždavinys: Sudaryti veikiantį prognozės modelį, kurio prognozės vidutinė absoliutinė procentinė paklaida (MAPE) būtų mažesnė kaip 20%.

Metodika: Sudaromas parametrinis SARIMA modelis, pagal 2.1 skyriuje nurodytą metodiką. Prognozės modelio kokybė įvertinama nustatant prognozės paklaidas, lyginant prognozuojamas reikšmes su validacijai skirtais realiais duomenimis. SARIMA modelio paklaidos lyginamos su neparimetrinio LSTM modelio paklaidomis.

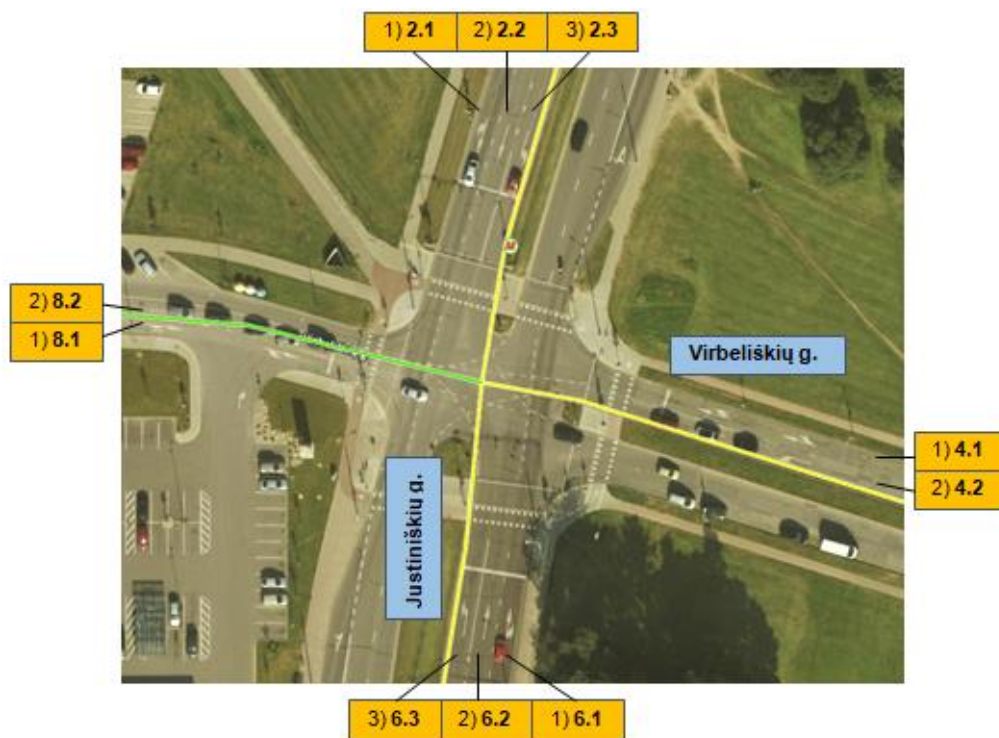
3.1.1 Pradiniai duomenys

Pradiniai duomenys gauti iš savivaldybės įmonės *Susisiekimo paslaugos*, kuri organizuoja eismo valdymą Vilniaus mieste. Duomenys surinkti keturšalėje Justiniškių–Virbeliškių gatvių sankryžoje, 2018 m. rugsėjo–spalio mėn., naudojant vaizdo jutiklius (vaizdas apdorojamas programine įranga). Sankryža yra vakarinėje miesto dalyje, Justiniškių miegamajame mikrorajone. Sankryža nepasižymi dideliu tranzitiniu eismu, yra mažo intensyvumo. Šalia sankryžos – gyvenamieji namai, parduotuvė. Netoli sankryžos prasideda Justiniškių g. Virbeliškių g. jungia Medeinos bei Budiniškių gatves. Justiniškių g. abiem kryptimis išplatėja į 3 eismo juostas, Virbeliškių – po 2 (3.1 pav.). Kiekviena eismo juosta prieš sankryžą turi po jutiklį. Jutiklių aprašas pateiktas 3.1 lentelėje.

Justiniškių g. automobilių srautas abiem kryptimis yra daug didesnis nei Virbeliškių g. (duomenys pateikti remiantis *Susisiekimo paslaugų* informacija 2019 m. spalio mėn.):

- Vidutinis Justiniškių g. paros intensyvumas (link Laisvės pr.) – 6360 aut./parą. Priešinga kryptimi (nuo Laisvės pr.) – 5560 aut./parą.
- Virbeliškių g. (nuo Budiniškių g.) paros intensyvumas – 2320 aut./parą. Priešinga kryptimi (nuo Medeinos g.) – 2820 aut./parą.

Pradiniai duomenys fiksuoti 90 s dažniu, matuojant vidutinį valandos intensyvumą. Iš viso fiksuota 58560 matavimų. Šios reikšmės perskaičiuotos į faktinį 90 s intensyvumą. Duomenyse pastebėta 16 neatitikimų: praleisti ar netinkamu periodu fiksuoti matavimai. Praleisti matavimai buvo pakeisti prieš tai buvusiomis reikšmėmis. Netinkamo periodiškumo matavimams atkurta laiko žyma. Kadangi iš 58560 matavimų tik 16 neatitiko (0,027% imties), tai neturėtų sudaryti reikšmingo skirtumo sudarant prognozės modelį.



3.1 pav. Justiniškių–Virbeliškių g. orofoto (Vilniaus miesto vairuotojų informacinė sistema 2020)

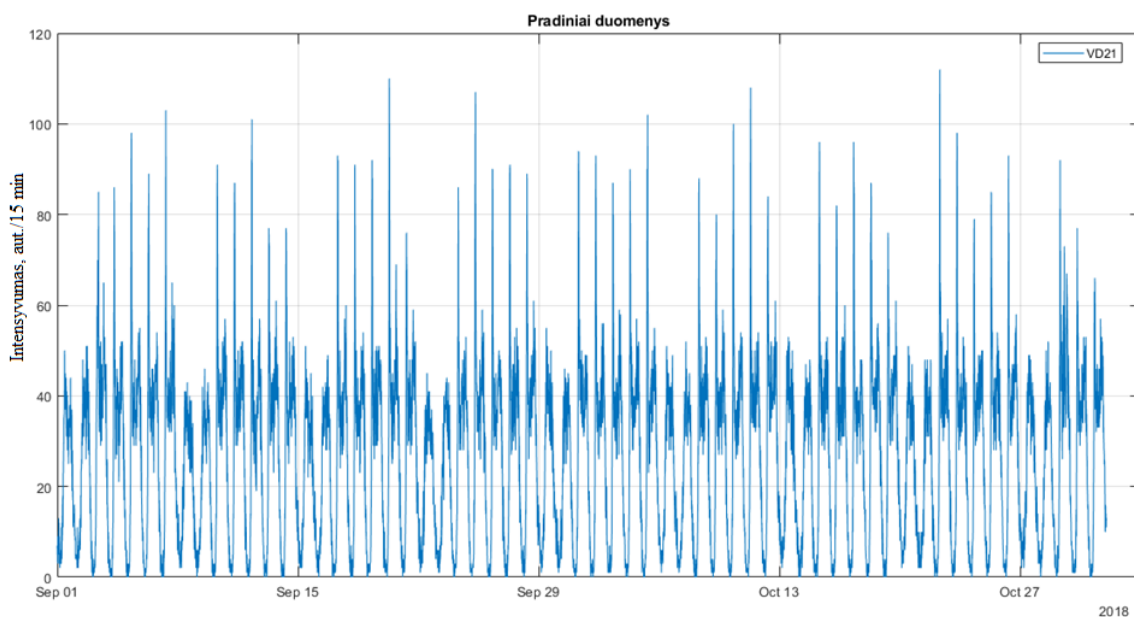
3.1 lentelė. Justiniškių–Virbeliškių g. sankryžoje naudojami eismo srautų jutikliai

Eil. Nr.	Žymuo paveiksle	Jutiklio pavadinimas	Gatvė	Eismo juosta	Eismo juostos kryptis
1	1) 2.1	VD2.1	Justiniškių g. (link Laisvės pr.)	1	Tiesiai arba į dešinę
2	2) 2.2	VD2.2		2	Tiesiai
3	3) 2.3	VD2.3		3	Į kairę
4	1) 4.1	VD4.1	Virbeliškių g. (nuo Medėinos g.)	1	Tiesiai arba į dešinę
5	2) 4.2	VD4.2		2	Į kairę
6	1) 6.1	VD6.1	Justiniškių g. (nuo Laisvės pr.)	1	Tiesiai arba į dešinę
7	2) 6.2	VD6.2		2	Tiesiai
8	3) 6.3	VD6.3		3	Į kairę
9	1) 8.1	VD8.1	Virbeliškių g. (nuo Budiniškių g.)	1	Tiesiai arba į dešinę
10	2) 8.2	VD8.2		2	Į kairę

Prognozės modelis sudaromas naudojant Justiniškių g. (link Laisvės pr.) pirmos eismo juostos intensyvumą fiksuojančio jutiklio VD2.1 duomenis. Duomenys, surinkti nuo 2018-09-01 iki 2018-10-23, naudojami prognozės modeliui sudaryti, o nuo 2018-10-24 iki 2018-10-31 – prognozės modeliui validuoti. Siekiant sumažinti kompiuterinius resursus, reikalingus modeliui sudaryti ir įvertinti, keičiamas duomenų fiksavimo dažnis ir pradiniai duomenys yra agreguojami 900 s (15 min) intervalais. Taip sumažinamas vieno jutiklio užfiksuotų reikšmių kiekis nuo 58560 iki 5856 (3.2 lentelė).

3.2 lentelė. Modeliui sudaryti naudojami pradiniai duomenys

Vieta	Justiniškių–Virbeliškių g. sankryža, Vilnius
Duomenų laikotarpis	2018-09-01 – 2018-10-31
Matuojamas parametras	Intensyvumas, aut./15 min.
Matavimo dažnis (vieno intervalo trukmė)	900 sekundžių (15 minučių)
Užfiksuota reikšmių	5856
Parą atitinkantis reikšmių kiekis	96
Duomenų imtis modeliui sudaryti	2018-09-01 – 2018-10-23
Reikšmių kiekis modeliui sudaryti	5088
Duomenų imtis modeliui validuoti	2018-10-24 – 2018-10-31
Reikšmių kiekis modeliui validuoti	768

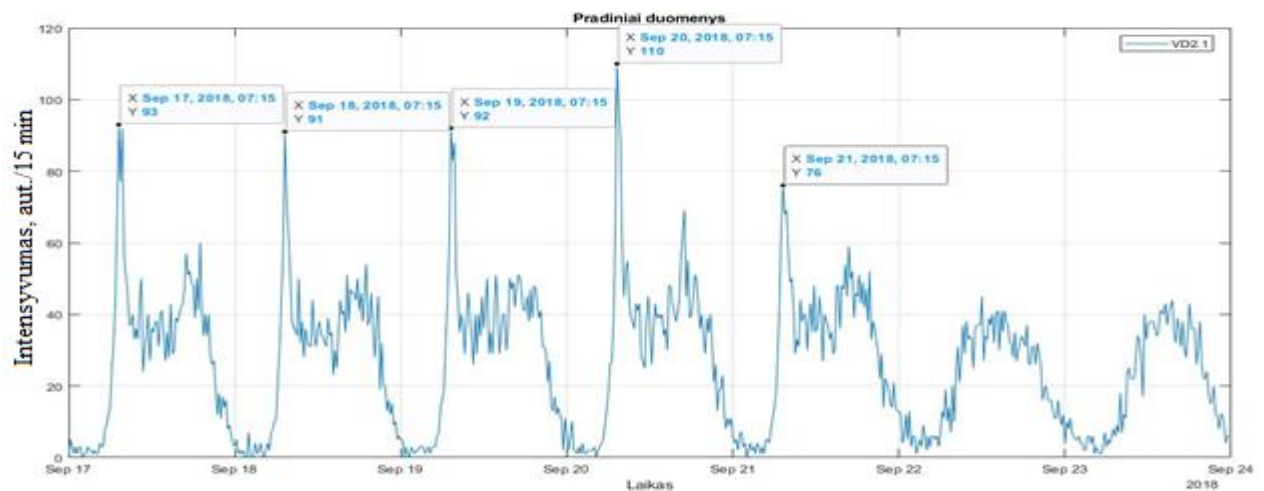


3.2 pav. Pradiniai Justiniškių g. (link Laisvės pr.) pirmos eismo juostos jutiklio VD2.1 duomenys

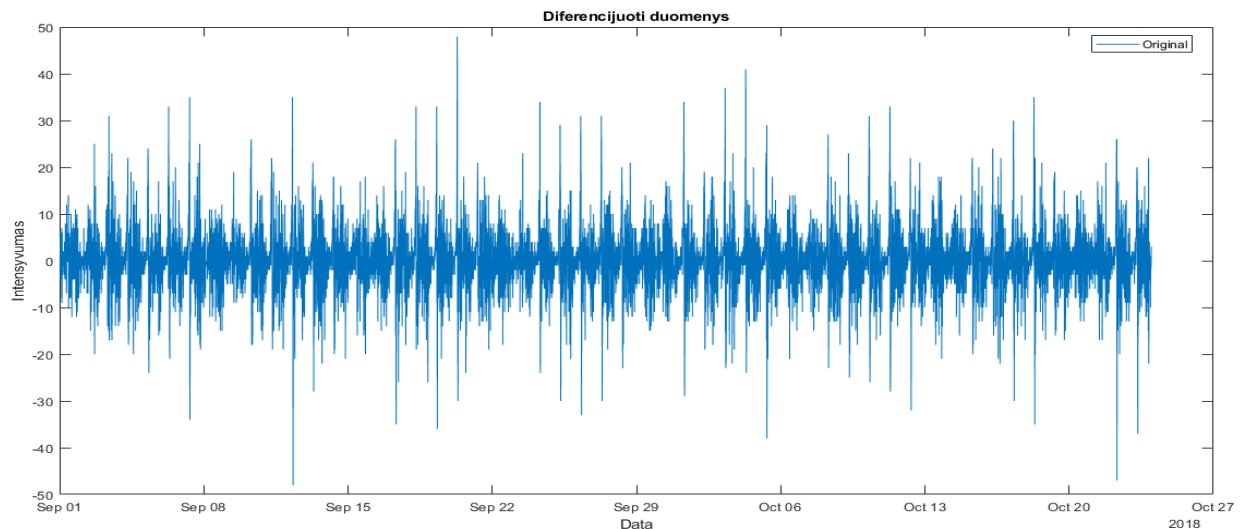
Vizualiai vertinant pradinių duomenų laiko eilutę (3.2 pav.), pastebimi savaitiniai bei paros dėsningumai. Rytinio piko metu eismo srautas būna didžiausias, siekia 90–110 aut./15 min. Piko metu kreivė kyla greitai, didžiausia reikšmė pasiekama per 1–2 laiko intervalus. Žvelgiant į konkrečios savaitės duomenis (3.3 pav.) pastebima, kad didžiausios piko reikšmės buvo pasiektos tuo pačiu laiko momentu – 7 val. 15 min. Tai rodo nusistovėjusius eismo srautus dėl kasdienių (rutininių) kelionių. Vėliau kreivė greitai pradeda žemėti ir apie 9 valandą ryto eismo srautas stabilizuojasi, intensyvumas siekia 35–40 aut./15 min. reikšmę. Vakarinio piko metu, apie 16 valandą, pastebimas nedidelis eismo intensyvumo padidėjimas, tačiau reikšmės ne tokios ekstremalios kaip rytinio piko metu, siekia 50–60 aut./15 min. Vėliau eismo intensyvumas mažėja lėtai ir tolygiai, po vidurnakčio pasiekama minimali reikšmė, artima nuliui. Duomenų kreivė rodo stabilią būseną iki 6–7 valandos ryto, kai intensyvumo reikšmė staiga pradeda didėti.

Vertinant vienos savaitės duomenis, pastebimi dažni (cikliški) svyravimai tarp gretimų matavimų. Atliekamas pirmos eilės laiko eilutės diferencijavimas (3.4 pav.), siekiant įvertinti

intensyvumo skirtumus tarp gretimų matavimų. Diferencijuotoje laiko eilutėje duomenys išlaiko periodiškai pasikartojančius dėsningumus. Intensyvumo verčių skirtumai tarp gretimų matavimų yra dideli, daugeliu atvejų svyruoja apie 10–20 reikšmių, pastebimi ir didelės vertės (>30) šuoliai. Jų skaičius nėra didelis (vertinant, kad imtį sudaro virš 5800 reikšmių), tačiau iš ankstesnių grafikų yra žinoma, kad vidutiniškai intensyvumo reikšmė svyruoja apie 30–40 aut./15 min, o didžiausias intensyvumas siekia apie 110 aut./15 min. Tokie dideli šuoliai tarp matavimų gali iškreipti prognozės modelio rezultatus arba dėl didelės duomenų dispersijos nebus galima parinkti tinkamos modelio struktūros.



3.3 pav. Pradiniai VD2.1 jutiklio duomenys. Savaitės imtis (2018-09-17 – 2018-09-24)



3.4 pav. Diferencijuotos laiko eilutės duomenys

3.1.2 Duomenų stacionarumas

ARIMA prognozės modelis gali būti sudarytas tik stacionariai laiko eilutei. Duomenų stacionarumo apibrėžimas bei metodai stacionarumui užtikrinti pateikti 2.1.2 skyrelyje.

Pirmiausiai duomenų stacionarumas vertinamas Augmented Dickey–Fuller (ADF) testu. Esant stacionariams duomenis, testo H_0 hipotezė, teigianti apie esančią vienetinę šaknį (nestacionarius duomenis dėl absoliučios regresijos su ankstesnėmis reikšmėmis) – atmetama ir duomenis galima tirti toliau. Tada atliekama koreliogramų analizė (3.5 ir 3.6 pav.)

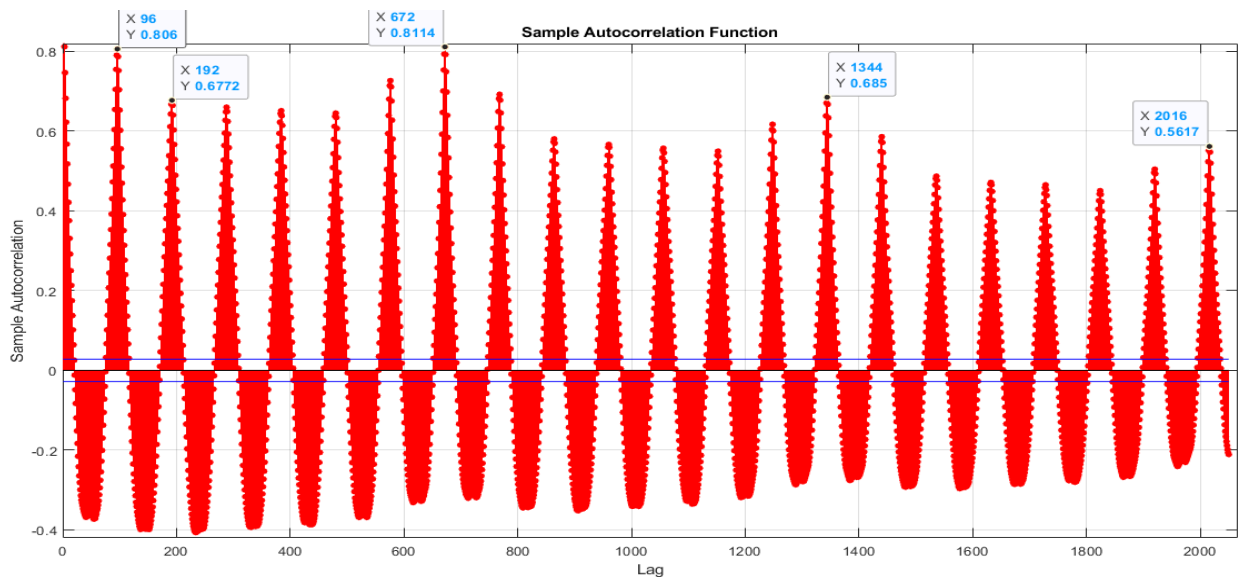
ADF testo statistika yra mažesnė už kritinę reikšmę, todėl nulinė hipotezė, reiškianti duomenų nestacionarumą – atmesta (3.3 lentelė). Testo rezultatai negarantuoja, kad duomenys yra stacionarūs, todėl reikalinga atlikti koreliogramų analizę.

3.3 lentelė. Pradinių duomenų ADF testo reikšmė

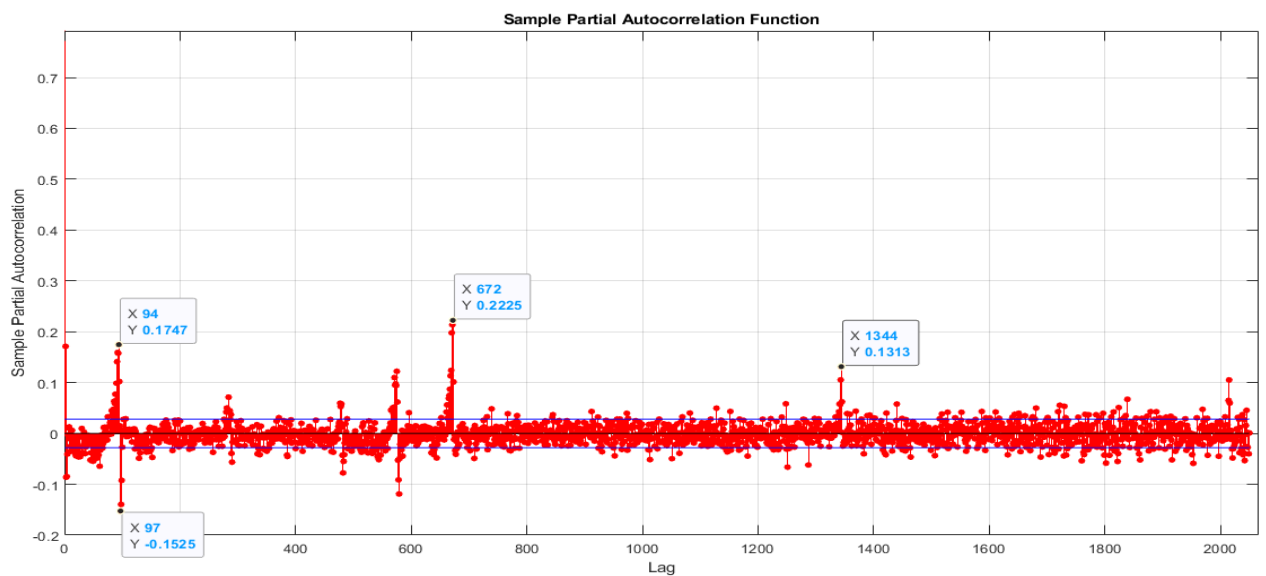
Testo kritinė reikšmė c	-1,9416
Testo statistika	-8,7876
H_0 hipotezė	Atmesta

Pradinių duomenų autokoreliacijos funkcija (3.5 pav.) pasižymi sinusoidės forma pasikartojančiais svyravimais, todėl negalima išskirti konkrečių laiko eilutės vėlavimų, svarbių esamam laiko momentui. Dėl šios priežasties negalima nustatyti ARIMA modelio parametrų eilės (detaliau apie ARIMA modelio eilės nustatymą žr. 2.1.3 skyrelyje).

Vieną svyravimų periodą atitinka 96 reikšmės, t. y. 1 paros duomenys. Pastebimas svyravimų padidėjimas ties 672 reikšme bei jos kartotiniais (1344, 2016). Šis duomenų kiekis atitinka vieną savaitę, tai rodo stiprų savaitės periodo duomenų ryšį. Siekiant užtikrinti laiko eilutės stacionarumą turi būti atliktas sezoninis duomenų diferencijavimas bei naudojamas sezoninis ARIMA modelis – $SARIMA(p,d,q)(P,D,Q)_s$, kurio vieną sezoną atitinka 672 reikšmės (viena savaitė).



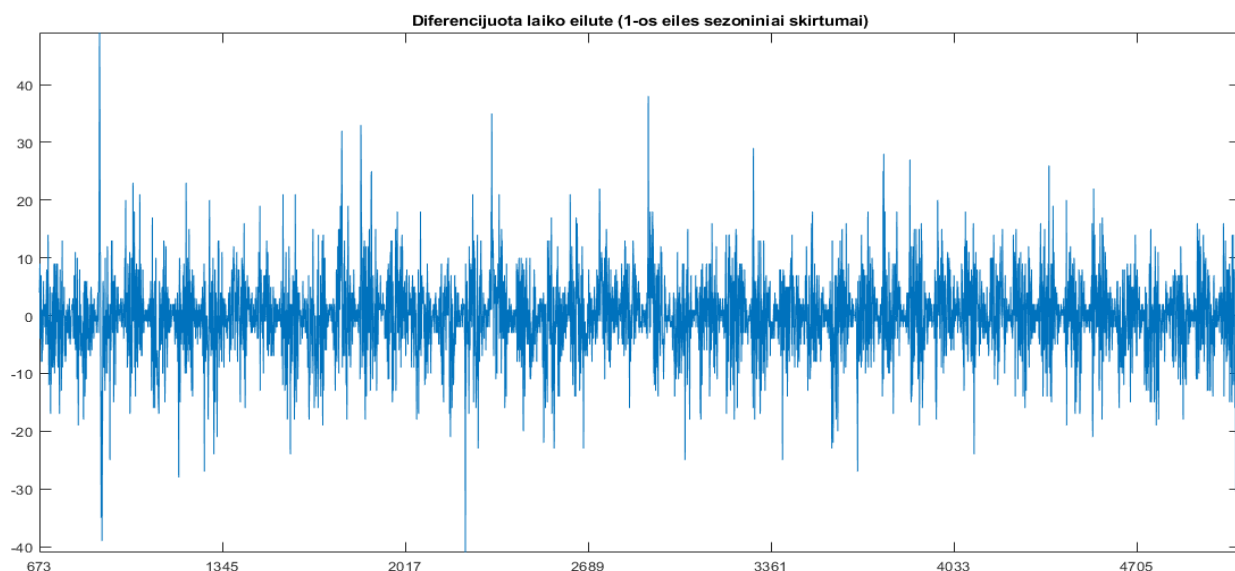
3.5 pav. Pradinių duomenų autokoreliacijos funkcija



3.6 pav. Pradinių duomenų dalinės autokoreliacijos funkcija

Laiko eilutės diferencijavimas

Duomenų diferencijavimas atliekamas pagal 2.1.2 skyrelyje aptartą metodiką. Atlikus duomenų diferencijavimą tiriamos ne tiesiogiai išmatuotos laiko eilutės reikšmės, o jų skirtumai. Kadangi atliekant pradinių duomenų koreliogramų analizę pastebėti periodiškai pasikartojantys laiko eilutės dėsningumai, pirmiausiai atliekamas sezoninis laiko eilutės diferencijavimas. Skirtumai nustatomi tarp reikšmių, kurias skiria viena savaitė (672 reikšmės).



3.7 pav. Sezoniskai diferencijuota laiko eilutė

3.4 lentelė. Sezoniskai diferencijuotos laiko eilutės ADF testo rezultatai

Testo statistika	-60,56
Testo kritinė reikšmė c	-1,9416
H0 hipotezė	Atmesta

Sezoniskai diferencijuotos laiko eilutės grafike (3.7 pav.) pastebima, kad išliko nemaži sezoniniai skirtumai, tačiau sumažėjo didelės vertės pikų skaičius (lyginant su 3.4 pav.) ir laiko eilutė pasižymi mažesne dispersija.

Pakartotinai atlikus ADF testą sezoniskai diferencijuotiems duomenims, H_0 hipotezė, teigianti apie duomenų nestacionarumą – atmesta (3.4 lentelė).

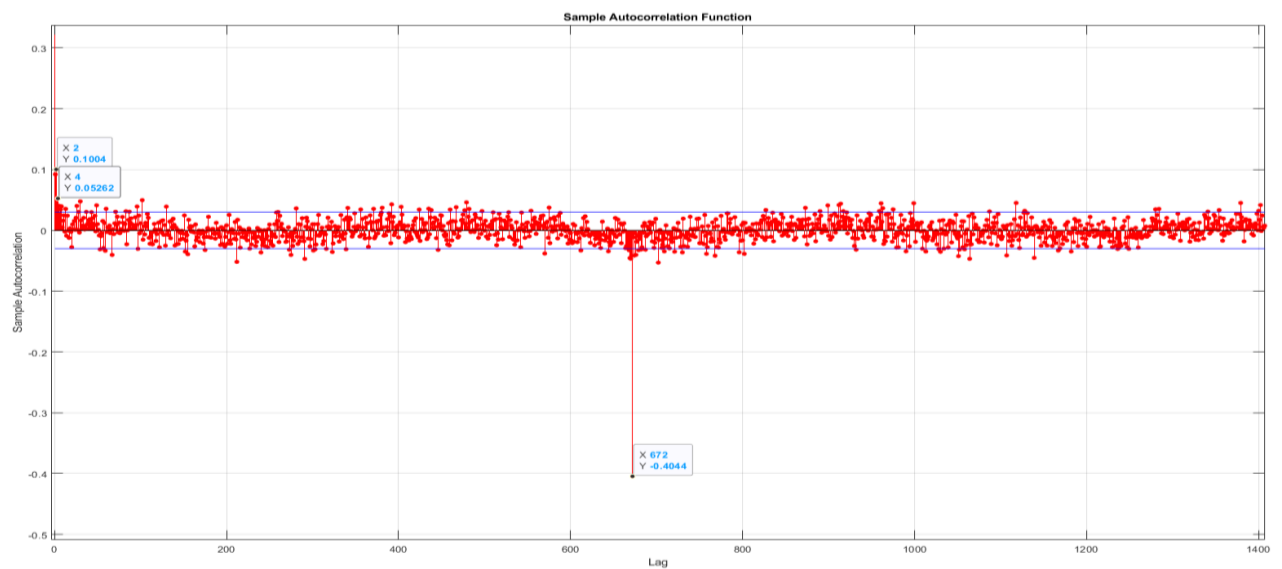
Rementis koreliogramomis (3.8 ir 3.9 pav.), sezoniskai diferencijuotų duomenų laiko eilutė yra stacionari. Autokoreliacijos funkcijoje pastebimas vienas sezoną atitinkantis pikas ties 672 reikšme, daugiau jo kartotinių nėra. Artimų vėlavimų autokoreliacija nutrūksta greitai. Pirmas ryškus kritimas įvyksta po 2 vėlavimų, sekantis po 4. Vėliau yra chaotiško pobūdžio reikšmių, kertančių pasikliautinojo intervalo ribą, tačiau jų amplitudė nėra didelė ir nėra aiškaus dėsningumo.

Dalinės autokoreliacijos funkcija nutrūksta greitai, po 2 vėlavimų (3.9 pav.). Tai rodo, kad abu vėlavimai yra reikšmingi, koreliuoja su esamu laiko momentu. Vėliau yra chaotiškų svyravimų, kurie kerta pasikliautinojo intervalo ribą, tačiau jų vertės nėra didelės. Dalinės autokoreliacijos funkcijoje pastebimi periodiškai pasikartojantys pikai, atitinkantys savaitės periodą (672, 1344, 2016) – sezoną atitinkančių vėlavimų dalinė autokoreliacija gęsta lėtai.

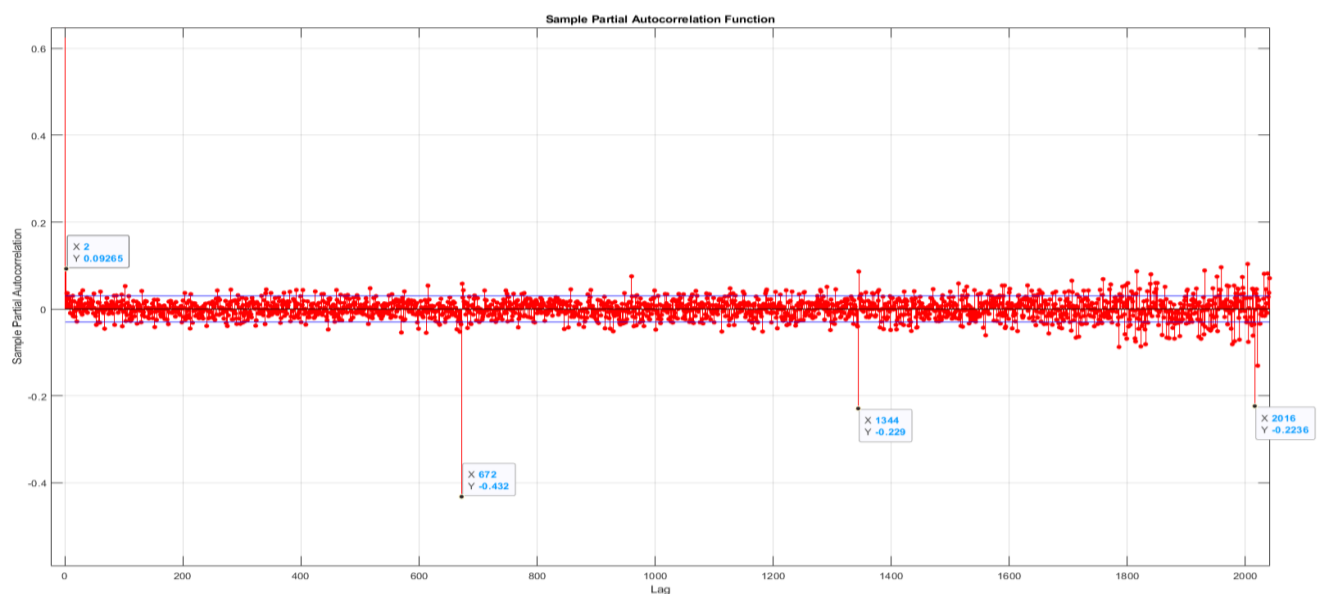
Kadangi autokoreliacijos ir dalinės autokoreliacijos funkcijos turi kritimą po 2 vėlavimo, tai rodo galimas AR(2) ir MA(2) modelio eiles. Modelio diagnostikos etape derėtų patikrinti ir mažesnes modelio eiles (AR(1), MA(1)), kadangi dalis ankstesnio vėlavimo autokoreliacijos gali

būti paaiškinta dėl pastarojo koreliacijos su aukštesnės eilės vėlavimu bei atlikti modelio liekamųjų paklaidų analizę.

Sezoninė laiko eilutės koreliacija yra gana aiški. Sezoninė autokoreliacija nutrūksta po 1, sezoną atitinkančio vėlavimo, o dalinėje autokoreliacijos funkcijoje sezoną atitinkančių vėlavimų koreliacija gęsta lėtai. Sezioniškai koreliuoti duomenys gali būti apibūdinti naudojant 1 eilės sezoninį slankiųjų vidurkių procesą – SMA(1).



3.8 pav. Sezioniškai diferencijuotų duomenų autokoreliacijos funkcija



3.9 pav. Sezioniškai diferencijuotų duomenų dalinės autokoreliacijos funkcija

Remiantis koreliogramomis, duomenys yra stacionarūs ir galima sudaryti prognozės modelį. Literatūroje rekomenduojama rinktis mažiausio laipsnio diferencijavimą, kuris užtikrina duomenų stacionarumą. Šio tyrimo metu siekiama atkreipti dėmesį į skirtingos modelio struktūros ypatumus prognozuojant eismo srautus, todėl papildomai atliekamas pirmos eilės diferencijavimas (sezoniškai diferencijuotiems duomenims), įvertinant laiko eilutės skirtumus tarp gretimų reikšmių. Pirma eile ir sezoniškai diferencijuotų duomenų grafikas pateiktas 2 priede.

Patikrinus pirma eile ir sezoniškai diferencijuotą laiko eilutę ADF testu, nestacionarumo požymių nėra (3.5 lentelė).

3.5 lentelė. Pirma eile ir sezoniškai diferencijuotos laiko eilutės ADF testo rezultatai

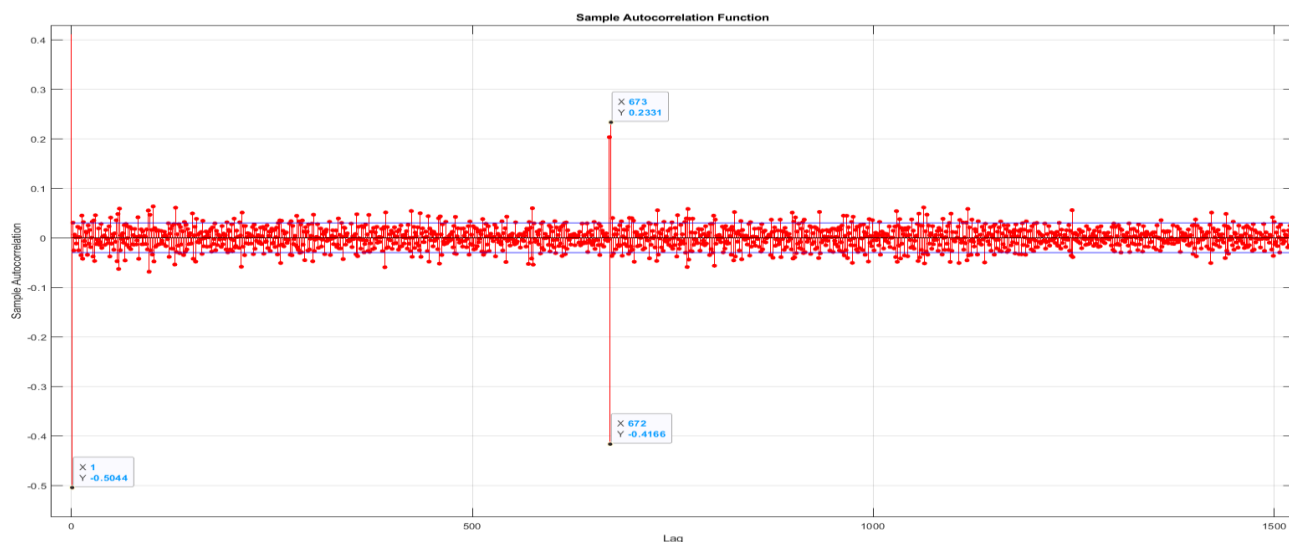
Testo kritinė reikšmė c	-1,9416
Testo statistika	-115,74
H0 hipotezė	Atmesta

Pirma eile ir sezoniškai diferencijuotos laiko eilutės koreliogramos rodo, kad duomenys yra stacionarūs (3.10 ir 3.11 pav.). Autokoreliacijos funkcijoje sumažėjo gretimų koreliuotų vėlavimų skaičius, liko tik 1 vėlavimas, kurio reikšmė yra neigiama, siekia -0,5. Tai rodo galimai perteklinį duomenų diferencijavimą, dėl to sudarant modelį reikia padidinti slankiųjų vidurkių MA(q) modelio eilę. Vertinant autokoreliacijos sezoninę dalį, pastebimas pikas ties 672 reikšme, koreliuoja ir ją lydintys vėlavimai (671, 673). Tai rodo galimą SMA(1) modelį. Vėliau koreliacija iškart nutrūksta ir daugiau sezoną atitinkančių kartotinių nėra. Atlikus tik sezoninį laiko eilutės diferencijavimą, šie vėlavimai (671, 673) nekoreliavo.

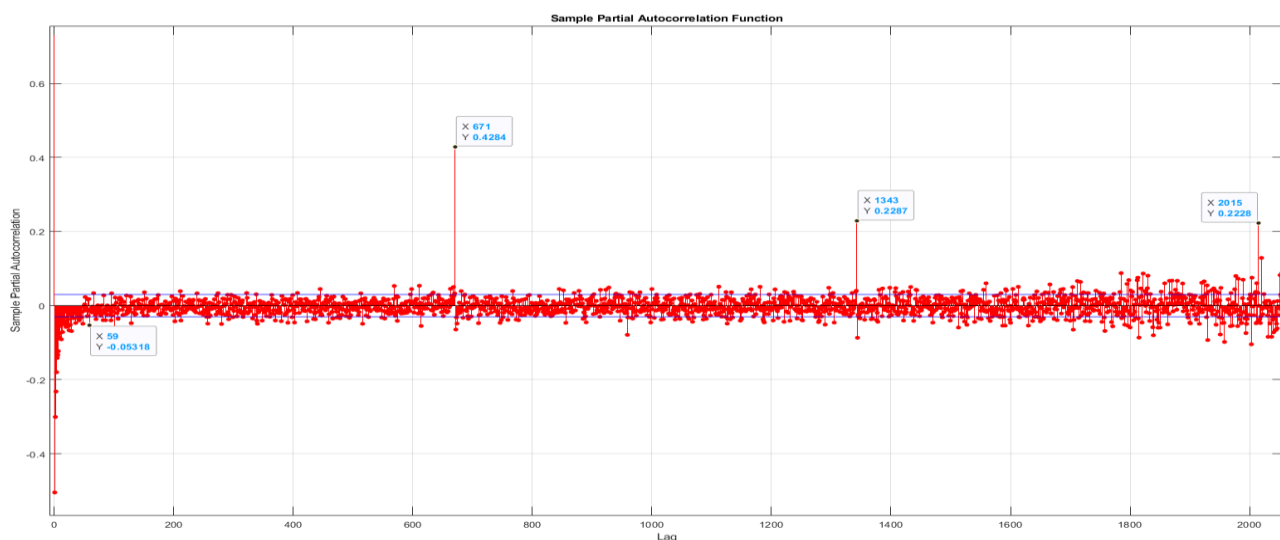
Dalinės autokoreliacijos funkcija gęsta lėtai ir netvarkingai, yra neigiama, prie pasikliautinojo intervalo ribų priartėja tik po 50–60 vėlavimų. Vertinant sezoninę dalį, yra nemažai sezoną atitinkančių pikų, šių vėlavimų koreliacija gęsta lėtai. Ties sezoną atitinkančiais vėlavimais dalinės autokoreliacijos funkcijoje atsiranda chaotiško pobūdžio koreliacija, kurios vertės nėra didelės.

Atlikus pirmos eilės diferencijavimą pastebima, kad duomenys gali būti apibūdinti naudojant tik slankiųjų vidurkių modelį. Kadangi autokoreliacija nutrūksta po 1 vėlavimo, o dalinės autokoreliacijos funkcija gęsta lėtai, tikėtinas 1 eilės slankiųjų vidurkių modelis – MA(1).

Vertinant sezoniškumą, autokoreliacijos funkcijoje yra vienas, sezoną atitinkantis vėlavimas, o dalinės autokoreliacijos funkcijoje pastebimas lėtas, sezoniškai pasikartojančių vėlavimų koreliacijos gėsimas. Dėl to ir šiuo atveju nustatoma, kad sezoninė koreliacija gali būti apibūdinta slankiųjų vidurkių modeliu – SMA(1).



3.10 pav. Pirma eile ir sezoniškai diferencijuotų duomenų autokoreliacijos funkcija



3.11 pav. Pirma eile ir sezoniškai diferencijuotų duomenų dalinės autokoreliacijos funkcija

3.1.3 Modelio identifikacija

Koreliogramų analizės metu nustatyta, kad sezoniškai arba sezoniškai ir pirma eile diferencijuoti duomenys yra stacionarūs ir gali būti naudojami prognozės modeliui sudaryti.

Kadangi pritaikius sezoninį duomenų diferencijavimą laiko eilutė tapo stacionari, galima būtų nebeįtraukti diferencijavimo pirma eile. Tačiau siekiant palyginti skirtingu laipsniu integruotų modelių tikslumą bei greitaveiką ir sudaryti lyginamuosius modelių tolimesniems bandymams, yra tiriami abu atvejai. Remiantis koreliogramomis, nustatomos galimos SARIMA modelio eilės (3.6 lentelė)

3.6 lentelė. Tikėtinos SARIMA modelio eilės

ARIMA(p,d,q)(P,D,Q) ₆₇₂ parametras	Tikėtina modelio eilė								
	Sezoniškai integruotas modelis						Sezoniškai ir pirma eile integruotas modelis		
p	1	2	0	0	1	2	0	0	1
d	0	0	0	0	0	0	1	1	1
q	0	0	1	2	1	2	1	2	1
P	0						0		
D	1						1		
Q	1						1		

3.1.4 Modelio įvertinimas ir diagnostika

Šiame etape atliekamas modelio regresijos koeficientų nustatymas ir alternatyvių modelių palyginimas. Naudojant pasirinktas SARIMA modelio parametrų eiles nustatomos regresijos lygties koeficientų reikšmės ir tikrinama, ar sudarytas modelis tinkamai apibūdino duomenų imtį. Siekiama nustatyti geriausiai duomenis apibūdinantį modelį, kurio liekamosios paklaidos būtų mažiausios ir nekoreliuotų tarpusavyje. Modelio paklaidas įvertinti galima informaciniais kriterijais (AIC, BIC). Geriausias yra tas modelis, kurio informacinių kriterijų reikšmės yra mažiausios. Tada tikrinamos modelio liekamosios paklaidos. Šiame darbe liekamųjų paklaidų koreliacija tikrinama vizualiai vertinant liekamųjų paklaidų autokoreliacijos funkciją bei naudojant Ljung–Box testą, kuriuo nustatoma ar modelio liekamosios paklaidos yra atsitiktinio pobūdžio ir sudaro baltąjį triukšmą. Plačiau apie informacinius kriterijus, modelio liekamasias paklaidas ir jų vertinimą žr. 2.1.5 skyrelyje.

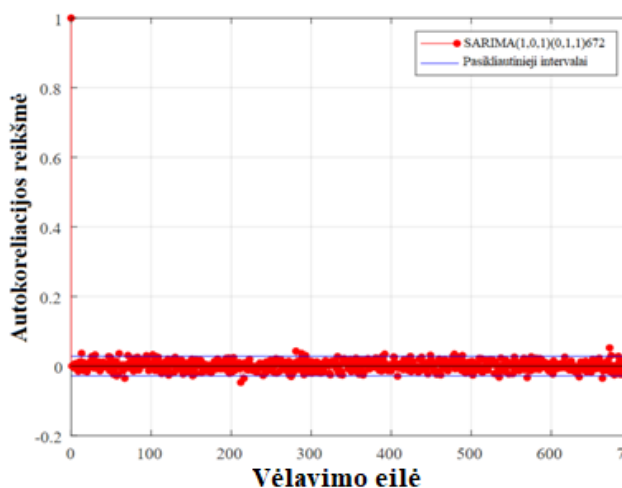
Vertinant sudarytą modelį rekomenduojama atkreipti dėmesį į nustatytų regresijos koeficientų reikšmingumą. Jei aukštesnės eilės vėlavimo regresijos koeficientai yra nereikšmingai maži (regresijos koeficiento t statistikos vertė yra mažesnė už $|2|$), derėtų vengti šių modelio eilių, siekiant supaprastinti modelio struktūrą ir išvengti galimai netinkamo duomenų įvertinimo. Išimtis taikoma, kai aukštesnės eilės vėlavimo regresijos koeficientai yra svarbūs, o žemesnio – ne.

Jei modelio diagnostikos metu pastebima liekamųjų paklaidų autokoreliacija, didinama AR(p) arba MA(q) vėlavimo eilė, priklausomai nuo ankščiau nagrinėtų koreliogramų. Jei vienas iš regresijos koeficientų yra statistiškai nereikšmingas, rekomenduojama mažinti atitinkamą modelio eilę iki statistiškai reikšmingos.

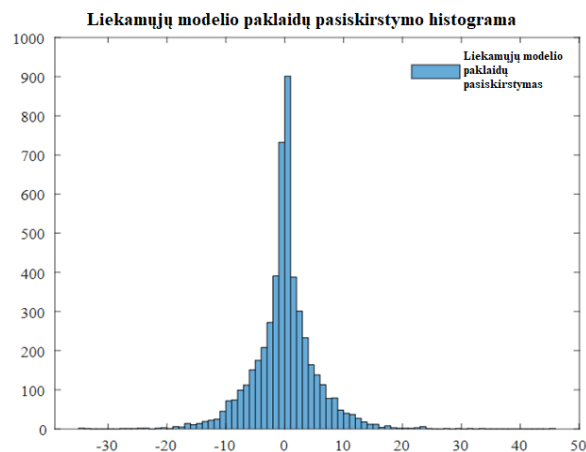
Išbandytų modelių diagnostikos rezultatų suvestinė pateikta 3 priede.

Remiantis 3 priede pateikta suvestine, nustatytus kriterijus tenkina ir įvertintas mažiausiais informacinių kriterijų įverčiais SARIMA(1,0,1)(0,1,1)₆₇₂ modelis. Šio modelio liekamųjų paklaidų autokoreliacijos funkcijoje yra keletas atsitiktinių vėlavimų, kertančių pasikliautinojo intervalo ribą,

tačiau jų vertės nėra didelės (3.12 pav.). Funkcija tvarkinga, nėra didelės vertės atsitiktinių šuolių. Jei yra reikšminga paklaidų autokoreliacija, ji bus nustatyta atliekant Ljung–Box testą. Remiantis liekamųjų paklaidų histograma (3.13 pav.), modelio liekamosios paklaidos yra pasiskirsčiusios apie nulio reikšmę, pagal normalųjį skirstinį.



3.12 pav. SARIMA(1,0,1)(0,1,1) modelio liekamosios paklaidos



3.13 pav. SARIMA(1,0,1)(0,1,1) modelio liekamųjų paklaidų histograma

Papildomai naudojant pirmos eilės diferencijavimą (kartu su sezoniniu), modelio kokybę apibūdinančių informacinių kriterijų reikšmės nepagerėjo. Nustatytos modelio eilės yra artimos pirmajam atvejui, naudojant tik sezoninį diferencijavimą. Nustatyti AR(1) ir MA(2) modeliai. Taip pat tikėtinas modelis, kurio eilės – AR(2), MA(2). Slankiųjų vidurkių modelio MA(q) dalis padidėjo vienu vienetu dėl padidintos diferencijavimo eilės, o tai atitinka prieš tai pasirinktą, tik sezoniskai integruotą modelį. Sezoninėje dalyje pokyčių nėra.

Siekiant įvertinti skirtingos struktūros modelių veikimą, palyginimui pasirenkami šie modeliai, atitinkantys nustatytus kriterijus ir turintys mažus informacinių kriterijų įverčius:

- SARIMA(0,0,5)(0,1,1)₆₇₂;
- SARIMA(2,1,2)(0,1,1)₆₇₂;
- SARIMA(1,1,2)(0,1,1)₆₇₂.

4 priede pateiktos šių modelių lygtys, liekamųjų paklaidų autokoreliacijos funkcijos, liekamųjų paklaidų histogramos.

Nustatyti ar modelio liekamosios paklaidos yra atsitiktinės ir sudaro baltąjį triukšmą galima naudojant Ljung–Box testą. Testo metu tiriamas liekamųjų paklaidų skirstinys. Nustatoma H_0 hipotezė, teigianti, kad modelio liekamosios paklaidos sudaro baltąjį triukšmą. Jei H_0 hipotezė patvirtinama, sudarytas modelis tinkamai apibūdina tiriamų duomenų imtį ir gali būti naudojamas

prognozuoti tiriamą procesą. Atmetus nulinę hipotezę, bandomi kiti modeliai, keičiama modelio struktūra.

Ljung–Box testo metu nurodomas vėlavimų skaičius, kurio liekamųjų paklaidų koreliacija yra tikrinama. Tiriant paprasto ARIMA modelio liekamųjų paklaidų autokoreliaciją įprastai tikrinama 20–30 vėlavimų. Kadangi nagrinėjamas sezoninis procesas, reikia įvertinti ar teisingai nustatyta sezoninė modelio dalis, todėl tikrinama daugiau vėlavimų nei sudaro vieną sezoną.

Visų sudarytų modelių H_0 hipotezė, teigianti kad paklaidos yra atsitiktinės ir nekoreliuotos, yra patvirtinta (3.7 lentelė). Visi sudaryti modeliai yra tinkami eismo srautų prognozavimui.

3.7 lentelė. Ljung–Box liekamųjų paklaidų testo rezultatai

Eil. Nr.	Modelis	H0 hipotezė	Patikimumo lygmuo	Tikrinti vėlavimai	p reikšmė
1	SARIMA(1,0,1)(0,1,1) ₆₇₂	Patvirtinta	0,05	30	0,684
				700	0,1942
2	SARIMA(0,0,5)(0,1,1) ₆₇₂	Patvirtinta		30	0,618
				700	0,1786
3	SARIMA(2,1,2)(0,1,1) ₆₇₂	Patvirtinta		30	0,371
				700	0,1330
4	SARIMA(1,1,2)(0,1,1) ₆₇₂	Patvirtinta		30	0,820
				700	0,1215

Modelio diagnostikos metu pastebėta, kad esant tik sezoniškai diferencijuotiems duomenims, geriausiai imtį apibūdina sujungtas AR(1) ir MA(1) modelis. Bandant taikyti tik autoregresinio proceso AR(p) modelį, liekamųjų paklaidų autokoreliacija neišnyko net ir didinant modelio eilės numerį. Su kiekviena aukštesne p eile, koreliuotas liekamųjų paklaidų vėlavimas persistumdavo viena eile. Bandant tik slankiųjų vidurkių modelį MA(q), buvo reikalinga didelė modelio eilė (vėlavimas) siekiant panaikinti modelio liekamųjų paklaidų autokoreliaciją. Nustatytas slankiųjų vidurkių MA(5) modelis – SARIMA(0,0,5)(0,1,1). Vertinant nustatytą MA(q) koeficientų reikšmingumą (t statistikos atžvilgiu), visi koeficientai yra reikšmingi, kadangi jų vertė buvo didesnė už |2|. Dėl reikšmingo slankiųjų vidurkių proceso ir mažų informacinių kriterijų testų reikšmių, šis modelis bus naudojamas palyginti pasirinkto SARIMA(1,0,1)(0,1,1) modelio tikslumą ir greitaveiką.

Taikant sujungtą AR(1) ir MA(1) modelį buvo pasiektos mažiausios AIC ir BIC informacinių kriterijų reikšmės. Papildomai vertinant šio modelio regresijos koeficientų reikšmingumą t statistikos atžvilgiu, visi koeficientai yra reikšmingi. Padidinus parametrų eiles iki 2, modelis galimai tapo perteklinis, kadangi nebuvo reikšmingų koeficientų, visų regresijos lygties koeficientų reikšmės t statistikos atžvilgiu buvo mažesnės už |2|. Vertinant sezoninę dalį, pasirinkta tinkama sezoninio diferencijavimo eilė ir sezoninis slankiųjų vidurkių modelis SMA(1). Yra nedidelė sezoninės dalies liekamųjų paklaidų koreliacija ties 673 vėlavimu, tačiau jos vertė nėra didelė, paklaidų nereikšmingą koreliaciją įrodo Ljung–Box testas (3.6 lentelė). Daugiau pikų

liekamųjų paklaidų autokoreliacijos funkcijoje nėra. Kadangi SARIMA(1,0,1)(0,1,1) modelis geriausiai apibūdino turimų duomenų imtį (mažiausios informacinių kriterijų reikšmės), šis modelis pasirenkamas kaip pagrindinis tyrimo objektas.

SARIMA(1,0,1)(0,1,1)₆₇₂ modelio lygtis (3.1) ir regresijos koeficientų vertės (3.8 lentelė):

$$(1 - \varphi_1 L)(1 - L^{672})y_t = (1 + \theta_1 L)(1 + \theta_{672} L^{672})\varepsilon_t \quad (3.1)$$

3.8 lentelė. SARIMA(1,0,1)(0,1,1) modelio regresijos koeficientų vertės

Parametras	Vertė	t statistikos reikšmė
Konstanta	0	
AR{1}	0.71897	23.4205
MA{1}	-0.60654	-17.6478
SMA{672}	-0.59605	-80.7877
Variance	28.8954	87.5322

Palyginimui skirtų modelių lygtys bei regresijos koeficientų vertės pateiktos 4 priede.

3.1.5 Modelio validacija

Modelio validacija atliekama lyginant prognozuojamas intensyvumo vertes su realiai fiksuotu intensyvumu 2018-10-24 – 2018-10-31 laikotarpiu. Atliekamos vieno žingsnio, 15 min laikotarpio prognozės. Iš viso sudaromos 768 vieno žingsnio prognozės. Nustatomos suminės paklaidos. Prognozės tikslumas vertinamas MAPE paklaida. Kiekviena prognozė sudaroma remiantis realių duomenų imtimi, o ne ankstesnėmis prognozėmis. SARIMA modelio pakartotinis regresijos koeficientų įvertinimas po kiekvieno žingsnio nėra atliekamas, t. y. modelis nebevertina susidariusios paklaidos tarp prognozuojamos reikšmės ir realių duomenų. Kadangi validacijai turima duomenų imtis siekia virš vienos savaitės duomenų (672 reikšmių), galima objektyviai įvertinti, kaip sudarytas modelis geba prognozuoti eismo srautus ir prisitaikyti prie savaitės eigoje kintančios eismo srautų dinamikos.

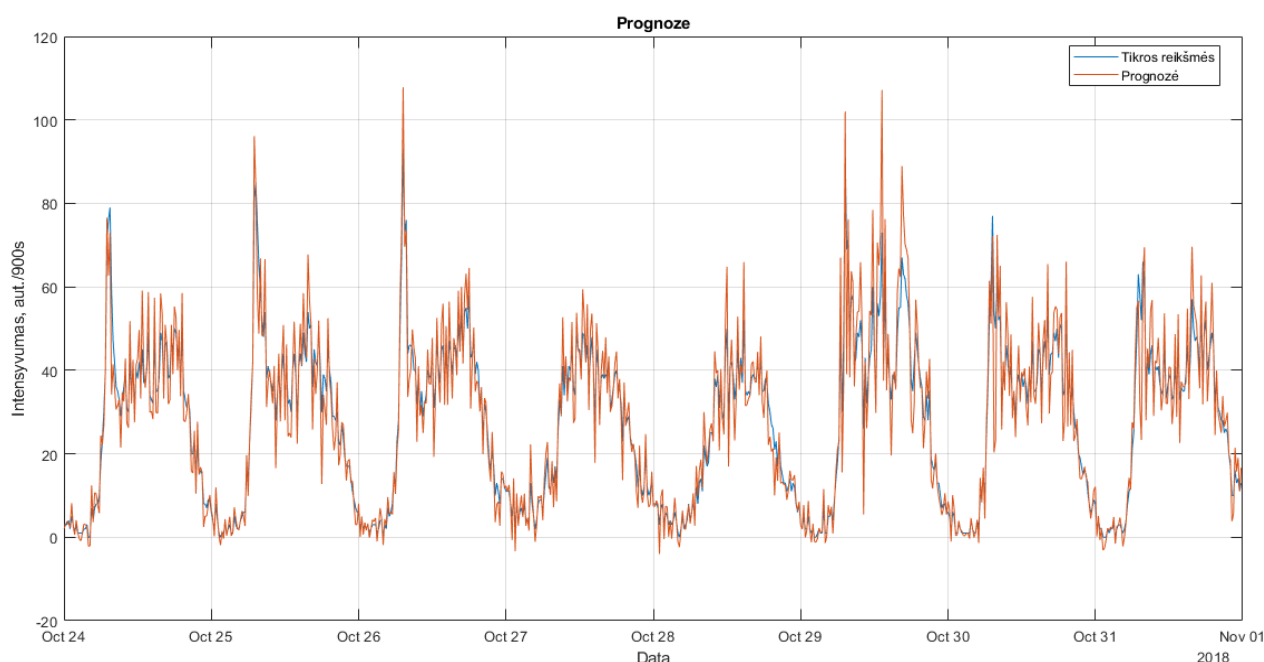
Siekiant palyginti SARIMA modelio prognozavimo gebą, pagal 2.2 poskyrį sudarytas LSTM modelis. Šio modelio prognozė pateikta 3.15 pav. 5 priede pateikiami grafikai, apibūdinantys modelio mokymosi procesą minimizuojant RMSE.

Validacijos metu buvo laikomasi šių principų:

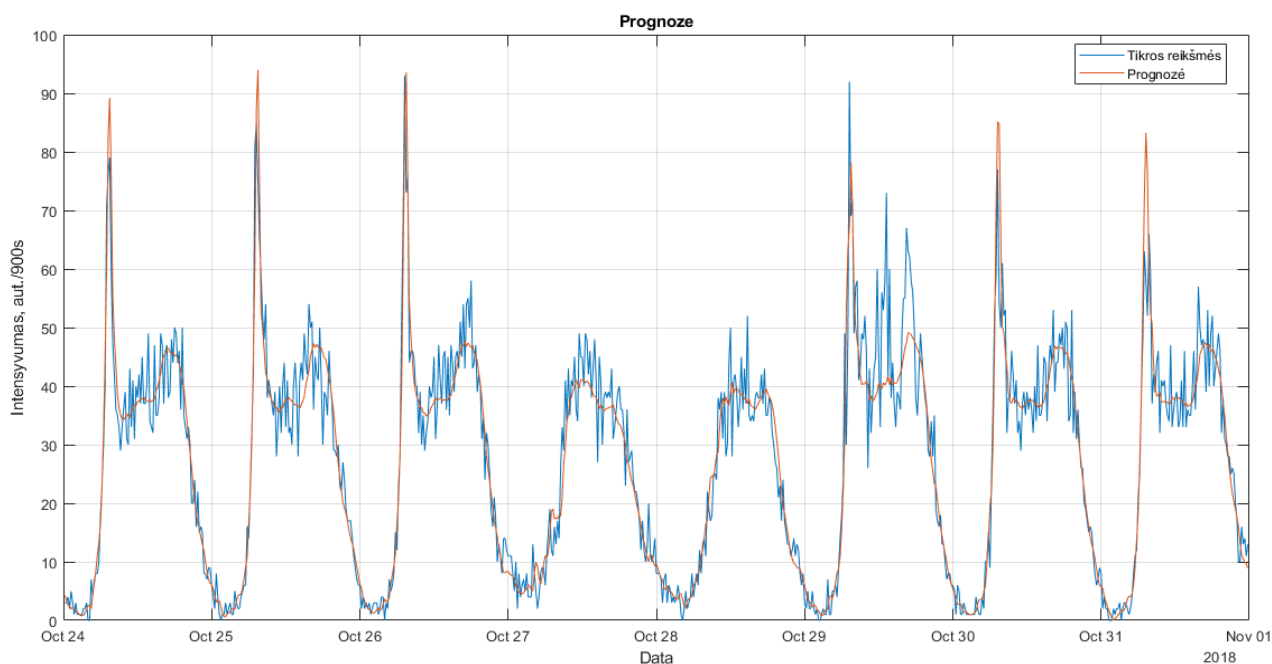
1. Nustatant prognozės modelio paklaidas, visos realių duomenų reikšmės lygios nuliui buvo pakeistos į 1 (artimiausią sveikąjį skaičių), kadangi skaičiuojant MAPE paklaidas reikalinga atlikti dalybą iš realių duomenų, o dalyba iš nulio negalima.
2. SARIMA modelio prognozės grafike (3.14 pav.) pastebima, kad kai kuriais atvejais prognozuojamos reikšmės buvo neigiamos, kadangi modelio sudarymo metu negalima nustatyti apribojimo veikti tik natūraliųjų skaičių aibėje. Esant neigiamai prognozės vertei, ji buvo keičiama į 1 ir galutiniai rezultatai skaičiuojami nuo patikslintų verčių.
3. Modeliuose nebuvo galima nustatyti apribojimo, kad visi skaičiai būtų sveikieji. Kadangi automobiliai yra skaičiuojami natūraliaisiais skaičiais, skaičiuojant paklaidas prognozės reikšmės buvo apvalinamos laikantis apvalinimo taisyklių, iki artimiausio sveikojo skaičiaus.

Galutinė prognozė yra laikoma ta, kuri yra patisklinta pagal šiuos principus.

Palyginimui sudarytų SARIMA modelių prognozės pateiktos 6 priede. 7 priede pateiktas 1 dienos skaitinių prognozės reikšmių pavyzdys.



3.14 pav. SARIMA(1,0,1)(0,1,1)₆₇₂ modelio prognozė



3.15 pav. LSTM modelio prognozė

3.9 lentelė. Prognozės paklaidų ir skaičiavimo trukmės rezultatai

Eil. Nr.	Modelis	Modelio paklaidos			Modelio greitaveika	
		MAPE, %	MAE	RMSE	Modelio įvertinimo trukmė, s	Vienos reikšmės prognozavimo trukmė, s
1	SARIMA(1,0,1)(0,1,1) ₆₇₂	17,60	3,88	5,83	180	0,01
2	SARIMA(0,0,5)(0,1,1) ₆₇₂	17,96	3,88	5,83	311	0,01
3	SARIMA(2,1,2)(0,1,1) ₆₇₂	17,88	3,89	5,84	470	0,02
4	SARIMA(1,1,2)(0,1,1) ₆₇₂	17,88	3,88	5,84	407	0,01
5	LSTM	24,36	3,94	5,66	1065	0,03

Žvelgiant į SARIMA(1,0,1)(0,1,1) modelio prognozės grafiką (3.14 pav.) pastebima, kad modelis gerai atkartojo cikliškus realių duomenų svyravimus. Atkartotas sezoniškumas, modelis sugebėjo įvertinti savaitgalio metu sumažėjusius eismo srautus. Šio modelio prognozė yra linkusi į didesnius nuokrypius keičiantis eismo srauto intensyvumui bei piko metu, tuo metu pastebimi didesni nuokrypiai nuo realių duomenų. SARIMA(1,0,1)(0,1,1) modelio MAPE paklaida buvo mažiausia, siekė 17,60% (3.9 lentelė). Šis modelis pasižymi paprasta struktūra, nedideliu parametų skaičiumi, todėl jo greitaveika taip pat geresnė nei kitų modelių. Vertinant prognozuojamo laikotarpio žingsnį – 15 min, vienos reikšmės prognozavimas užtrunka nereikšmingai mažą laiko tarpą – 0,01 s. Modelio įvertinimas truko 180 s (3 min), todėl naudojant 1,5 mėn. duomenų imtį ir prognozuojant 15 min žingsniu, šį modelį galima pakartotinai atnaujinti (perskaiciuoti regresijos lygties koeficientų reikšmes) kiekvieną kartą atliekant prognozavimą.

SARIMA(0,0,5)(0,1,1) modelis, pasižymintis vien tik slankiųjų vidurkių modelio naudojimu, taip pat sudarė tikslias prognozes. Paklaida nuo SARIMA(1,0,1)(0,1,1) modelio skyrėsi nedaug, siekė 17,96%. Tačiau šio modelio greitaveika buvo daug blogesnė. Modelio įvertinimas

užtruko 319 s, t. y. trečdaliu ilgiau nei pirmojo modelio. Prastesnė modelio greitaveika paaiškinama didele slankiųjų vidurkių modelio eile, kadangi šio proceso metu vertinami ne tiesiogiai išmatuoti duomenys, o paklaidos ir reikalinga atlikti papildomus skaičiavimo veiksmus.

Analizuojant pirma eile integruotus modelius, paklaidos labai nesiskyrė nuo pirmojo, siekia 17,88%, tačiau dėl papildomai atlikto integracijos proceso modelių įvertinimo trukmė buvo dvigubai didesnė nei SARIMA(1,0,1)(0,1,1) modelio, siekė 407–470 s.

LSTM modelio paklaida buvo didesnė nei SARIMA modelių, siekė 24,37%. Vertinant šio modelio prognozės grafiką (3.15 pav.) pastebima tolygesnė prognozės kreivė, modelis neatkartojo cikliško realiųjų duomenų svyravimų, laikėsi vidutinių reikšmių. Šio modelio įvertinimo (mokymo) trukmė yra daug didesnė, siekė 1065 s (17,8 min), vienos reikšmės prognozavimo trukmė – 0,03 s. Modelis tinkamas prognozuoti realiuoju laiku, tačiau atnaujinimo (pakartotinio mokymo) procesas turėtų būti atliekamas nepriklausomai nuo prognozavimo.

Žvelgiant į sudarytus prognozės grafikus (3.14 ir 3.15 pav.) pastebimas didelis triukšmas realiuose duomenyse, šuoliai tarp gretimų reikšmių. SARIMA modelis gerai atkartojo šį triukšmą, o LSTM modelis pasižymėjo tolygesne kreive, nebuvo cikliško svyravimų ir dėl susidariusio nuolatinio skirtumo tarp realiųjų duomenų MAPE paklaidos vertė buvo didesnė.

Siekiant didesnio prognozavimo tikslumo ir laiku pastebėti kintančius eismo srautus ilgalaikėje perspektyvoje, reikalinga periodiškai atnaujinti modelius, atliekant pakartotinį modelių įvertinimą su papildyta duomenų imtimi. Šį veiksmą rekomenduojama atlikti nuolatos, kaip įmanoma mažesniu periodiškumu. SARIMA modelių atveju, modelio įvertinimo trukmė buvo mažesnė nei prognozuojamas laikotarpis, todėl modelio įvertinimą galima kartoti kiekvieną kartą, atliekant prognozavimą realiuoju laiku. Jei modelio įvertinimo laikas viršija prognozuojamo laikotarpio trukmę (kaip nagrinėtu LSTM modelio atveju), patartina pakartotiną modelių įvertinimą (regresijos koeficientų perskaičiavimą, mokymą) atlikti mažiausiu įmanomu periodiškumu arba bent kartą per parą, siekiant laiku pastebėti ilgalaikes eismo srautų tendencijas.

3.1.6 Dalinės išvados

1. Apibendrinant skyrelio rezultatus, sudarytas SARIMA(1,0,1)(0,1,1) prognozės modelis, kurio MAPE paklaida siekia 17,6%. Kadangi šio modelio paklaida mažesnė nei siekiama 20% riba, modelio prognozė yra vertinama kaip tiksli. Modelis pasižymėjo gera greیتaveika. Prognozuojant vieną reiškmę užtrunkama 0,01 s, o modelio įvertinimas truko 180 s. Prognozuojant 15 min (900 s) žingsniu, šis modelis gali būti pakartotinai atnaujintas (perskaičiuotos regresijos koeficientų reikšmės) po kiekvienos prognozės, kadangi duomenų įvertinimo trukmė yra mažesnė už prognozės žingsnį.
2. LSTM modelio paklaidos buvo didesnės nei SARIMA modelių, siekė 24%. Šio modelio greیتaveika taip pat buvo prastesnė, modelio įvertinimo procesas truko 1065 s (17 min 45 s), todėl esamomis sąlygomis modelio pakartotinis mokymas turėtų būti vykdomas nepriklausomai nuo prognozavimo proceso.
3. Vertinant papildomai sudarytus SARIMA modelius, perteklinė modelio struktūra gali būti nenaudinga sudarant prognozės modelį, kadangi mažesnė paklaida nėra pasiekama, o modelio greیتaveika labai sulėtėja. Esant galimybei reikėtų rinktis mažesne eile integruotus modelius.
4. Nagrinėjant visų modelių greیتaveiką, vienos reikšmės prognozavimo trukmė yra labai maža (0,01–0,03 s) ir lyginant su prognozės žingsniu – 15 min, visi modeliai yra tinkami prognozavimui realiuoju laiku.
5. Remiantis literatūros apžvalga, Smith *et al.* (2002), Williams, Hoel (2003) atliktuose tyrimuose, prognozuojant eismo srautus greیتkelyje, SARIMA(1,0,1)(0,1,1) modelis taip pat buvo pripažintas kaip tiksliausiai prognozavęs ARIMA grupės modelis.

3.2 Prognozės modelio universalumo įvertinimas

Uždavinys: Patikrinti sudaryto prognozės modelio gebą prognozuoti kitų eismo juostų intensyvumą.

Metodika: SARIMA(1,0,1)(0,1,1) modelio regresijos koeficientai įvertinami naudojant kitų eismo juostų duomenis. Atliekamas prognozavimas. Tikrinama, kaip pasikeitė prognozės paklaidos. Prognozės rezultatai palyginami su kitais SARIMA bei LSTM modeliais. Nustačius reikšmingą paklaidų padidėjimą, papildomai vertinamas SARIMA(1,0,1)(0,1,1) modelio tinkamumas artimiausio jutiklio duomenims.

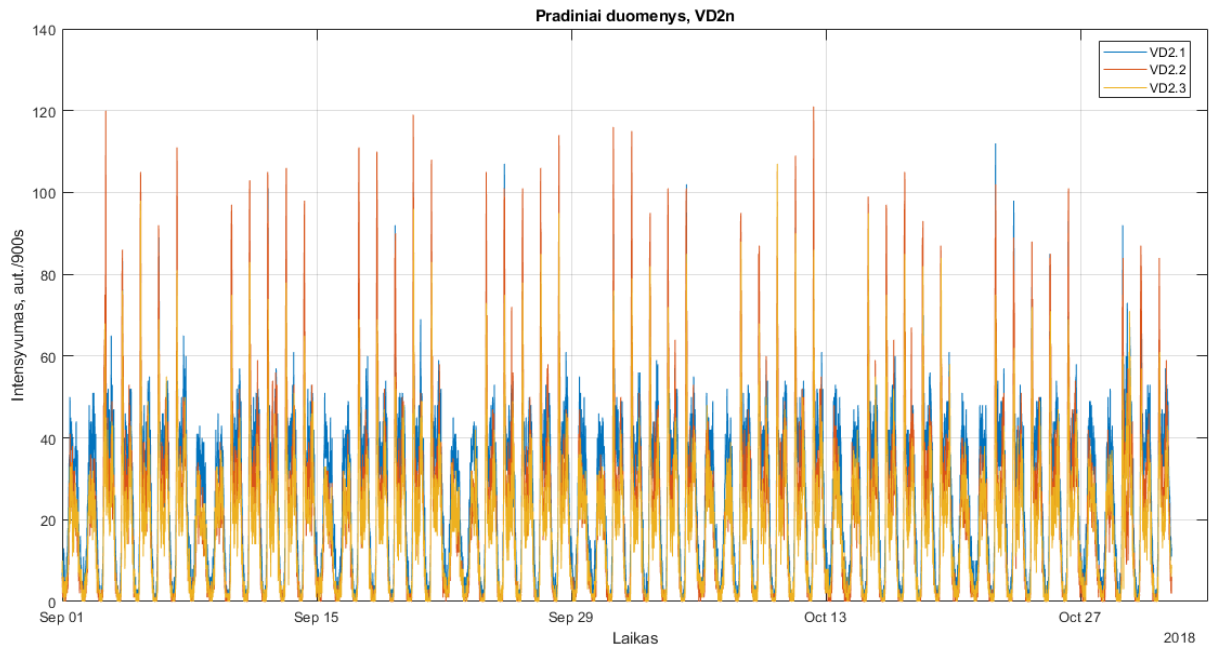
Trumpas aprašas

Šiame skyriuje vertinamas sudaryto SARIMA(1,0,1)(0,1,1) modelio universalumas, tikrinant modelio gebą prognozuoti kitų Justiniškių–Virbeliškių g. sankryžos eismo juostų intensyvumą.

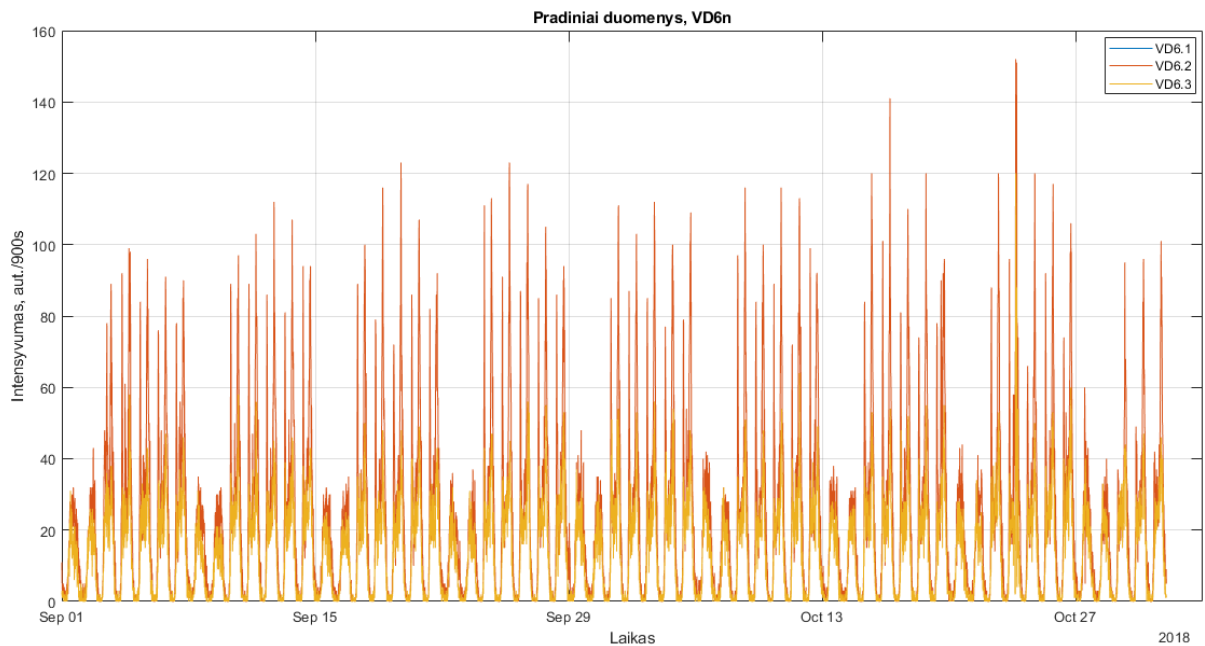
Kiekvienas duomenų rinkinys pasižymi unikaliomis ypatybėmis, skirtingu duomenų pasiskirstymu, todėl nagrinėjant skirtingų eismo juostų intensyvumą, kiekvieno jutiklio duomenys turėtų būti vertinami individualiai. Tai reiškia, kad kiekvienu atveju iš naujo turi būti nustatyta prognozės modelio struktūra ir įvertintas modelio tinkamumas prognozuoti konkrečios eismo juostos intensyvumą. Šis aspektas svarbus taikant parametrinius prognozės modelius, kadangi šių modelių sudarymas nėra visiškai automatizuotas, reikalinga rankiniu būdu nustatyti koreliuotus vėlavimus, parinkti modelio struktūrą ir ją patikrinti. Modelio struktūros nustatymas yra sudėtingiausias žingsnis SARIMA modelio sudarymo metu, šis veiksmas reikalauja laiko bei pastangų. Siekiant supaprastinti parametrinių prognozės modelių taikymą būtų naudinga taikyti standartizuotą prognozės modelį, gebantį vertinti skirtingų eismo juostų duomenis, taip išvengiant unikalių modelių parengimo kiekvienam atvejui atskirai.

Ankstesniame skyriuje, naudojant Justiniškių g. pirmos eismo juostos duomenis (VD2.1 jutiklio), sudarytas SARIMA (1,0,1)(0,1,1) prognozės modelis. Šis modelis bandomas su kitų Justiniškių–Virbeliškių g. eismo juostų duomenimis. Modelio regresijos koeficientai nustatomi remiantis kiekvieno jutiklio duomenimis atskirai. Nustatomos prognozės paklaidos ir palyginama su kitų prognozės modelių rezultatais. Kadangi LSTM modelis kiekvienu atveju yra parengiamas pagal atskiros jutiklio duomenis, nepriklausomai nuo ankščiau nustatytų regresinių ryšių, šis modelis yra indikatorius, apibūdinantis kiekvieno jutiklio duomenų imties savybę. LSTM tinklo parametrai, aptarti 2.2 poskyryje, nėra keičiami.

3.2.1 Grafinė analizė



3.16 pav. Justiniškių g. (link Laisvės pr.) eismo juostų intensyvumas



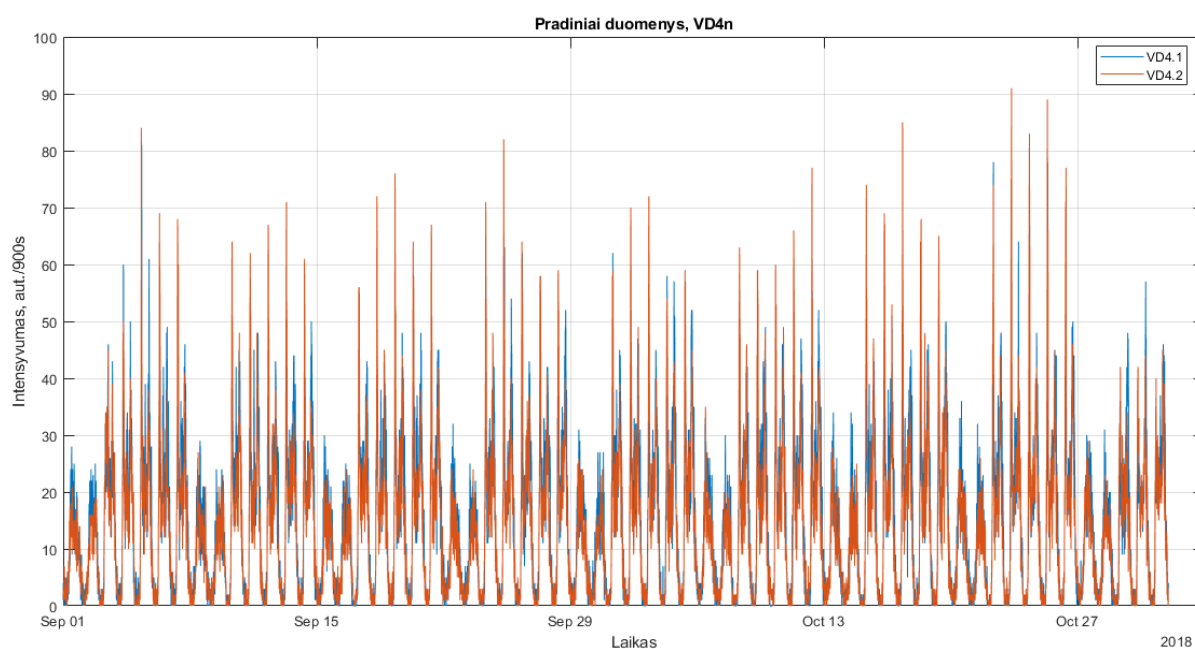
3.17 pav. Justiniškių g. (nuo Laisvės pr.) eismo juostų intensyvumas

(3.16) pav. pateikti Justiniškių g. (link Laisvės pr.) visų eismo juostų intensyvumo duomenys. Žvelgiant į grafiką pastebima, antros (VD2.2) bei trečios (VD2.3) eismo juostos intensyvumo vertės išlaiko panašius dėsningumus laike kaip ir pirmą eismo juosta (VD2.1). Antros eismo juostos jutiklio fiksuotos reikšmės ne piko metu yra šiek šiek mažesnės už pirmos, siekia 35–40 aut./15 min, trečios – 25–30 aut./15 min. Piko metu visų eismo juostų intensyvumo vertės susilygina, siekia apie 90–100 aut./15 min. Kadangi trečios eismo juostos intensyvumas dienos metu būna mažesnis, piko metu pastebimas didesnis šuolis. Remiantis grafiku, antros eismo

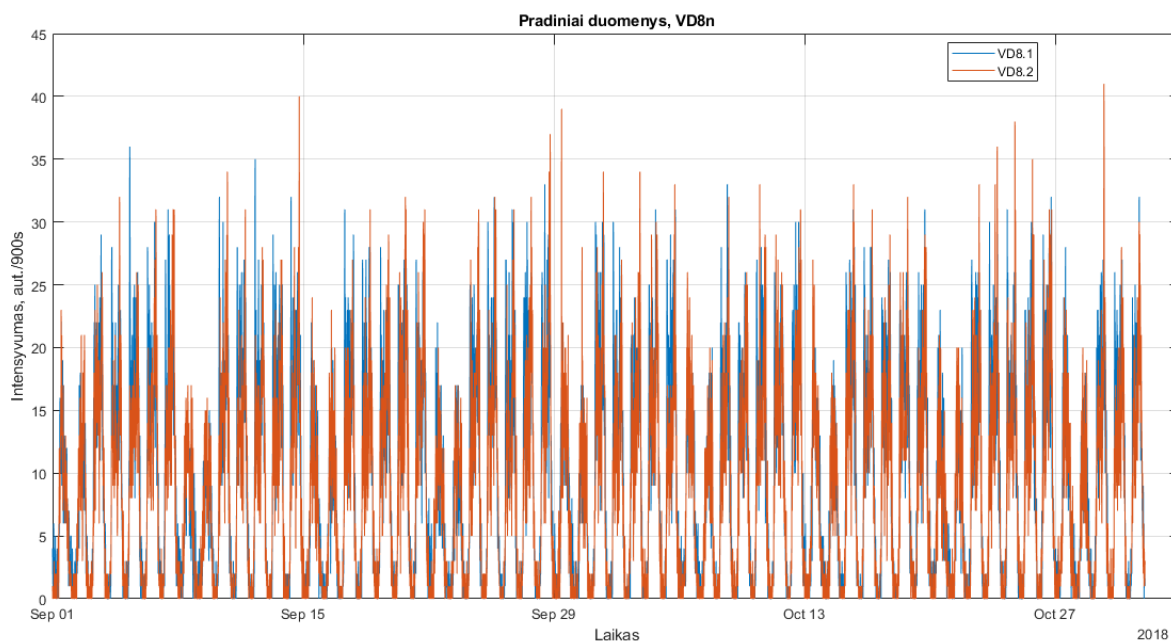
juostos duomenys gana panašūs į pirmos, tačiau trečios eismo juostos jutiklio fiksuoto intensyvumo vertės skiriasi.

Žvelgiant į Justiniškių g. priešingos krypties (nuo Laisvės pr.) eismo juostų duomenis (3.17 pav.) pastebima, kad intensyvumo vertės dienos metu yra truputį mažesnės, siekia apie 30 aut./15 min. Skirtingai nei priešingos krypties juostose, didžiausias intensyvumas fiksuotas vakarinio piko metu, siekia 100–120 automobilių. Vertinant pirmos (VD6.1) ir antros (VD6.2) eismo juostos duomenis pastebima, kad šių jutiklių duomenys yra identiški. Tai galima sisteminė klaida.

Virbeliškių g. intensyvumas daug mažesnis nei Justiniškių g., siekia 10–20 aut./15 min. Važiuojančių iš Medeinos g. (3.18 pav.) intensyvumas padidėja rytinio piko metu, siekia 70–80 aut./15 min. Vėliau būna gana stabilus. Eismo srautas iš Budiniškių g. (3.19 pav.) – stabilus, piko metu intensyvumas labai nepadidėja, siekia 30–40 aut./15 min. Mažiausias intensyvumas dienos metu fiksuojamas Virbeliškių g. antroje eismo juostoje (nuo Budiniškių g.), siekia 10 aut./15 min.



3.18 pav. Virbeliškių g. (nuo Medeinos g.) eismo juostų intensyvumas



3.19 pav. Virbeliškių g. (nuo Budiniškių g.) eismo juostų intensyvumas

3.2.2 Prognozės rezultatai

Taikant SARIMA(1,0,1)(0,1,1) prognozės modelį, prognozuojant kitų eismo juostų intensyvumą, paklaidos padidėjo nuo 18% iki 24–25% (3.10 lentelė). Didžiausia paklaida fiksuota Virbeliškių g. antros eismo juostos (nuo Budiniškių g.) VD8.2 jutiklio duomenims – 29,3%. Remiantis (3.19 pav.), ši eismo juosta pasižymi mažiausiu intensyvumu ir duomenų kreivė nuo pradinio atvejo reikšmingai skiriasi.

Kitų SARIMA modelių paklaidos taip pat padidėjo. SARIMA(0,0,5)(0,1,1) paklaidos buvo artimos pirmajam modeliui, kai kuriais atvejais keliomis dešimtosiomis procento ir mažesnės. Pirma eile integruotų modelių paklaidos pasižymėjo didesne sklaida, siekė 21–39%. Modeliai prasčiau vertino didesnio intensyvumo eismo juostas, pavyzdžiui Justiniškių g. (nuo Laisvės pr.) pirmos ir antros eismo juostų paklaidos siekė 38% (šių jutiklių duomenys vienodi dėl galimos sisteminės klaidos).

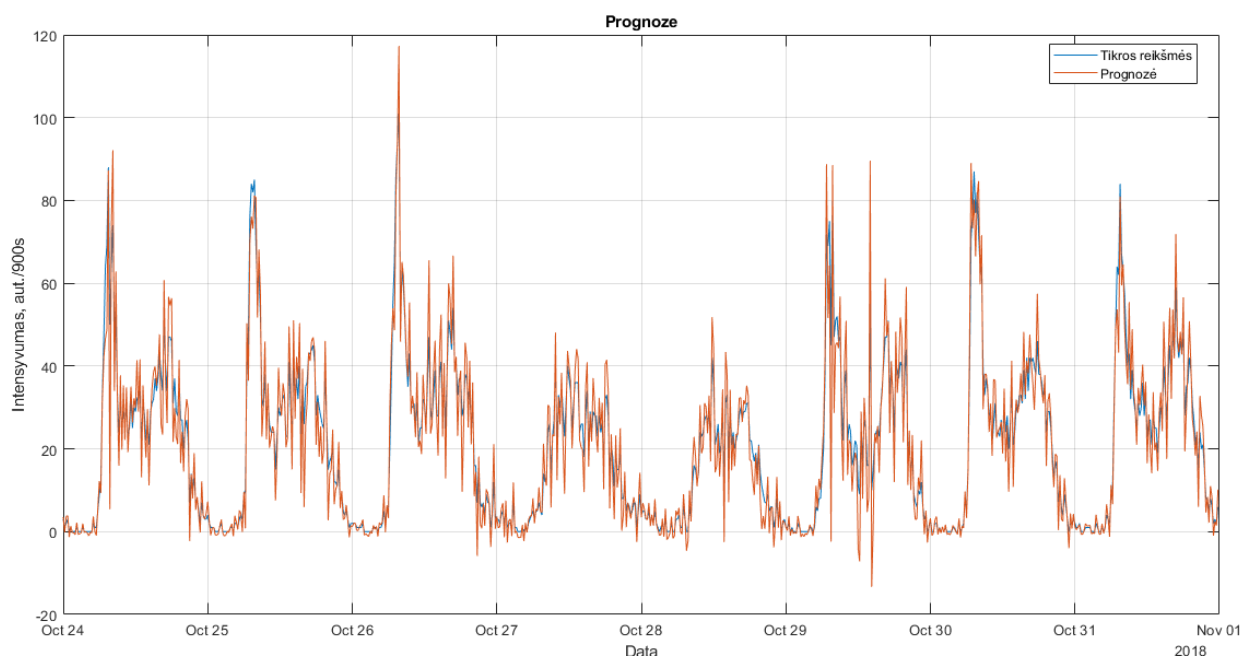
3.10 lentelė. Skirtingų eismo juostų prognozių paklaidos

Jutiklis	Prognozės modelis														
	1.SARIMA(1,0,1) (0,1,1) ⁶⁷²			2.SARIMA(0,0,5) (0,1,1) ⁶⁷²			3.SARIMA(2,1,2) (0,1,1) ⁶⁷²			4.SARIMA(1,1,2) (0,1,1) ⁶⁷²			5. LSTM		
	MA PE	MA E	RM SE	MA PE	MA E	RM SE	MA PE	MA E	RM SE	MA PE	MA E	RM SE	MA PE	MA E	RM SE
VD2.1 ¹	17,6	3,88	5,83	17,9 6	3,88	5,83	17,8 8	3,89	5,84	17,8 8	3,88	5,84	24,3 6	3,94	5,66
VD2.2	24,4 3	3,82	5,89	24,5 4	3,83	5,89	29,2 5	3,93	5,96	29,2 9	3,9	5,96	33,3 1	3,99	6,05
VD 2.3	23,0 8	3,81	6,17	23,1 9	3,81	6,16	25,7 6	3,92	6,32	25,8 4	4,02	6,49	35,4	4,03	6,29
VD 4.1	25,8 3	3,21	4,96	25,7	3,21	4,96	27,1 6	3,26	5,02	27,7	3,26	5,03	32,7 2	3,5	5,17
VD 4.2	24,7 5	3,11	5,13	24,3 5	3,09	5,11	24,0 6	3,08	5,14	26,7 2	3,18	5,19	34,1 6	3,32	5,21
VD 6.1	25,7 3	4,74	7,72	25,5 8	4,72	7,7	36,1 8	4,96	7,81	38,6 5	5,02	7,8	33,6 8	4,64	7,41
VD 6.2	25,7 3	4,74	7,72	25,5 8	4,72	7,7	36,1 8	4,96	7,81	38,6 5	5,02	7,8	31,4 6	4,29	6,84
VD 6.3	24,0 6	3,08	4,83	24,2 4	3,05	4,8	21,3 8	3,01	4,78	21,4 5	3	4,77	30,8 3	2,94	4,27
VD 8.1	23,5 5	2,29	3,39	23,8 3	2,28	3,39	22,9 3	2,26	3,37	22,9 9	2,26	3,37	34,4 2	2,52	3,48
VD 8.2	29,3 4	2,47	3,72	28,9 9	2,46	3,72	28,7 6	2,45	3,7	28,7 2	2,45	3,7	38,2 9	2,53	3,61

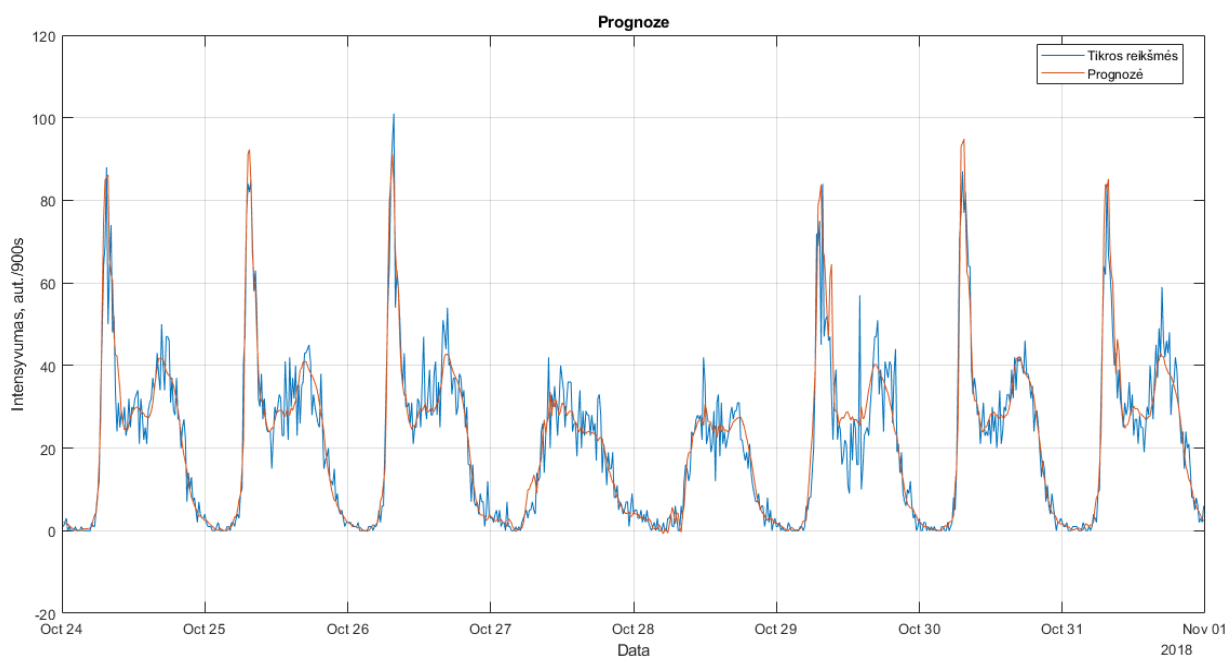
Vizualiai vertinant SARIMA(1,0,1)(0,1,1) prognozuojamą Justiniškių g. (link Laisvės pr.) antros eismo juostos intensyvumą (VD2.2 jutiklis), pastebimai padidėjo svyravimai apie realius duomenis (3.20 pav.). Prognozės modeliui sunkiai sekėsi išlaikyti vidutines reikšmes tarp pikų, kai eismo srautai būna gana stabilūs. Besikeičiant eismo srautams, didėjant ar mažėjant intensyvumui, modelis pervertino pokyčius, nuokrypiu nuo realių duomenų buvo didesni. Kai kuriais atvejais pastebimas nesutapimas, kai realių duomenų kreivė pradeda kilti, o prognozės modelis vis dar prognozuoja mažesnio intensyvumo vertes arba atvirkščiai, pavyzdžiui spalio 29 dienos vidurdienį prognozės modelis neįvertino stabilius eismo srautus dienos metu, prognozuojamo intensyvumo vertės buvo mažesnės, netgi viršijo nulio ribą. Tikrajam intensyvumui pradėjus tolygiai didėti, prognozuojama reikšmė greitai kilo į viršų ir pervertino realus duomenis.

LSTM modelio prognozės kreivė tolygi (3.21 pav.), tačiau yra nedidelių šuolių. Modelio paklaidos smarkiai išaugo, siekė 33–38%. Modelis geriau vertino Justiniškių g. didesnio intensyvumo eismo juostas, prastai vertino nedidelio intensyvumo Virbeliškių g. duomenis.

¹ Naudojant VD2.1 duomenis buvo sudaryti SARIMA modeliai. Šio jutiklio rezultatai pateikti palyginti kitų jutiklių paklaidas.



3.20 pav. SARIMA(1,0,1)(0,1,1) modelio prognozė antrai Justiniškių g. (link Laisvės pr.) eismo juostai (VD2.2 jutiklis)



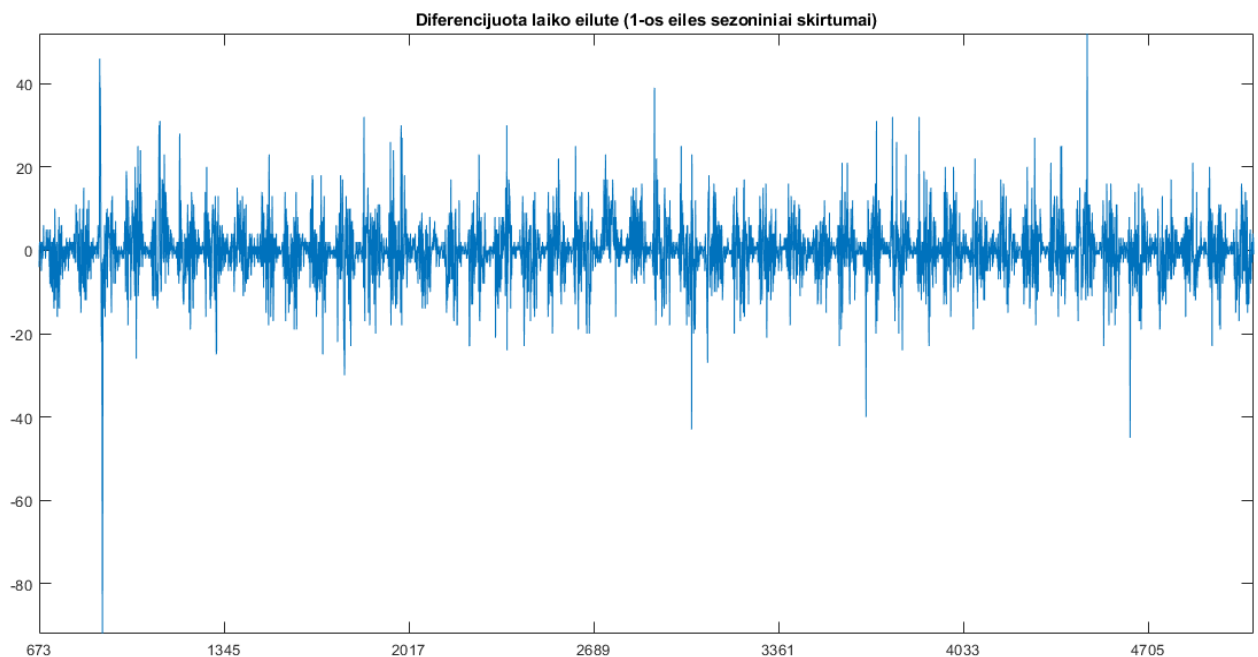
3.21 pav. LSTM modelio prognozė antrai Justiniškių g. (link Laisvės pr.) eismo juostai (VD2.2 jutiklis)

Kitų jutiklių intensyvumo prognozės grafikai pateikti 8 priede.

3.2.3 SARIMA modelio įvertinimas su kito jutiklio duomenimis

SARIMA(1,0,1)(0,1,1) modelio struktūra buvo nustatyta naudojant Justiniškių g. (link Laisvės pr.) pirmos eismo juostos duomenis (VD2.1 jutiklio). Taikant prognozės modelį gretimai eismo juostai (VD2.2 jutiklio duomenims), paklaida padidėjo iki 24,4%. Skirtumas yra reikšmingas, vertinant kad šių eismo juostų duomenys yra artimi dėl vienodai taikomų eismo valdymo signalų ir pasižymi panašiomis intensyvumo vertėmis. Kadangi VD2.2 jutiklio duomenys turėtų būti artimiausi VD2.1 jutiklio duomenimis, atliekama šių duomenų detalesnė analizė.

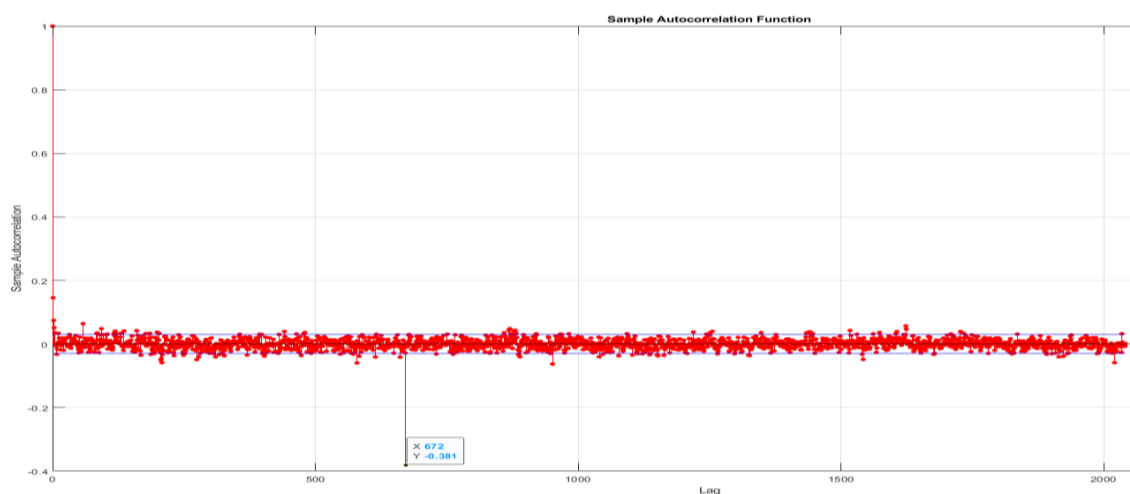
Pastebima, kad VD2.2 jutiklio sezoninių skirtumų grafikas yra ganėtinai stabilus, tačiau yra pikų, kurių amplitudės labai didelės, yra skirtumų, kurių vertė viršija 40 automobilių tarp matavimų (3.22 pav.). Vertinant, kad vidutiniškai intensyvumas siekia apie 30 automobilių per nagrinėjama laiko intervalą, o pikų reikšmės siekia apie 100–120 automobilių, šie skirtumai yra reikšmingi ir rodo didesnę duomenų dispersiją nei praeitame skyriuje vertinto VD2.1 jutiklio duomenų (3.7 pav.).



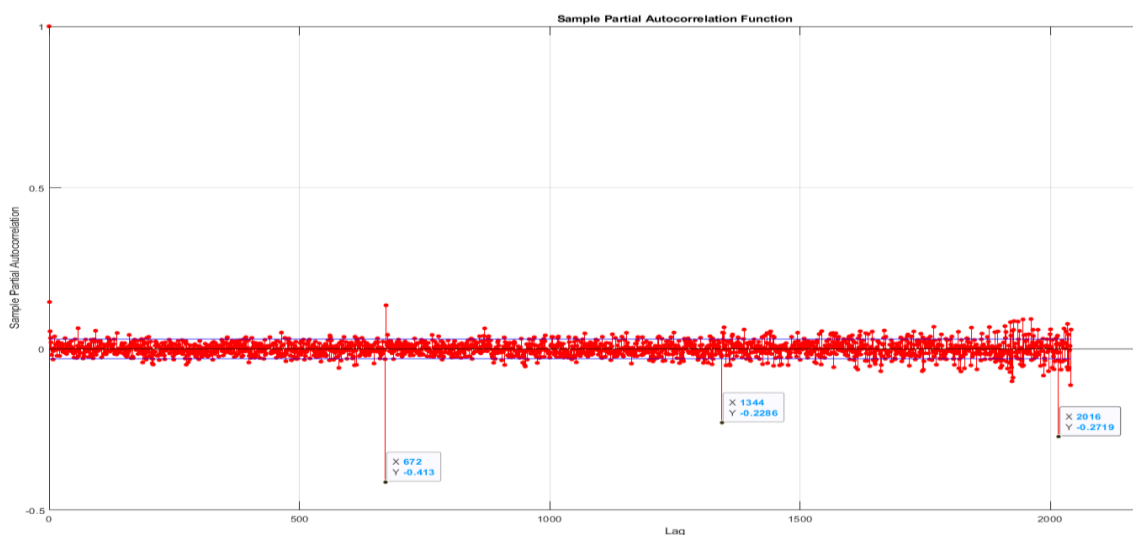
3.22 pav. Sezoniniškai diferencijuotų VD2.2 jutiklio duomenų kreivė

VD2.2 jutiklio autokoreliacijos funkcija (3.23 pav.) nutrūksta po 4 vėlavimo, tačiau žymus kritimas matomas jau po 1 vėlavimo. Panaši tendencija yra ir VD2.1 jutiklio sezoniškai diferencijuotų duomenų autokoreliacijos funkcijoje (3.8 pav.). Sezoninė koreliacija yra tik su 1, sezoną atitinkančiu vėlavimu (672 vėlavimas). Dalinė autokoreliacija (3.24 pav.) nutrūksta po 2 vėlavimo, kurio koreliacijos vertė yra artima pasikliautinojo intervalo ribai, o didžiausias trūkis matomas iškart po 1 vėlavimo. Sezoninė autokoreliacija gęsta lėtai, pastebimi pasikartojantys pikai, atitinkantis sezono trukmę.

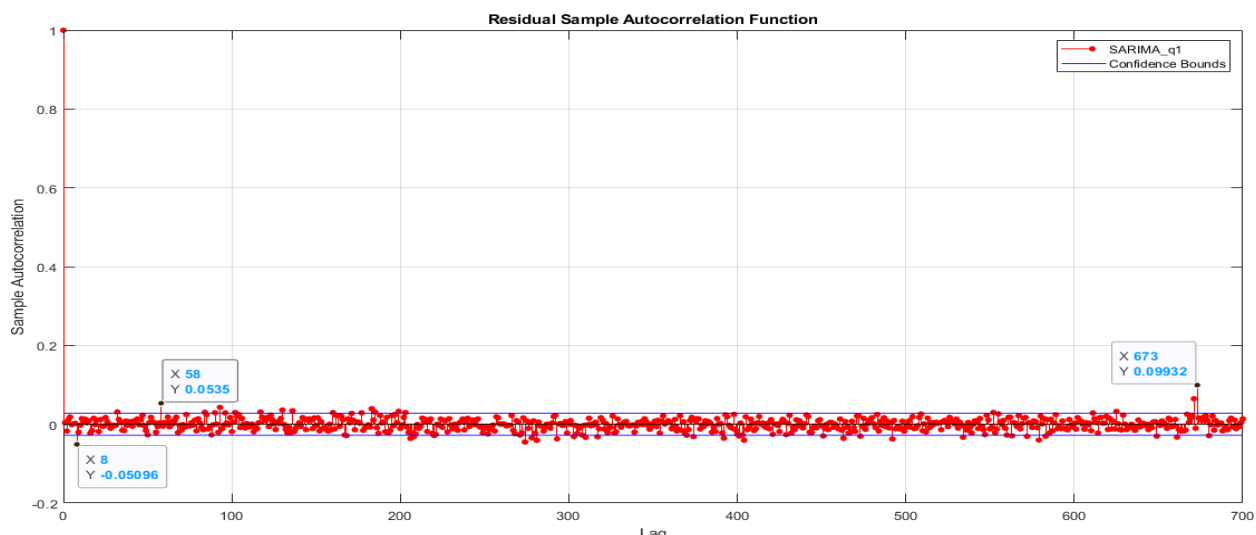
Remiantis koreliogramomis, VD2.2 jutiklio duomenims būtų sudaromas analogiškos struktūros SARIMA modelis, t. y. SARIMA(1,0,1)(0,1,1).



3.23 pav. VD2.2 jutiklio sezoniškai diferencijuotų duomenų autokoreliacijos funkcija



3.24 pav. VD2.2 jutiklio sezoniškai diferencijuotų duomenų dalinės autokoreliacijos funkcija



3.25 pav. SARIMA modelio liekamųjų paklaidų autokoreliacija su VD2.2 jutiklio duomenimis

3.11 lentelė. Ljung–Box liekamųjų paklaidų koreliacijos testo rezultatai (VD2.2 jutiklis)

Modelis	H0 hipotezė	Patikimumo lygmuo	Tikrinti vėlavimai	p reikšmė
SARIMA(1,0,1)(0,1,1) ₆₇₂	Atmesta	0,05	30	0,322
			700	1,438e-09

SARIMA(1,0,1)(0,1,1) modelio liekamųjų paklaidų autokoreliacija (3.25 pav.), naudojant VD2.2 jutiklio duomenis, panaši į VD2.1 duomenų (3.12 pav.). Tačiau pastebimi keli išsiskiriantys koreliuoti vėlavimai: 8, 56, 671, 673. Atlikus Ljung–Box liekamųjų paklaidų autokoreliacijos testą (3.11 lentelė), artimų vėlavimų liekamųjų paklaidų koreliacija nereikšminga, tačiau tikrinant sezoną apimančius vėlavimus, buvo nustatyta reikšminga tiriamo modelio liekamųjų paklaidų autokoreliacija. Dėl to SARIMA(1,0,1)(0,1,1) modelis neturėtų būti taikomas VD2.2 jutiklio duomenimis.

3.2.4 Dalinės išvados

1. Apibendrinant šio poskyrio rezultatus, vienodo SARIMA prognozės modelio taikymas skirtingų eismo juostų intensyvumui prognozuoti – nėra galimas. Tai rodo padidėjusios SARIMA modelių paklaidos, šiuos modelius taikant kitų eismo juostų intensyvumo prognozavimui.
2. Atsižvelgiant į tai, kad LSTM modelis buvo parengtas nepriklausomai nuo ankstesnių rezultatų, modelio mokymosi metu kiekvieno jutiklio duomenys įvertinti nenaudojant ankčiau sudarytų priklausomybių, padidėjusios šio modelio paklaidos rodo, kad kitų jutiklių duomenys yra galimai mažiau stacionarus dėl cikliškų svyravimų, didelės duomenų dispersijos. Dėl to šios sankryžos kiekvienos eismo juostos duomenys ir jų tinkamumas prognozės modeliui sudaryti turi būti vertinama individualiai.

3.3 Prognozuojamo laikotarpio įvertinimas

Uždavinys: Įvertinti prognozuojamo laikotarpio poveikį prognozės tikslumui.

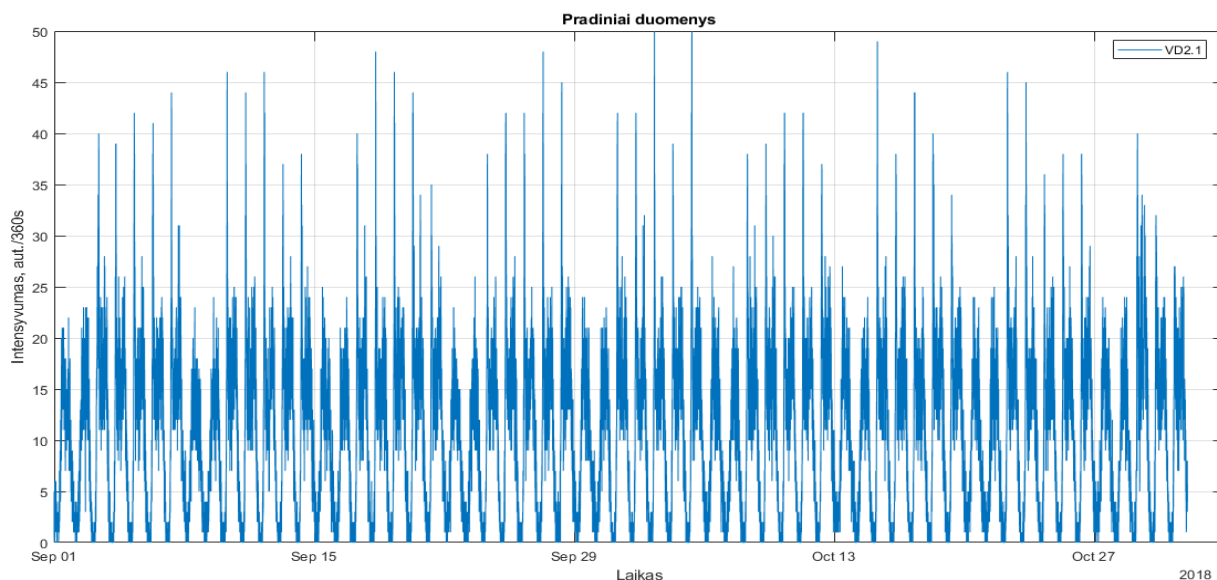
Metodika: Prognozuojamo laikotarpio trukmė yra mažinama performuojant pradinę laiko eilutę. Prognozavimas yra vykdomas vienu laiko žingsniu, atitinkančiu laiko intervalą tarp duomenų matavimų. Todėl sumažinama šio intervalo trukmė, padidinant duomenų fiksavimo dažnį. Performuotos laiko eilutės yra taikomos prognozės modelių įvertinimui. Tikrinama, kaip pasikeitė prognozės paklaidos. Atliekamas paklaidų palyginimas su kitais SARIMA bei LSTM modeliais.

Trumpas aprašas

Planuojant prognozės modelių integraciją į eismo valdymo sistemą, būtina įvertinti periodiškumą, kuriuo bus atliekamas prognozavimas ir šviesoforų darbo ciklo optimizavimas. Prognozuojamo laikotarpio trukmė gali būti keičiama prognozuojant kelis žingsnius į priekį ar keičiant vieno prognozuojamo žingsnio ilgį. Sudarant vieno žingsnio prognozes, prognozuojamas laikotarpis atitinka laiko trukmę tarp duomenų matavimų. Siekiant pakeisti šio laikotarpio trukmę, reikalinga performuoti turimų duomenų imtį, keičiant intervalus tarp matavimų, t. y. didinant duomenų fiksavimo dažnį.

Remiantis teorinėje dalyje aptartu prognozės patikimumo apibrėžimu, sumažinus prognozuojamo laikotarpio trukmę, prognozės tikslumas turėtų padidėti, kadangi sumažėja netikėtų pokyčių tikimybė. Tuo pačiu padaugėja ir duomenų kiekis, kurį turi įvertinti prognozės modelis. Dėl to ilgiau užtrunka duomenų įvertinimas, ilgėja prognozės skaičiavimo trukmė ir reikalinga daugiau kompiuterinių resursų, siekiant užtikrinti prognozės modelių greitaveiką.

Šiame skyriuje siekiama įvertinti galimybę naudoti sudarytus modelius prognozuojant trumpesnio laikotarpio intensyvumo vertes ir bandoma nustatyti, kaip keisis prognozės tikslumas. Tyrimui naudojami pradiniai VD2.1 jutiklio duomenys. Šių duomenų laiko eilutė yra perskaičiuojama 540 s (9 min) ir 360 s (6 min) laiko intervalais. Kadangi laiko eilutė yra performuojama, pasikeičia ir duomenų kiekis, fiksuojamas per vieną savaitę. Dėl to keičiama SARIMA modelio sezono trukmė. Esant 540 s intervalais suformuotai laiko eilutei, sezoną atitinka 1120 reikšmės, o 360 s – 1680 reikšmės.



3.26 pav. 360 s intervalu suformuota laiko eilutė

Sumažinus laiko eilutės intervalų trukmę, tarp matavimų fiksuojama mažiau transporto priemonių (3.26 pav.). Dėl to intensyvumo skalė yra mažesnė, labiau juntami skirtumai tarp gretimų matavimų, pastebimi aukšto dažnio svyravimai. Remiantis 360 s intervalais suformuota laiko eilutė, vidurdienio metu intensyvumas vidutiniškai svyruoja nuo 10 iki 25 aut./6 min, piko metu siekia apie 40–50 aut./6 min.

3.3.1 Prognozės rezultatai

Prognozuojant trumpesnę laikotarpį, MAPE paklaida padidėjo (3.12 lentelė). Sumažinus prognozuojamą laikotarpį iki 540 s (6 min), SARIMA modelių paklaidos padidėjo 3%, LSTM modelio – 1%. Prognozuojant 360 s laikotarpį, visų modelių paklaidos padidėjo 6%.

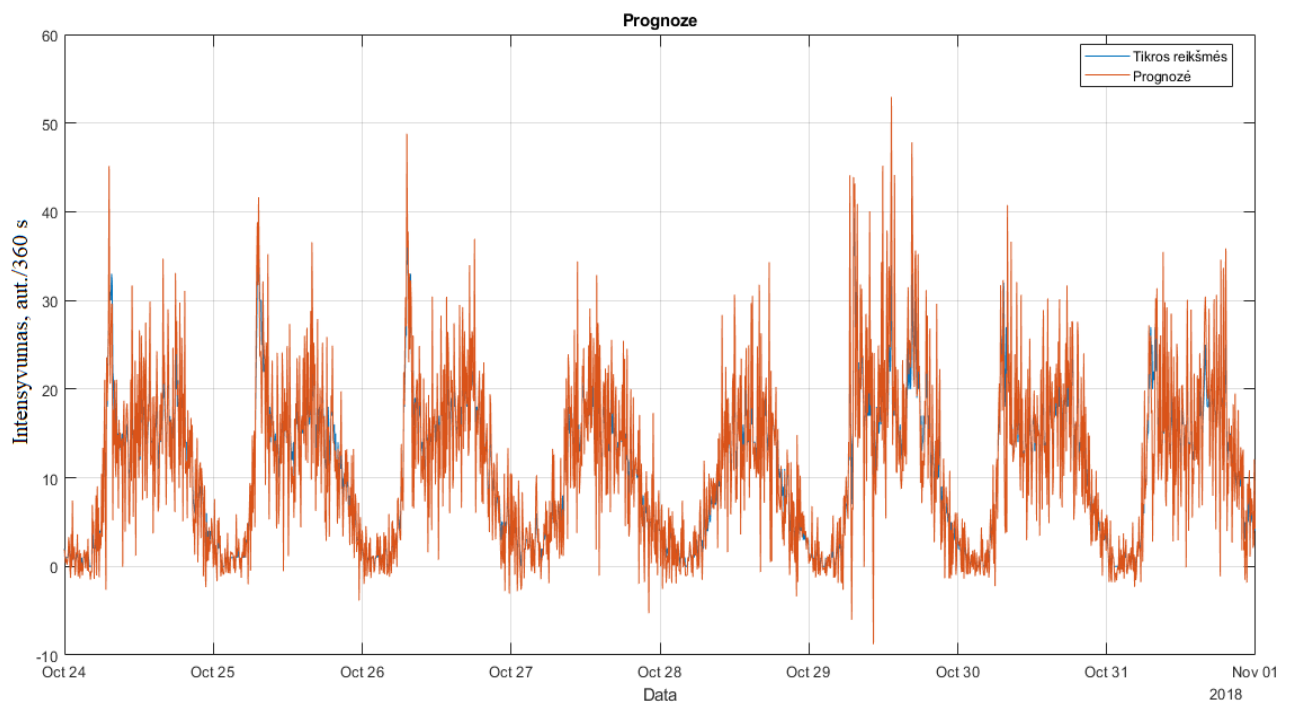
Performavus laiko eilutę, per trumtesnę laikotarpį yra fiksuojama daugiau matavimų. Dėl to prognozės modelio įvertinimo ir prognozavimo metu turėjo būti įvertinama daugiau duomenų. Prognozavimo trukmė padidėjo, tačiau šis laikas yra daug mažesnis nei tiriami laiko intervalai ir nėra detalčiau nagrinėjamas (3.12 lentelė). Tačiau modelio įvertinimo trukmė reikšmingai padidėjo. Prognozuojant 540 s žingsniu, visų modelių įvertinimo trukmė padidėjo apie 2–2,5 karto. Prognozuojant 360 s žingsniu, SARIMA modelių įvertinimo trukmė padidėjo 7–7,5 karto, LSTM modelio – apie 3 kartus. Remiantis šiais rezultatais, tik SARIMA(1,0,1)(0,1,1) modelis gali būti pakartotinai perskaičiuojamas realiuoju laiku, prognozuojant 540 s laikotarpį, kadangi jo įvertinimo trukmė šiuo atveju buvo mažesnė nei prognozuojamas laikotarpis. Visų kitų modelių atnaujinimo procesas turėtų vykti nepriklausomai nuo prognozavimo nustatytu periodiškumu.

3.12 lentelė. Trumpesnio laikotarpio prognozės rezultatai

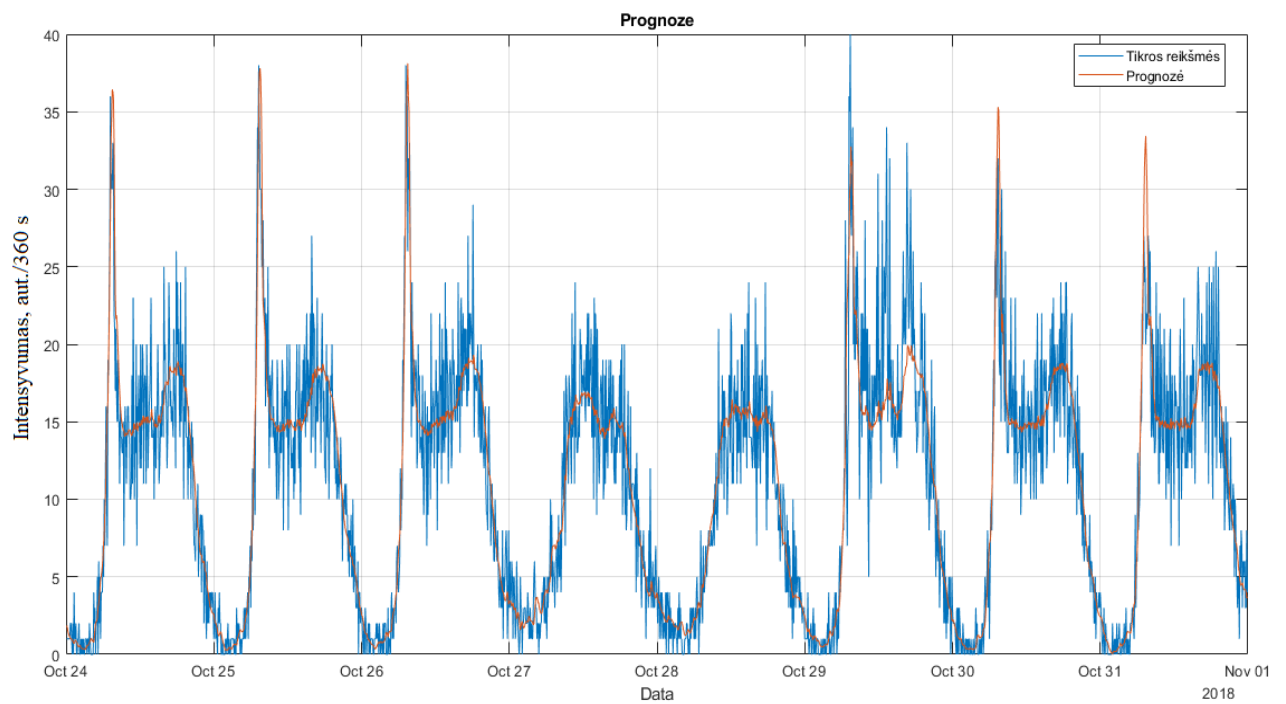
Prognozuojamas laikotarpis	Modelis	MAPE, %	MAE	RMSE	Modelio greitaveika	
					Modelio įvertinimo trukmė, s	Vienos reikšmės prognozavimo trukmė, s
900 s	1.SARIMA(1,0,1)(0,1,1) ₆₇₂	17,60	3,88	5,83	180	0,01
	2.SARIMA(0,0,5)(0,1,1) ₆₇₂	17,96	3,88	5,83	311	0,01
	3.SARIMA(2,1,2)(0,1,1) ₆₇₂	17,88	3,89	5,84	470	0,02
	4.SARIMA(1,1,2)(0,1,1) ₆₇₂	17,88	3,88	5,84	407	0,01
	5. LSTM	24,36	3,94	5,66	1065	0,03
540 s	1.SARIMA(1,0,1)(0,1,1) ₁₁₂₀	20,52	2,79	4,09	526	0,03
	2.SARIMA(0,0,5)(0,1,1) ₁₁₂₀	20,48	2,78	4,08	994	0,03
	3.SARIMA(2,1,2)(0,1,1) ₁₁₂₀	22,04	2,8	4,11	1098	0,03
	4.SARIMA(1,1,2)(0,1,1) ₁₁₂₀	21,8	2,8	4,11	1133	0,03
	5. LSTM	25,97	2,81	3,99	1807	0,04
360 s	1.SARIMA(1,0,1)(0,1,1) ₁₆₈₀	23,06	2,24	3,29	1426	0,07
	2.SARIMA(0,0,5)(0,1,1) ₁₆₈₀	23,2	2,25	3,31	2973	0,08
	3.SARIMA(2,1,2)(0,1,1) ₁₆₈₀	23,53	2,25	3,3	4138	0,07
	4.SARIMA(1,1,2)(0,1,1) ₁₆₈₀	23,71	2,26	3,31	2794	0,07
	5. LSTM	30,8	2,35	3,23	3195	0,03

SARIMA(1,0,1)(0,1,1) prognozės grafike pastebimi dideli svyravimai (3.27 pav.). Modelis nustatė pikus ir prognozių laiko momentas nepersislinkęs, tačiau vidurdienį prasčiau įvertinti nusistovėję srautai ir prognozės kreivė pasižymėjo labai dideliais svyravimais apie realius duomenis. Pastebimas tam tikras realių duomenų pervertinimas, pavyzdžiui spalio 29 d. SARIMA(1,0,1)(0,1,1) modelis, reaguodamas į mažėjantį eismo srautą, prognozavo per mažą intensyvumą ir netgi viršijo nulio ribą. Modelis neturi apribojimo veikti tik natūraliųjų skaičių aibėje, prognozuojamos neigiamos reikšmės yra keičiamos į 1 (plačiau šie modelio ribojimai aptarti sudarant pirminį modelį, 3.1.6 skyrelyje).

LSTM modelio prognozės kreivė tolygesnė nei SARIMA modelio (3.28 pav.). Šiame grafike gerai matomi realių duomenų aukšto dažnio svyravimai, triukšmas. LSTM modelis nesistengė atkartoti šių svyravimų, nustatė vidutines reikšmes ir jų laikėsi. Dėl to vizualiai ši kreivė atrodo tvarkingiau, tačiau dėl skirtumų su realiais duomenimis susidarė didesnės modelio paklaidos.



3.27 pav. SARIMA(1,0,1)(0,1,1) modelio prognozė, prognozuojant 360 s žingsniu



3.28 pav. LSTM modelio prognozė, prognozuojant 360 s žingsniu

3.3.2 Dalinės išvados

1. Sumažinus prognozuojamo laikotarpio trukmę, prognozės modelių paklaidos padidėjo. Prognozuojant 540 s laikotarpį, SARIMA modelių MAPE paklaidos padidėjo 3%, LSTM 1%. Prognozuojant 360 s laikotarpį, visų modelių paklaidos padidėjo 6%.
2. Prognozuojamo laikotarpio trukmė buvo mažinama perfuormuojant pradinę laiko eilutę, keičiant laiko intervalus tarp matavimų. Per vieną matavimą buvo fiksuojamas mažesnis transporto priemonių skaičius, dėl to sumažėjo intensyvumo vertės amplitudė, buvo labiau juntami skirtumai tarp matavimų. Reiktų įvertinti, kad padidėjusias paklaidas galėjo lemti ši aplinkybė.
3. Sumažinus prognozuojamo laikotarpio trukmę, prognozės paklaidos viršijo 20% ribą. Todėl prognozės turėtų būti vertinamos tik kaip patenkinamos ir netinkamos realiam taikymui.
4. Realiose duomenyse buvo pastebėti dideli svyravimai tarp gretimų matavimų. Tai sukelia netolygumą laiko eilutėje, dispersiją ir pablogėjusį prognozės modelių tikslumą. Sumažinus šių svyravimų poveikį modelių tikslumas turėtų būti tinkamas prognozuojant trumpesnę laikotarpį. Svyravimų poveikis gali būti sumažintas taikant duomenų glodinimo metodus (šių metodų taikymas atliekamas 3.5 poskyryje).
5. Vertinant modelių greitaveiką, prognozavimas gali būti atliekamas realiuoju laiku, kadangi vienos reikšmės prognozavimas yra daug trumpesnis nei prognozuojamas laikotarpis. Tačiau padidėjęs reikšmių kiekis lėmė pailgėjusią modelio įvertinimo trukmę ir dauguma tirtų modelių nebegalėtų atlikti pakartotino modelio įvertinimo prognozavimo metu. Esant 540 s prognozuojamo laikotarpio trukmei, SARIMA(1,0,1)(0,1,1) modeliu galima atlikti pakartotiną regresijos koeficientų perskaičiavimą realiuoju laiku, tačiau kitų modelių atveju pakartotinis modelių įvertinimas turi būti vykdomas atskirai nuo prognozavimo, nustatytu periodiškumu.

3.4 Modeliui sudaryti reikalingo duomenų kiekio įvertinimas

Uždavinys: Įvertinti duomenų kiekį, reikalingą prognozės modeliui sudaryti.

Metodika: SARIMA ir LSTM modelių įvertinimo procesas atliekamas su skirtingą laikotarpį fiksuotų duomenų imtimis: 1,5 mėn., 1 mėn., 2 sav. Palyginama, kaip modelių tikslumas ir greitaveika priklauso nuo pradinių duomenų imties dydžio.

Trumpas aprašas

Įvertinimo etape modelis yra supažindinamas su duomenimis. SARIMA modelių atveju, yra nustatomos regresijos lygties koeficientų vertės. LSTM atveju, atliekamas modelio mokymas, modelis susipažįsta su egzistuojančiais dėsningumais duomenyse, sudaromas matematinis mechanizmas, padedantis atpažinti skirtingas laiko eilutės būsenas.

Įvertinimas yra atliekamas pirmą kartą sudarant modelį bei periodiškai atnaujinant turimus duomenis, siekiant išlaikyti prognozės modelių adekvatumą. Šio proceso kokybė labai priklauso nuo sukauptos duomenų imties. Didelė duomenų imtis užtikrina tinkamą duomenų įvertinimą. SARIMA modelio atveju, mažėja klaidos tikimybė nustatant regresijos koeficientus. LSTM modelio atveju, modelis geriau išmoka prisitaikyti prie duomenų kaitos. Tačiau didinant duomenų imtį daugėja reikšmių kiekis, kurias reikia įvertinti. Dėl to ilgėja modelių įvertinimo (mokymo) laikas. Šis laikas svarbus periodiškai atnaujinant modelius, kadangi remiantis modelio įvertinimo trukme nustatomas minimalus periodiškumas, kuriuo būtų galima atlikti pakartotiną modelio atnaujinimą. Esant nedideliai modelio įvertinimo trukmei, modelį galima atnaujinti mažesniu periodiškumu, o kai kuriais atvejais ir realiuoju laiku atliekant prognozavimą.

Vertinant fizinius duomenų ribojimus, SARIMA modelių atveju yra naudojamas aiškiai apibrėžtas vėlavimų skaičius, slenkantis laiko eilute. Šis vėlavimų skaičius ir nustato, koks minimalus skaičius reikšmių turi sudaryti pradinių duomenų imtį (3.2 formulė). LSTM modelio atveju, nėra griežtų ribojimų, nurodančių minimalų pradinių duomenų imties dydį. Tačiau remiantis Nagy, Simon (2018), šie modeliai veikia stabiliai ir tinkamai tik tada, kai yra susipažinę su kuo didesniu įvairių atvejų skaičiumi ir geba atpažinti skirtingas eismo srautų būsenas.

Šioje darbo dalyje siekiama nustatyti, kaip kinta prognozės modelių tikslumas, mažinant pradinių duomenų imtį, skirtą modelių įvertinimui. SARIMA modelių struktūra nėra keičiama (taikomi 3.1 poskyryje sudaryti modeliai). Modelių validacija atliekama su tais pačiais duomenimis, fiksuotais 2018-10-24 – 2018-10-31 laikotarpiu. Modelio įvertinimo (mokymo) imtis nustatoma atbuline data nuo spalio 24 dienos.

Minimali duomenų imtis, reikalinga SARIMA modeliui sudaryti, priklauso nuo naudojamo modelio parametrų sumos (Hyndman, Kostenko 2007):

$$n_{min} = p + d + q + P + Q + s \cdot D \quad (3.2)$$

čia p – autoregresinio proceso (AR) eilė;
 d – integravimo eilė (nesezoninio);
 q – slankiųjų vidurkių proceso (MA) eilė;
 P – sezoninio autoregresinio proceso (SAR) eilė;
 Q – sezoninio slankiųjų vidurkių proceso (SMA) eilė;
 s – sezono trukmė;
 D – sezoninio diferencijavimo eilė.

SARIMA(1,0,1)(0,1,1)₆₇₂ modeliui reikia:

$$n_{min} = 1 + 0 + 1 + 0 + 672 + 672 \cdot 1 = 1346 \text{ reikšmių}$$

3.4.1 Prognozės rezultatai

Modeliui sudaryti naudojant 1 mėn. laikotarpio duomenis, SARIMA modelių paklaidos sumažėjo 0,1–0,3% (3.13 lentelė). Žvelgiant į perskaičiuotus SARIMA(1,0,1)(0,1,1) modelio regresijos koeficientus pastebima, vieno mėnesio regresijos koeficientų vertės yra didesnės už 1,5 mėn. (3.14 lentelė). LSTM modelio atveju, naudojant 1 mėn. laikotarpio duomenis pastebimas reikšmingas paklaidos padidėjimas nuo 24% iki 39%.

SARIMA(1,0,1)(0,1,1) modelio atveju, minimali duomenų imtis apima dviejų savaitių laikotarpiu fiksuotus duomenis. Vertinant šią imtį, modelio paklaida padidėjo iki 19,2%. Žvelgiant į modelio regresijos koeficientų vertes (3.14 lentelė), autoregresinio proceso dalies AR koeficientas tapo neigiamu, MA dalies teigiamu. Skaitinių verčių modulis šiek tiek sumažėjo. Sezoninės dalies (SMA) koeficiento vertė išliko panaši kaip ir pradiniu atveju. Modeliui sudaryti naudojant 2 sav. duomenis, tiksliausiai veikė SARIMA(0,0,5)(0,1,1) modelis, kurio MAPE paklaida siekė 18,9%. Apibendrinant, SARIMA modelių tikslumas išliko panašus. LSTM modelio atveju, MAPE paklaida padidėjo iki 51%, t. y. beveik dvigubai, lyginant su pradiniu atveju, modelio mokymui naudojant 1,5 mėn. laikotarpio duomenis.

Vertinant modelių greitaveiką, SARIMA(0,1,1)(0,1,1) modelio atveju didelio skirtumo nebuvo. Esant minimaliai 2 sav. duomenų imčiai, modelio įvertinimo procesas truko 172 s, o tai vos 8 s greičiau nei pradiniu atveju. Žvelgiant į kitus SARIMA modelius, SARIMA(1,1,2)(0,1,1) modelio atveju pastebimas laiko trukmės sumažėjimas nuo 407 s (esant 1,5 mėn. duomenims) iki 258 s (esant 2 sav. duomenims). Kitų SARIMA modelių rezultatai buvo prieštaringi. SARIMA(0,0,5)(0,1,1) modelio atveju, esant 1,5 mėn. laikotarpio duomenims, modelio įvertinimas

truko 311 s, o 1 mėn. ir 2 sav. duomenų atveju – modelio įvertinimo laikas pailgėjo, atitinkamai iki 318 ir 337 s. SARIMA(2,1,2)(0,1,1) modelio atveju, esant 1 mėn. laikotarpio duomenims, modelio įvertinimo laikas sutrumpėjo iki 304 s, tačiau esant 2 sav. imčiai – pailgėjo iki 407 s. Skirtumai nėra dideli ir juos galėjo lemti kompiuterio resursų išnaudojimas. Visi bandymai buvo atliekami panašiomis sąlygomis, tačiau kompiuterio resursų paskirstymas nebuvo fiksuojamas, todėl galima klaidinga rezultatų interpretacija.

LSTM modelio atveju skirtumas tarp modelio mokymo trukmės yra didelis. Esant 1 mėn. laikotarpio duomenims, mokymo laikas sutrumpėjo iki 633 s, o esant 2 sav. – iki 316 s. Tačiau pastebimas reikšmingas šio modelio tikslumo sumažėjimas.

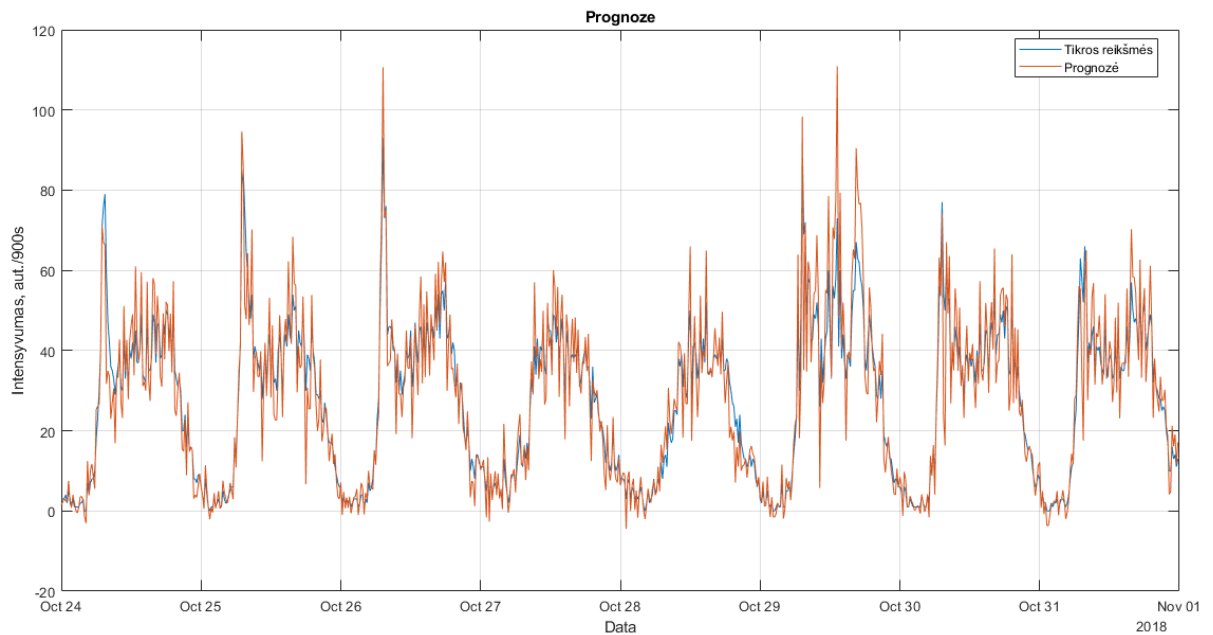
3.13 lentelė. Prognozės rezultatai, modeliui sudaryti taikant skirtingo laikotarpio duomenų imtį

Duomenų fiksavimo laikotarpis	Duomenų imties pradžia	Duomenų imties pabaiga	Modelis	MAPE, %	MAE	RMSE	Modelio greitaveika	
							Modelio įvertinimo trukmė, s	Vienos reikšmės prognozavimo trukmė, s
1,5 mėn.	01/09/2018 00:00	23/09/2018 23:59	1.SARIMA(1,0,1)(0,1,1)	17,60	3,88	5,83	180	0,01
			2.SARIMA(0,0,5)(0,1,1)	17,96	3,88	5,83	311	0,01
			3.SARIMA(2,1,2)(0,1,1)	17,88	3,89	5,84	470	0,02
			4.SARIMA(1,1,2)(0,1,1)	17,88	3,88	5,84	407	0,01
			LSTM	24,36	3,94	5,66	1065	0,03
1 mėn.	23/09/2018 00:00		1.SARIMA(1,0,1)(0,1,1)	17,53	3,95	6	172	0,01
			2.SARIMA(0,0,5)(0,1,1)	17,64	3,95	6,01	318	0,03
			3.SARIMA(2,1,2)(0,1,1)	17,47	3,96	6,01	304	0,01
			4.SARIMA(1,1,2)(0,1,1)	17,56	3,97	6,08	341	0,04
			LSTM	39,42	2,66	3,84	633	0,03
2 sav.	09/10/2018 00:00		1.SARIMA(1,0,1)(0,1,1)	19,24	4,22	6,43	172	0,01
			2.SARIMA(0,0,5)(0,1,1)	18,87	4,18	6,64 ₅	337	0,01
			3.SARIMA(2,1,2)(0,1,1)	21,22	4,24	6,39	407	0,01
			4.SARIMA(1,1,2)(0,1,1)	21,64	4,27	6,43	258	0,01
			LSTM	51,23	3,37	5,31	316	0,02

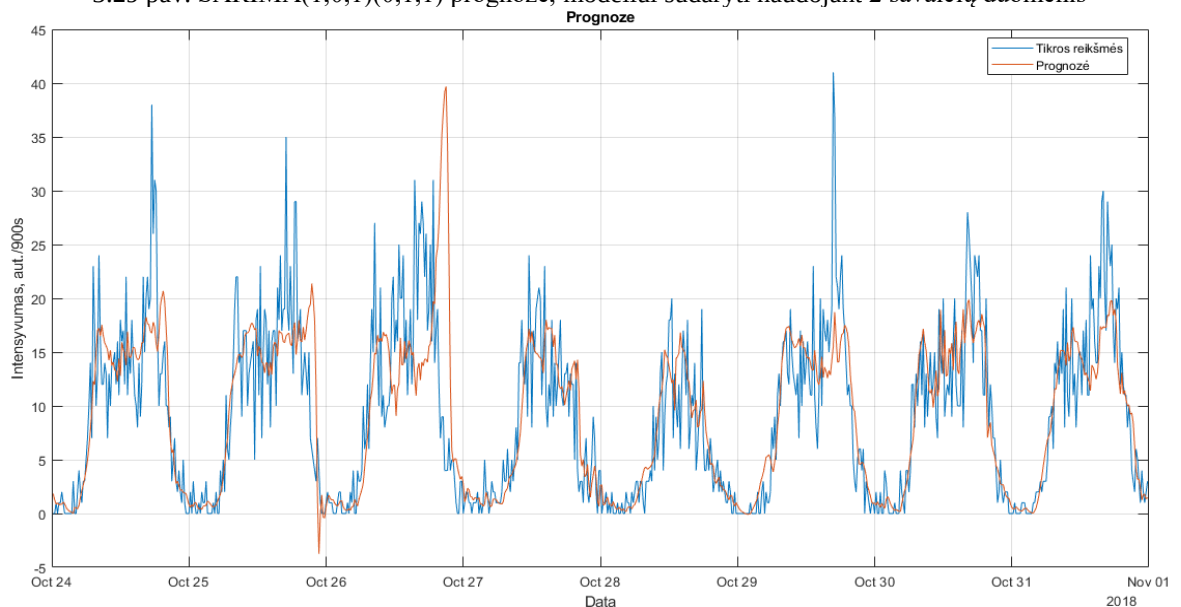
3.14 lentelė. SARIMA(1,0,1)(0,1,1) modelio regresijos lygties koeficientai

	1,5 mėn.	1 mėn.	2 sav.
Constant	0	0	0
AR{1}	0.71897	0,88138	-0,46291
MA{1}	-0.60654	-0,840111	0,424813
SMA{672}	-0.59605	-0,58575	-0,545165
Variance	288.954	24,9758	21,8721

Vizualiai vertinant SARIMA(1,0,1)(0,1,1) modelio prognozės grafiką (3.29 pav.), modeliui sudaryti naudojant 2 sav. duomenis, pastebimų skirtumų lyginant su pradiniu atveju, naudojant 1,5 mėn. laikotarpio duomenis (3.14 pav.) – nėra. LSTM modelio atveju (3.30 pav.), pastebimi dideli kreivės svyravimai ir prognozės kreivė nebėra tolygi kaip pradiniu atveju (3.15 pav.). Modelis bandė atkartoti svyravimus, tačiau matomos prognozavimo klaidos, kai susidarė dideli nuokrypiai nuo realių duomenų kreivės. Pastebimas prognozės kreivės inertiškumas, per lėtas reagavimas į besikeičiančius eismo srautus.



3.29 pav. SARIMA(1,0,1)(0,1,1) prognozė, modeliui sudaryti naudojant 2 savaitių duomenis



3.30 pav. LSTM prognozė, modeliui sudaryti naudojant 2 savaitių duomenis

3.4.2 Dalinės išvados

1. Sumažinus modeliui sudaryti skirtos duomenų imties dydį SARIMA(1,0,1)(0,1,1) modelio paklaidos padidėjo nedaug, greیتaveika beveik nepakito. Šiuo atveju geriausia naudoti 1 mėn. laikotarpio duomenis, kadangi paklaidos buvo mažesnės, nei su 1,5 mėn. duomenimis, MAPE paklaida siekė 17,5%, o modelio įvertinimo trukmė sumažėjo 8 sekundėmis. Nors tai nesudaro žymaus skirtumo, tačiau sumažinamas archyvuojamų duomenų kiekis, reikalinga mažiau resursų duomenų administravimui.
2. LSTM modelio atveju, sumažinus modelio mokymui skirtos duomenų imties dydį, modelio paklaidos reikšmingai padidėjo, mokymosi laikas sutrumpėjo. Esant 2 sav. laikotarpio duomenims, šio modelio paklaida padidėjo iki 51% MAPE, o mokymosi trukmė sumažėjo iki 316 s. Vertinant šiuos rezultatus, tikėtina pagerinti LSTM modelio tikslumą, modeliui sudaryti naudojant didesnę duomenų imtį.
3. Vertinant palyginimui skirtų SARIMA modelių greیتaveiką buvo pastebėti neatitikimai. Esant trumpesnei duomenų imčiai, šių modelių įvertinimo procesas truko ilgiau nei pradiniu atveju. Skaičiavimo trukmė priklauso nuo skirtų kompiuterinių resursų. Kadangi kompiuteriniai resursai atliekant šį bandymą nebuvo stebimi, vertinti SARIMA modelių įvertinimo trukmę reikėtų atsargiai, kadangi galima klaidinga rezultatų interpretacija. LSTM modelio atveju, modelio mokymo trukmės skirtumai buvo dideli, todėl kompiuterinių resursų įvertinimas nėra toks aktualus.

3.5 Duomenų glodinimo įvertinimas

Uždavinys: Įvertinti glodinimo metodų taikymą, siekiant sumažinti duomenų svyravimus ir padidinti prognozės modelių tikslumą.

Metodika: Atliekamas visos pradinio duomenų laiko eilutės glodinimas. Įvertinamas glodinimo metodų bei taikomo glodinimo pločio poveikis prognozės modelių tikslumui. Naudojamas SARIMA(1,0,1)(0,1,1) modelis, palyginimas atliekamas su kitais, SARIMA bei LSTM modeliais. Nustačius paklaidų sumažėjimą, papildomai vertinamas glodinimo metodų taikymas sudarant trumpesnio laikotarpio prognozes.

Trumpas aprašas

Sudarius pradinį prognozės modelį bei atliekant prognozuojamo laikotarpio poveikio analizę, buvo pastebėti dažni duomenų svyravimai tarp gretimų reikšmių, triukšmas. Šie svyravimai eismo sraute gali susidaryti dėl įvairių aplinkybių, tokių kaip netikėtos kliūtys, nepavyzdingas eismo dalyvių elgesys ar dėl eismo valdymo signalų. Dėl to prognozės pasižymėjo dideliais nuokrypiais, susidarė didelės paklaidos.

Sumažinus pradinio duomenų svyravimus, duomenų kreivės dispersija turėtų sumažėti ir dėl to padidėti prognozės modelių tikslumas. Glodinimo įgyvendinimas prognozuojant intensyvumą realiuoju laiku galėtų būti atliktas taikant duomenų filtrą, sudaryta remiantis tirtais glodinimo metodais. Taikant šį filtrą realiuoju laiku, kiekviena naujai užfiksuota reikšmė taptų paskutiniu glodinimo pločio nariu ir į laiko eilutę būtų įrašyta perskaičiuota jos vertė.

Šio skyriaus tikslas – patikrinti glodinimo metodų taikymą, siekiant suvienodinti laiko eilutės svyravimus ir sumažinti atsitiktinių nuokrypių poveikį prognozės tikslumui. Bandymui naudojami pradiniai duomenys, aptarti 3.1 poskyryje. Atliekamas visos laiko eilutės glodinimas, t. y. ir duomenų, skirtų modelio mokymui, ir duomenų, skirtų modelio validavimui. Nustačius paklaidų sumažėjimą, glodinimo metodų taikymas įvertinamas ir prognozuojant trumpesnio laikotarpio – 360 s, intensyvumą.

Glodinimas atliekamas taikant Gauso filtro bei slankiosios medianos metodus, vertinamas 3 ir 5 reikšmių glodinimo plotis. Glodinimo metodų aprašas pateiktas 2.3 poskyryje.

3.5.1 Prognozės rezultatai glodintiems duomenims

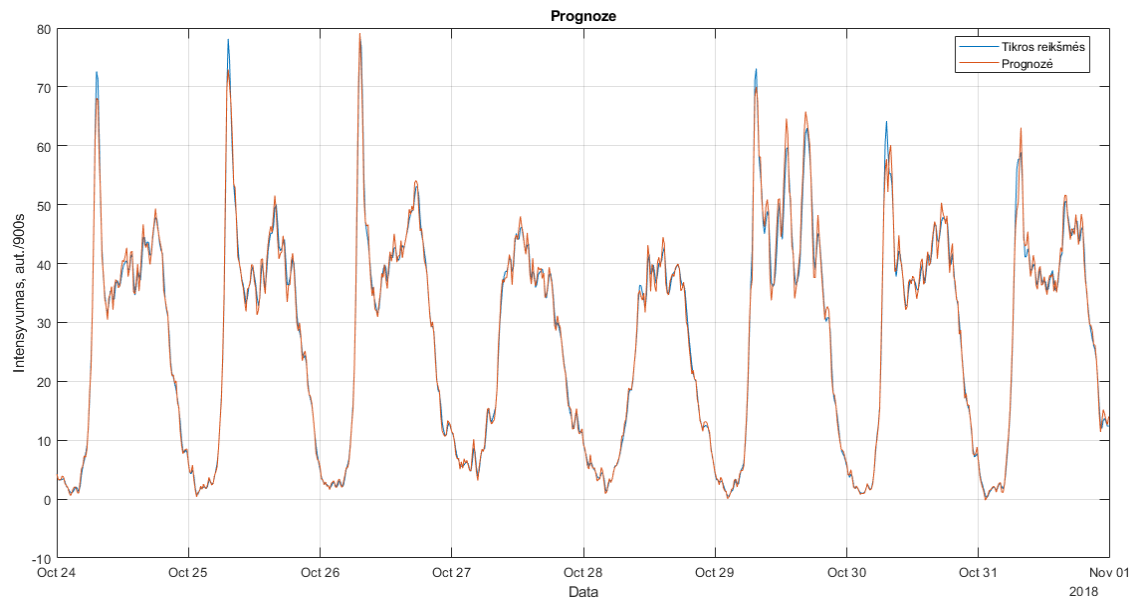
3.15 lentelė. Glodintų duomenų prognozės paklaidos

Glodinimo metodas	Glodinimo plotis, reikšmės	Modelis	MAPE, %	MAE	RMSE
Gauso filtras	3	1.SARIMA(1,0,1)(0,1,1)	14,58	2,5	3,65
		2.SARIMA(0,0,5)(0,1,1)	14,76	2,5	3,64
		3.SARIMA(2,1,2)(0,1,1)	15,6	2,5	3,64
		4.SARIMA(1,1,2)(0,1,1)	15,65	2,49	3,65
		LSTM	16,84	2,43	3,55
	5	1.SARIMA(1,0,1)(0,1,1)	6,56	0,96	1,38
		2.SARIMA(0,0,5)(0,1,1)	6,06	0,91	1,25
		3.SARIMA(2,1,2)(0,1,1)	6,27	0,91	1,26
		4.SARIMA(1,1,2)(0,1,1)	6,6	0,95	1,33
		LSTM	5,58	0,84	1,18
Slankioji mediana	3	1.SARIMA(1,0,1)(0,1,1)	11,92	2,27	3,48
		2.SARIMA(0,0,5)(0,1,1)	11,58	2,25	3,44
		3.SARIMA(2,1,2)(0,1,1)	11,98	2,26	3,47
		4.SARIMA(1,1,2)(0,1,1)	11,84	2,27	3,48
		LSTM	14,17	2,28	3,47
	5	1.SARIMA(1,0,1)(0,1,1)	9,5	1,83	2,97
		2.SARIMA(0,0,5)(0,1,1)	9,5	1,82	2,96
		3.SARIMA(2,1,2)(0,1,1)	10	1,85	2,98
		4.SARIMA(1,1,2)(0,1,1)	10,07	1,85	2,99
		LSTM	10,37	1,77	2,85

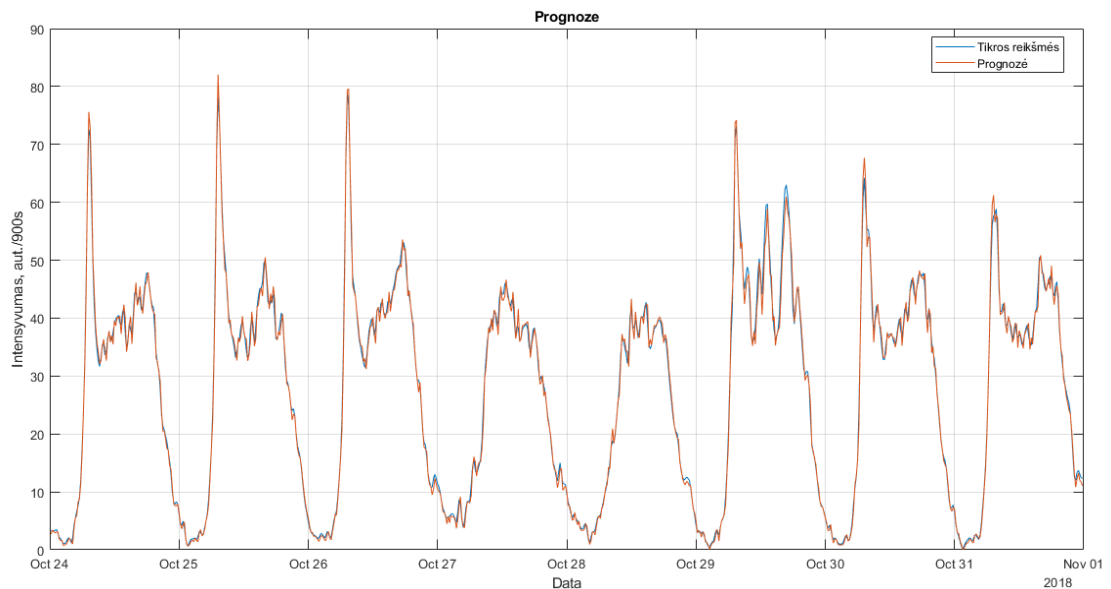
Taikant Gauso filtrą ir 3 reikšmių glodinimo plotį prognozės tikslumas padidėjo (3.15 lentelė). SARIMA modelių MAPE paklaidos sumažėjo iki 14–15%, LSTM modelio paklaida sumažėjo iki 17%. Padidinus glodinimo plotį iki 5 reikšmių, modelių paklaidos pasiekė 10% ribą. SARIMA(1,0,1)(0,1,1) modelio paklaida siekė 6,6%, o LSTM modelio tikslumas buvo didžiausias, paklaida siekė 5,6%.

Atlikus laiko eilutės glodinimą slankiosios medianos metodu pastebima, prognozės paklaidos pasižymi mažesne sklaida nei glodinimui taikant Gauso filtrą ir mažiau priklauso nuo glodinamo pločio. Esant 3 reikšmių pločiui, SARIMA modelių MAPE paklaidos siekė 11%, LSTM 14%. Padidinus plotį iki 5 reikšmių, modelių paklaidos šiek tiek sumažėjo, SARIMA modelių atveju siekė 9–10%, LSTM modelio – 10%.

Proгнозиų grafikuose pastebimi sumažėję realių duomenų svyravimai. SARIMA(1,0,1)(0,1,1) ir LSTM modeliai gerai atkartojo realių duomenų kreives. SARIMA modelio atveju pastebimi nedideli nuokrypiai, ne visiškai įvertinti realių duomenų pikai (3.31 pav.). LSTM modelio prognozės kreivė tolygiai atkartoja glodintų duomenų kreivę (3.32 pav.). Lyginant su neglodintų duomenų prognoze (3.15 pav.), LSTM modelis šiuo atveju atkartojo ir nedidelius svyravimus, nevertino tik vidutinių reikšmių.



3.31 pav. Gauso filtru, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant SARIMA(1,0,1)(0,1,1) modelį



3.32 pav. Gauso filtru, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant LSTM modelį

Kitos glodintų duomenų prognozės pateiktos 9 priede.

3.5.2 Glodinimo metodų taikymas, prognozuojant trumpesnio laikotarpio intensyvumą

Sudarant trumpesnio laikotarpio prognozes buvo performuota pradinių duomenų laiko eilutė, sumažinti laiko intervalai tarp matavimų (3.3 poskyris). Sumažinus laiko intervalus buvo pastebėtas aukšto dažnio triukšmas duomenyse, dėl kurio prognozės modeliams buvo sunku prisitaikyti prie duomenų svyravimų ir padidėjo prognozių paklaidos. 3.5.1 skyrelyje nustatčius reikšmingą prognozės paklaidų sumažėjimą taikant glodinimo metodus pradiniais duomenimis, tikrinamas šių metodų efektyvumas sudarant trumpesnio laikotarpio prognozes.

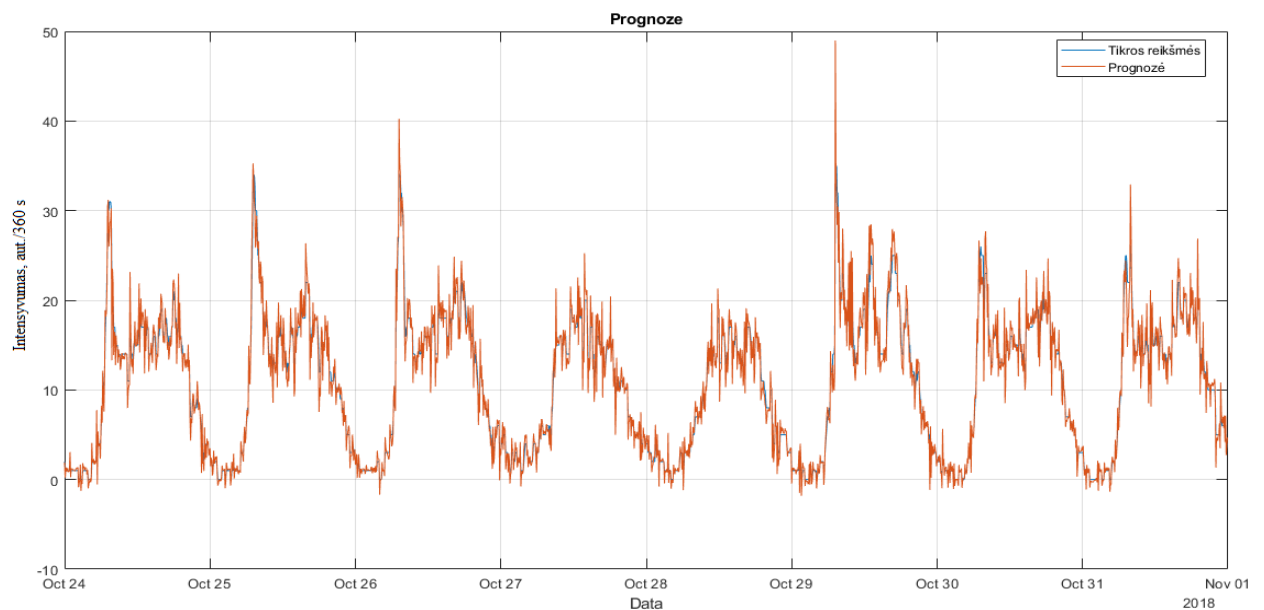
Vertinant nustatytas prognozės paklaidas (3.16 lentelė), prognozuojant 360 s laiko žingsniu stabilesnės paklaidos pastebimos tų duomenų, kurie buvo glodinami slankiosios medianos metodu. Esant 3 reikšmių glodinimo pločiui, SARIMA modelių paklaidos siekė 14–15%, LSTM – 17%. Padidinus glodinimo plotį iki 5 reikšmių, tiksliausiai veikė SARIMA(1,0,1)(0,1,1) modelis, kurio MAPE paklaida siekė 10,8%. Kitų modelių paklaidos buvo panašios.

3.16 lentelė. Prognozės paklaidos prognozuojant 360 s laikotarpį

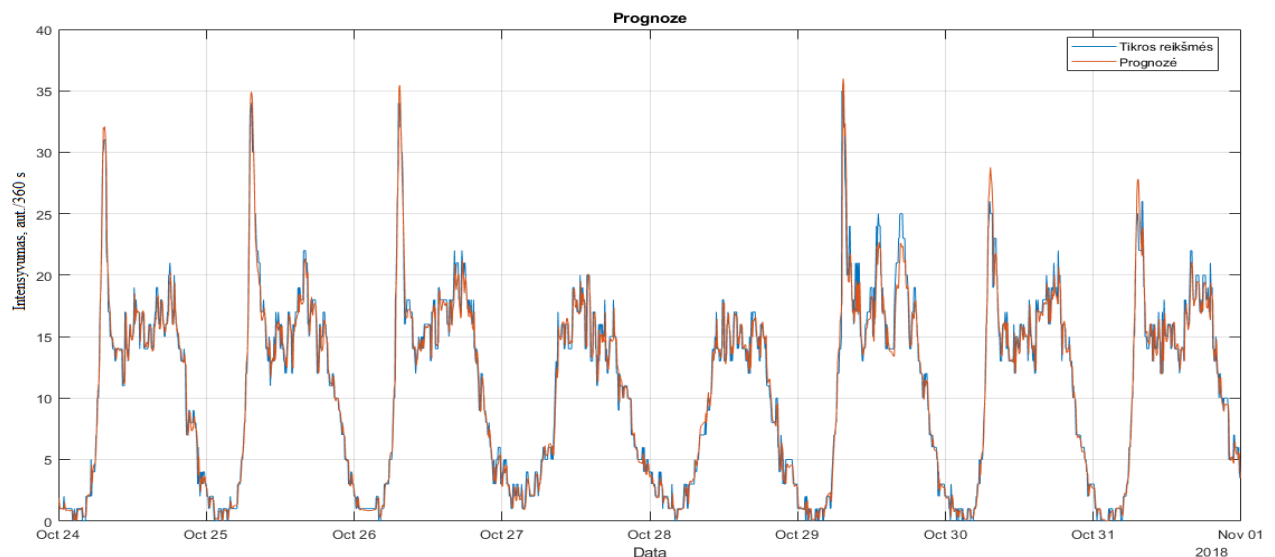
Glodinimo metodas	Glodinamas plotis, reikšmės	Modelis	MAPE	MAE	RMSE
Gauso filtras	3	1.SARIMA(1,0,1)(0,1,1)	28,76	1,52	2,12
		2.SARIMA(0,0,5)(0,1,1)	28,96	1,51	2,11
		3.SARIMA(2,1,2)(0,1,1)	29,1	1,51	2,11
		4.SARIMA(1,1,2)(0,1,1)	29,11	1,51	2,12
		LSTM	28,76	1,47	2,04
	5	1.SARIMA(1,0,1)(0,1,1)	18,99	0,63	0,79
		2.SARIMA(0,0,5)(0,1,1)	18,32	0,58	0,72
		3.SARIMA(2,1,2)(0,1,1)	18,34	0,6	0,744
		4.SARIMA(1,1,2)(0,1,1)	15,8	0,62	0,78
		LSTM	9,1	0,5	0,69
Slankioji mediana	3	1.SARIMA(1,0,1)(0,1,1)	14,8	1,24	1,88
		2.SARIMA(0,0,5)(0,1,1)	14,89	1,23	1,86
		3.SARIMA(2,1,2)(0,1,1)	14,88	1,25	1,88
		4.SARIMA(1,1,2)(0,1,1)	15,19	1,25	1,89
		LSTM	17,24	1,25	1,84
	5	1.SARIMA(1,0,1)(0,1,1)	10,77	0,88	1,38
		2.SARIMA(0,0,5)(0,1,1)	11,02	0,87	1,37
		3.SARIMA(2,1,2)(0,1,1)	11,15	0,88	1,39
		4.SARIMA(1,1,2)(0,1,1)	11,26	0,88	1,39
		LSTM	11,63	0,84	1,32

Glodinimui naudojant Gauso filtrą, prognozių paklaidos pasižymėjo didele sklaida. Esant 3 reikšmių glodinimo pločiui, SARIMA modelių paklaidos buvo didesnės nei pradiniu atveju (žr. 3.11 lentelė), siekė 28–29%. LSTM modelio atveju, pastebimas 1% pagerėjimas. Padidinus glodinimo plotį iki 5 reikšmių, SARIMA modelių paklaidos siekė 15–18%. Tiksliausiai tarp SARIMA modelių veikė SARIMA(1,1,2)(0,1,1) modelis. LSTM paklaida šiuo atveju buvo mažiausia, siekė 9,1%.

Vizualiai lyginant SARIMA(1,0,1)(0,1,1) bei LSTM modelių prognozes pastebimi išlikę nedideli svyravimai realiuose duomenyse. Sumažinus intervalą tarp matavimų iki 360 s, ne visi skirtumai tarp gretimų reikšmių buvo suvienodinti kaip 3.5.1 skyrelio atveju. SARIMA(1,0,1)(0,1,1) modelis atkartojė svyravimus, kai kuriais atvejais pastebimas realių duomenų pervertinimas, didesni nuokrypiai vidurdienio metu (3.33 pav.). LSTM modelis pasižymėjo tolygesnia kreive ir mažesniais svyravimais, labiau laikėsi vidutinių reikšmių (3.34 pav.). Lyginant abu grafikus pastebima, SARIMA modelis labiau linkęs pervertinti realių duomenų reikšmes, o LSTM modelis laikėsi vidutinių reikšmių ir vengė atsitiktinių svyravimų.



3.33 pav. Slankiaja mediana, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant SARIMA(1,0,1)(0,1,1)



3.34 pav. Slankiaja mediana, 5 reikšmių pločiu glodintos laiko eilutės prognozė, taikant LSTM modelį

3.5.3 Dalinės išvados

1. Glodinimo metodų taikymas padėjo reikšmingai sumažinti prognozės modelių paklaidas. Vertinant pradinių duomenų laiko eilutę, mažiausios paklaidos buvo pasiektos glodinimui taikant Gauso filtrą ir 5 reikšmių glodinimo plotį. SARIMA(1,0,1)(0,1,1) modelio MAPE paklaida siekė 6,6%, LSTM – 5,6%. Slankiosios medianos metodo paklaidos mažiau priklausė nuo glodinimo pločio. Mažiausios šio metodo paklaidos nustatytos taikant 5 reikšmių glodinimo plotį. SARIMA(1,0,1)(0,1,1) modelio MAPE paklaida siekė 9,5%, LSTM – 10,4%.
2. Išbandžius glodinimo metodų efektyvumą prognozuojant 360 s laikotarpį, modelių tikslumas taip pat pagerėjo. Tačiau pastebėtas paklaidų padidėjimas glodinimui taikant Gauso filtrą ir 3 reikšmių glodinimo plotį. Šiuo metodu glodintus duomenis geriausiai prognozavo LSTM modelis, kurio MAPE paklaida siekė 9,1%, SARIMA(1,0,1)(0,1,1) modelio atveju – 19%. Slankiosios medianos metodu glodintų duomenų prognozių paklaidos buvo stabilesnės. Esant 3 reikšmių glodinimo pločiui, SARIMA modelių MAPE paklaidos siekė 14–17%. Esant 5 reikšmių glodinimo pločiui, tiksliausiai prognozavo SARIMA(1,0,1)(0,1,1) modelis, kurio MAPE siekė 10,8%. LSTM modelio atveju – 11,6%.
3. Vertinant išbandytus glodinimo metodus, mažiausios paklaidos buvo nustatytos glodinimui taikant Gauso filtrą. Tačiau šiuo metodu glodintų duomenų prognozės paklaidos buvo nestabilios, pasižymėjo didele sklaida. Slankiosios medianos metodas pasižymėjo mažomis paklaidomis, kurios svyravo 9–17% intervale. Šio metodo elgesys labiau nuspėjamas, o skaičiavimo principas yra paprastesnis. Dėl šių priežasčių duomenų glodinimui rekomenduojama taikyti slankiosios medianos metodą.

IŠVADOS

Baigiamajame magistro darbe nagrinėti metodai, skirti trumpalaikiam eismo srautų prognozavimui. Teorinėje dalyje atlikta prognozės metodų klasifikacija, apžvelgti naudojami prognozės modeliai, aptartos jų savybės. Atlikus mokslo darbų analizę, buvo pastebėti parametrinių modelių privalumai prognozuojant eismo srautus dėl didesnės jų greیتaveikos, nesudėtingos modelių struktūros bei didelio tikslumo vertinant eismo srautus. Todėl tiriamojoje dalyje pasirinkta detaliau nagrinėti parametrinių ARIMA modelių grupę, o palyginimui naudotas neparametrinis LSTM modelis. Tyrimo metu buvo pasiekti šie rezultatai:

1. Sudarytas sezoninis SARIMA(1,0,1)(0,1,1) modelis, skirtas prognozuoti 15 min laikotarpio vienos eismo juostos intensyvumą. Šio modelio prognozės vidutinė absoliutinė procentinė paklaida (MAPE) siekia 17,6%. Kadangi nustatyta paklaida mažesnė nei 20%, šis modelis yra vertinamas kaip tikslus. Modelio greیتaveika vertinama kaip labai gera, vienos reikšmės prognozavimas truko 0,01 s, todėl modelį galima taikyti prognozuojant realiuoju laiku. Prognozavimą atliekant 15 min laiko žingsniu, galima papildomai atlikti regresijos koeficientų perskaičiavimą, kadangi modelio įvertinimo procesas trunka 3 min. Palyginimui naudoto LSTM modelio prognozės paklaida buvo didesnė, siekė 24,4%, o mokymo procesas truko ilgiau, 17,5 min. Šis modelis gali būti naudojamas prognozavimui realiuoju laiku, tačiau mokymo procesas turi būti vykdomas nepriklausomai. Vertinant šiuos rezultatus, SARIMA(1,0,1)(0,1,1) modelis buvo tinkamesnis nagrinėtos eismo juostos 15 min intensyvumui prognozuoti.
2. Buvo sudaryti skirtingos struktūros SARIMA modeliai, papildomai taikant pirmos eilės diferencijavimą bei skirtingų eilių (AR(p) ir MA(q)) dedamąsias. Esant perteklinei modelių struktūrai, prognozės rezultatai nepagerėjo, o greیتaveika modelių įvertinimo metu suprastėjo 1,5–2 kartus.
3. Atliekant prognozės modelio universalumo bandymą, buvo vertinama SARIMA(1,0,1)(0,1,1) modelio geba veikti su kitų eismo juostų duomenimis. Nustatytas žymus MAPE paklaidų padidėjimas, siekiantis 8–12%. Atlikus detalesnę analizę su gretimo jutiklio duomenimis, nustatyta modelio liekamųjų paklaidų koreliacija. Padidėjusios LSTM modelio paklaidos rodo, kad kitų jutiklių duomenys buvo mažiau stacionarūs, pasižymėjo didesne dispersija. Dėl šių priežasčių kiekvienos eismo juostos intensyvumas turi būti vertinamas nepriklausomai.
4. Taikant SARIMA bei LSTM modelius trumpesnio laikotarpio prognozėms, susidurta su dideliu triukšmu duomenyse. Sumažėjus laiko tarpui tarp duomenų matavimų, sumažėjo ir užfiksuotų transporto priemonių skaičius, dėl to skirtumai tarp gretimų matavimų buvo

labiau pastebimi. Prognozuojant 540 s laikotarpio žingsniu, modelių MAPE paklaidos padidėjo 1–3%, o 360 s atveju iki 6%. Vertinant grafinius rezultatus buvo pastebėtas didesnis duomenų nestacionarumas dėl dispersijos, prognozės modeliai negalėjo tiksliai įvertinti didelių duomenų svyravimų. Todėl esant nedideliame eismo intensyvumui reiktų prognozuoti didesniu laiko žingsniu, siekiant išvengti nestabilaus prognozės modelių veikimo.

5. Renkantis prognozuojamo laikotarpio trukmę (prognozavimo žingsnį), reiktų atsižvelgti į eismo reguliavimo prietaisų darbo ciklo optimizavimo periodiškumą. Per didelis prognozavimo žingsnis gali lemti per lėtą sistemos reakciją į eismo srautų dinamiką. Per mažas žingsnis gali sukelti nestabilumą eismo valdymo sistemoje, kadangi prognozės modeliai tokiomis sąlygomis netinkamai įvertina dažnus svyravimus duomenyse. Pasirinktas 15 min prognozavimo žingsnis yra tinkamas nagrinėjamos sankryžos eismo juostų intensyvumui prognozuoti, kadangi buvo sudarytos tikslios prognozės.
6. Prognozuojant trumpesnio laikotarpio žingsniu buvo performuota pradinė laiko eilutė, sumažinti laiko tarpai tarp duomenų matavimų. Dėl to padidėjo duomenų kiekis, kurį prognozės modeliai turėjo įvertinti, sulėtėjo modelių greیتaveika įvertinimo metu. Prognozuojant 540 s žingsniu, SARIMA(1,0,1)(0,1,1) modelio įvertinimas truko 526 s, LSTM modelio mokymas – 1807 s. Prognozuojant 360 s žingsniu, SARIMA(1,0,1)(0,1,1) modelio įvertinimas užtruko 1426 s, LSTM modelio atveju – 3195 s. Prognozės šiuo atveju gali būti sudaromos realiuoju laiku, tačiau pakartotinis modelių įvertinimas turi būti vykdomas nepriklausomai nuo prognozavimo.
7. Vertinant modeliui sudaryti reikalingos duomenų imties dydį nustatyta, kad esamos struktūros SARIMA(1,0,1)(0,1,1) modeliui pakanka 1 mėn. fiksuotų duomenų, su šia imtimi modelis veikė tiksliausiai. Esant minimaliai 2 sav. duomenų imčiai, prognozės paklaidos padidėjo apie 1,5%, bet neviršijo 20% MAPE. Tai rodo, kad ir esant minimaliai duomenų imčiai SARIMA modelis išlaiko pageidaujamą tikslumą. SARIMA modelio greیتaveika visais atvejais išliko panaši. LSTM modelio atveju, sumažinus mokymui skirtų duomenų imtį iki 2 sav., mokymosi procesas truko daug trumpiau – 316 s, tačiau paklaidos padidėjo iki 51%. Dėl to sudarant šio modelio pradinių duomenų imtį reikia atsižvelgti į tikslumą, kurio siekiama.
8. Glodinimo metodų taikymas sumažino prognozės modelių paklaidas. Atlikus glodinimą pradinių duomenų imčiai, mažiausios paklaidos buvo nustatytos taikant Gauso filtrą ir esant 5 reikšmių glodinimo pločiui. SARIMA(1,0,1)(0,1,1) modelio MAPE paklaida sumažėjo iki 6,6%, o LSTM iki 5,6%. Išbandžius glodinimo metodus su didesniu triukšmu pasižymėjusiais duomenimis, prognozuojant 360 s laiko žingsniu, Gauso filtro paklaidos

buvo nestabilios, pasižymėjo didele sklaida. Glodinimui taikant slankiosios medianos metodą ir 5 reikšmių glodinimo plotį buvo pasiektos stabilios 10–11% paklaidos. SARIMA(1,0,1)(0,1,1) modelio paklaida siekė 10,8%, LSTM modelio – 11,6%. Kadangi slankiosios medianos metodu glodintų duomenų prognozės buvo stabilesnės abiem atvejais ir šio metodo veikimas buvo labiau nuspėjamas, eismo srautų duomenims glodinti tinkamesnis yra šis metodas.

Vertinant pasiektus rezultatus, tyrimo metu buvo sudarytas užsibrėžto tikslumo SARIMA modelis bei įvertinti veiksniai, lemiantys prognozės tikslumą.

Rekomendacijos

1. Nedidelio intensyvumo transporto priemonių srautams prognozuoti patariama rinktis SARIMA modelius, kadangi šie modelis geba geriau įvertinti dideliais svyravimais pasižyminčią duomenų imtį. SARIMA modelių prognozės buvo tikslesnės, modeliai greičiau įvertina pradinis duomenis ir modeliams sudaryti reikalinga mažesnė duomenų imtis.
2. Esant nedideliems duomenų svyravimams rekomenduojama taikyti LSTM modelį arba atlikti papildomą duomenų glodinimą prieš taikant šį modelį.
3. Palyginti parametrinių bei nparametrinių modelių prognozavimo tikslumą esant dideliame sankryžos intensyvumui bei papildomai įvertinti skirtingo prognozuojamo laikotarpio prognozių tikslumą didelio intensyvumo sąlygomis.
4. Įvertinti šviesoforinio valdymo sankryžų darbo ciklo optimizavimo periodiškumą, siekiant užtikrinti stabilų eismo valdymo sistemos veikimą. Nustatytas periodiškumas taptų siekiamu prognozavimo žingsniu.
5. Įvertinti prognozuojamo kintamojo poveikį prognozės tikslumui. Nustatyti skirtumus tarp vidutinio eismo srauto greičio, kelionės trukmės ir intensyvumo prognozių.
6. Taikant prognozės modelius šviesoforų ciklui optimizuoti, periodiškai pakartoti duomenų įvertinimo procedūrą. Esant galimybei ir ilgam prognozavimo žingsniui, įvertinimo (mokymosi) procesą atlikti kiekvieną kartą prieš atliekant optimizavimą. Jei duomenų įvertinimas (modelio mokymo procesas) užtrunka ilgiau nei vieno prognozuojamo laikotarpio trukmė, mokymo procesą atlikti mažiausiu įmanomu periodiškumu, nepriklausomai nuo prognozavimo proceso.

Vygantas Verbilis



LITERATŪRA

1. Abbas, Z., *et al.* 2018. Short-Term Traffic Prediction Using Long Short-Term Memory Neural Networks, in *2018 IEEE International Congress on Big Data*, San Francisco, CA, USA: IEEE, 57–65.
2. Anokye, M., *et al.* 2013. Application of Queuing Theory to Vehicular Traffic at Signalized Intersection in Kumasi–Ashanti Region, Ghana, *American International Journal of Contemporary Research*, Vol. 3, No. 7; July 2013: 23–29.
3. Billings, D.; Yang, J. 2006. Application of the ARIMA Models to Urban Roadway Travel Time Prediction – A Case Study, in *2006 IEEE International Conference on Systems, Man, and Cybernetics*, Taipei, Taiwan, 2529–2534.
4. Bolishinsky, E.; Freidman, R. 2012. *Traffic Flow Forecast Survey*. Technion – Computer Science Department – Technical report [interaktyvus]. Haifa: Technion – Computer Science Department [žiūrėta 2020-05-03]. Prieiga per internetą:
<<http://www.cs.technion.ac.il/users/wwwb/cgi-bin/tr-get.cgi/2012/CS/CS-2012-06.pdf>>.
5. Brockwell, P.J.; Davis, A. R. 2002. *Introduction to Time Series and Forecasting*, Cham, Switzerland: Springer.
6. Dong, C., *et al.* 2014. Flow rate and time mean speed predictions for the urban freeway network using state space models, *Transportation Research Part C* (2014): 20–32.
7. Dong H., *et al.* 2009. Road traffic flow prediction with a time – oriented ARIMA model, in *2009 Fifth International Joint Conference on INC, IMS and IDC*, Seoul, Korea, 1649–1652.
8. Hinsbergen, C.P. IJ.; Lint, J.W.C.; Sanders, F.N. 2007. Short Term Traffic Prediction Models, in *Advance Traffic Monitoring* [interaktyvus]. Beijing, China. [žiūrėta 2020-05-03]. Prieiga per internetą:
<https://www.researchgate.net/publication/228882378_Short_Term_Traffic_Prediction_Models>
9. Hyndman, J. R.; Kostenko, A. V. 2007. Minimum sample size requirements for seasonal forecasting models, *Foresight Issue* 6: 12–15.
10. Yan, L., *et al.* 2013. Multi-Objective Optimization of Traffic Signal Timing for Oversaturated Intersection, *Mathematical Problems in Engineering*, Volume 2013: 1–9.
11. Younes, M. B.; Boukerche, A. 2016. Intelligent Traffic Light Controlling Algorithms Using Vehicular Networks, *IEEE Transactions on vehicular technology*, vol. 65, No. 8, August 2016: 5887–5899.

12. Yuan, X.; Li, L.; Wang, Y. 2018. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network, *IEEE Transactions on Industrial Informatics*, PP(99):1–1: 1–9.
13. Yue, Y., *et al.* 2007. Prediction Time Horizont and Effectiveness of Real-time Data on Short-term Traffic Predictability, in *Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference*, Seattle, WA, USA: 962–967.
14. Kerner, B. S. 2018. *Traffic Congestion, Spatiotemporal Features of*. New York: Springer.
15. Kumar, S. V.; Vanajakshi, L. 2015. Short-term traffic flow prediction using seasonal ARIMA modeli with limited input data, *Eur. Transp. Res. Rev.* (2015): 7–21.
16. Lin, Y.; Want, P.; Ma, M. 2017. Intelligent Transportation System (ITS): Concept, Challenge and Opportunity, in *2017 IEEE 3rd International Conference on Big Data Security on Cloud*, Beijing, China: IEEE, 167–172.
17. Mohamed, E. A., *et al.* 2015. Short-Term Travel Time Estimation: A Case Study, in *Proceeding of the IEEE 28th Canadian Conference on Electrical and Computer Engineering*, Halifax, Canada, 489–496.
18. Nagy, A. M.; Simon, V. 2018. Survey on traffic prediction in smart cities, *Pervasive and Mobile Computing* 50 (2018): 148–163.
19. Nellore, K.; Hancke, G.P. 2016. A survey on Urban traffic management system using wireless sensor networks, *Sensors* 2016, 16, 157: 1–25.
20. Qiao, S., *et al.* 2017. Short-Term Traffic Flow Forecast Based on Parallel Long Short-Term Memory Neural Network, in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China: IEEE, 253–257.
21. Shrestha, B. P., *et al.* 2010. Review of urban traffic management and the impacts of new vehicle technologies, *IET Intelligent Transport Systems*, Vol. 3, Issue 4: 419–428.
22. Smith, B. L.; Williams, B. M.; Oswald, K. R. 2002. Comparison of parametric and nonparametric models for traffic flow forecasting, *Transportation research part C emerging technologies*: 303–321.
23. Sumayya, P. A.; Shefeena, P. S. 2015. VANET Based Vehicle Tracking Module for Safe and Efficient Road Transportation System, in *International Conference on Information and Communication Technologies (ICICT 2014)*, Depok, Indonesia, 1173–1180.
24. Sun, S.; Zhang, C.; Yu, G. 2006. A Bayesian Network Approach to Traffic Flow Forecasting, *IEEE Transactions on intelligent transportation systems*, Vol. 7, No. 1: 124–132.

25. Zaidi, K., *et al.* 2016. Host-Based Intrusion Detection for VANETs: A Statistical Approach to Rogue Node Detection, *IEEE Transactions on vehicular technology*, Vol. 65, No. 8, August 2016: 6703–6714.
26. Zhang, F.; Grande, R. E. D.; Boukerche, A. 2016. Accuracy Analysis of Short-term Traffic Flow Prediction Models for Vehicular Clouds, in *PE-WASUN '16*, Malta, Malta, 19–26.
27. Zheng, Z., *et al.* 2016. LSTM network: a deep learning approach for short-term traffic forecast, *IET Intelligent Transport systems*, 2017, Vol. 11 Iss. 2: 68–75.
28. Zheng, X.; Recker, W. 2013. An adaptive control algorithm for traffic-actuated signals, *Transportation Research Part C* 30: 93–115.
29. Williams, B. M.; Hoel, L. A. 2003. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results, *Journal of Transportation Engineering*, Vol. 129, No. 6: 664–672.
30. Xu, D., *et al.* 2017. Real-time road traffic state prediction based on ARIMA and Kalman filter, *Front Inform Technol Electron Eng* 2017 18(2): 287–302.
31. Xu, Y.; Kong, Q.; Liu, Y. 2013. Short-Term Traffic Volume Prediction Using Classification and Regression Trees, in *2013 IEEE Intelligent Vehicles Symposium (IV)*, Gold Coast, Australia, 493–498.

PRIEDAI

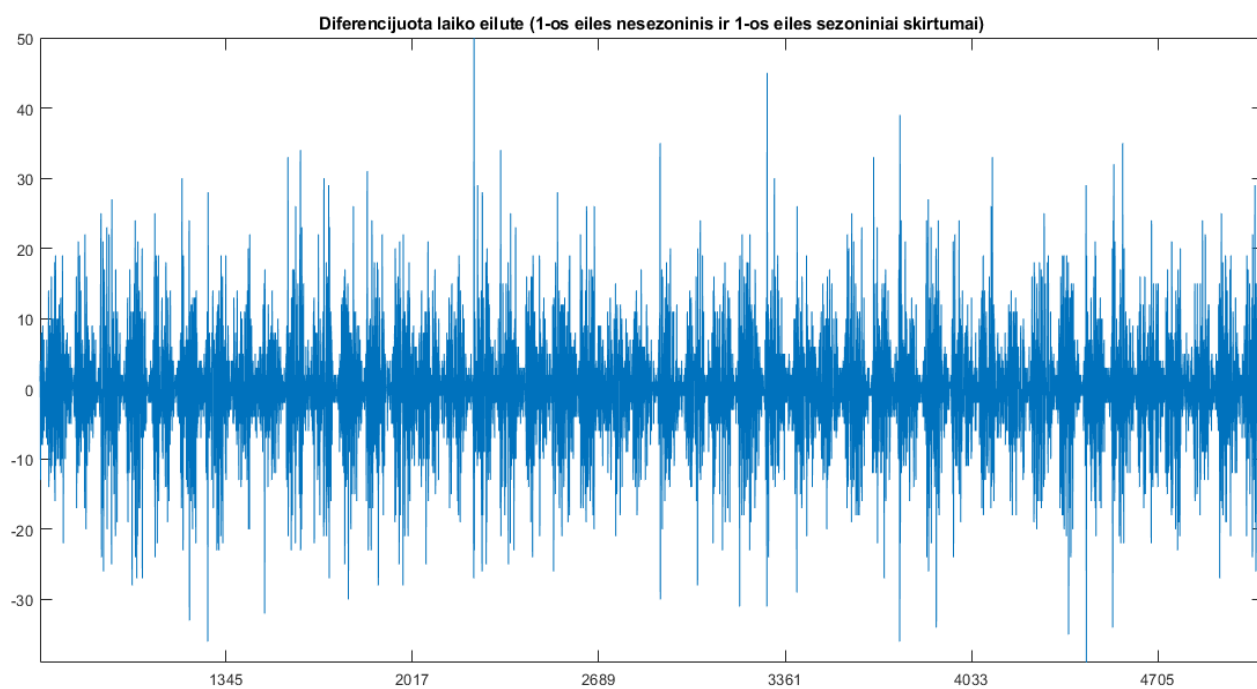
1 Priedas

Literatūros apžvalgos santrauka

Autoriai	Vieta	Prognozės modelis	Duomenų imtis modeliui sudaryti (laikotarpis)	Prognozuojamas laikotarpis, mėn	Prognozuojamas kintamasis
Abbas <i>et al.</i> (2018)	Greitkelis	NN, ARIMA, SVR, RNN	1 mėn.	10, 20, 30	Eismo tankis
Billings, Yang (2006)	Greitis	ARIMA	3 mėn.	40	Kelionės trukmė
Dong <i>et al.</i> (2014)	Greitkelis	ARIMA, Elman	1 mėn.	2	Intensyvumas
Dong <i>et al.</i> (2009)	Greitis	ARIMA	2 mėn.	5, 10	Intensyvumas, vidutinis greitis
Kumar, Vanajakshi (2015)	Miestas	SARIMA	3 d.	10	Intensyvumas
Qiao <i>et al.</i> (2017)	Nenurodyta	ARIMA, LSTM, SVR, Kalmano filtras	Nenurodyta	3	Intensyvumas
Smith <i>et al.</i> (2002)	Greitkelis	KNN, SARIMA, HA	2 mėn.	15	Intensyvumas
Zhang <i>et al.</i> (2016)	Greitkelis	SARIMA, SVR, DES, BP, KNN	1 mėn.	Nenurodyta	Intensyvumas
Zheng <i>et al.</i> (2016)	Miestas	ARIMA, LSTM, RNN, SVR	6 mėn.	15, 30, 45, 60	Intensyvumas
Williams, Hoel (2003)	Greitkelis	SARIMA, HA, RW	3, 5 mėn.	15	Intensyvumas
Xu <i>et al.</i> (2017)	Miestas	Hibridinis modelis (ARIMA sujungtas su Kalmano filtru), ARIMA, Kalmano filtras	4 d.	20	Vidutinis greitis, intensyvumas
Xu <i>et al.</i> (2013)	Greitkelis	CART, KNN, Kalmano filtras	1 mėn.	15	Intensyvumas

2 Priedas

Pirma eile ir sezoniškai diferencijuota laiko eilutė



3 Priedas

SARIMA prognozės modelių įvertinimo procedūra

Eil. Nr.	SARIMA(p,d,q) (P,D,Q) modelių eilės	AIC	BIC	Liekamųjų paklaidų autokoreliacija	Nustatyti modelio koeficientai reikšmingi (t statistika >2)	Komentaras
1	(1,0,0)(0,1,1)	31613	31633	Taip (2 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
2	(2,0,0)(0,1,1)	31571	31597	Taip (3 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
3	(3,0,0)(0,1,1)	31565	31597	Taip (4 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
4	(2,0,1)(0,1,1)	31562	31594	Ne	Ne (AR(2))	Galimas. Perteklinė AR(p) modelio struktūra
5	(3,0,1)(0,1,1)	31563	31602	Ne	Ne (AR(2); AR(3))	Galimas. Perteklinė AR(p) modelio struktūra
6	(0,0,1)(0,1,1)	31625	31645	Taip (2 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
7	(0,0,2)(0,1,1)	31582	31608	Taip (3 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
8	(0,0,3)(0,1,1)	31571	31603	Taip (4 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
9	(1,0,3)(0,1,1)	31563	31602	Ne	Ne (MA(2);MA(3))	Galimas. Perteklinė MA(q) dalies struktūra
10	(0,0,4)(0,1,1)	31569	31607	Taip (5 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
11	(0,0,5)(0,1,1)	31566	31611	Ne	Taip	Tinkamas
12	(1,0,1)(0,1,1)	31562	31587	Ne	Taip	Tinkamas
13	(2,0,2)(0,1,1)	31564	31602	Ne	Ne (nenustatytas nei vienas reikšmingas koeficientas)	Galimas. Tikėtina perteklinė modelio struktūra
14	(0,1,1)(0,1,1)	31694	31713	Taip (3 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
15	(0,1,2)(0,1,1)	31651	31677	Taip (2 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos

3 Priedo lentelės tęsinys

16	$(0,1,3)(0,1,1)$	31618	31650	Taip (3 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
17	$(0,1,4)(0,1,1)$	31610	31648	Ne	Taip	Tinkamas
18	$(0,1,5)(0,1,1)$	31609	31654	Ne	Taip	Tinkamas
19	$(1,1,5)(0,1,1)$	31568	31619	Ne	Ne (MA(2:5))	Galimas. Perteklinė MA(q) dalies struktūra
20	$(1,1,1)(0,1,1)$	31642	31668	Taip (2 vėlavimas)	Taip	Netinkamas dėl liekamųjų paklaidų koreliacijos
21	$(2,1,2)(0,1,1)$	31569	31607	Ne	Taip	Tinkamas
22	$(1,1,2)(0,1,1)$	31567	31599	Ne	Taip	Tinkamas

4 Priedas

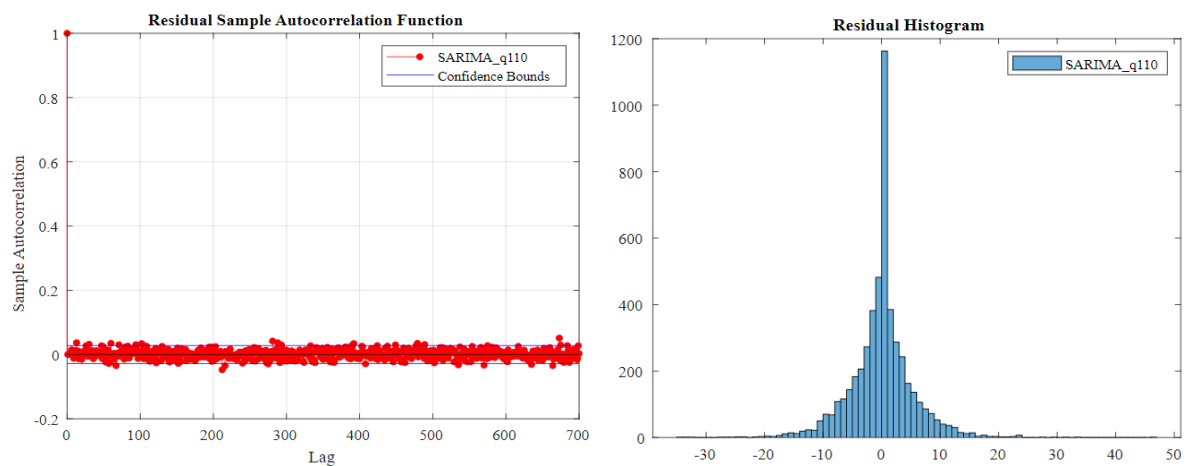
SARIMA lyginamųjų modelių liekamosios paklaidos

1. SARIMA(0,0,5)(0,1,1)

$$(1 - L^{672})y_t = (1 + \theta_1 L - \dots - \theta_5 L^5)(1 + \theta_{672} L^{672})\varepsilon_t$$

4 priedo 1 lentelė. SARIMA(0,0,5)(0,1,1) regresijos koeficientai

Parametras	Vertė	t Statistikos reikšmė
Konstanta	0	
MA{1}	0.10537	12.0073
MA{2}	0.096825	8.8765
MA{3}	0.054289	5.2638
MA{4}	0.03327	3.3552
MA{5}	0.029899	2.9925
SMA{672}	-0.59734	-80.7022
Variance	28.8879	87.2474



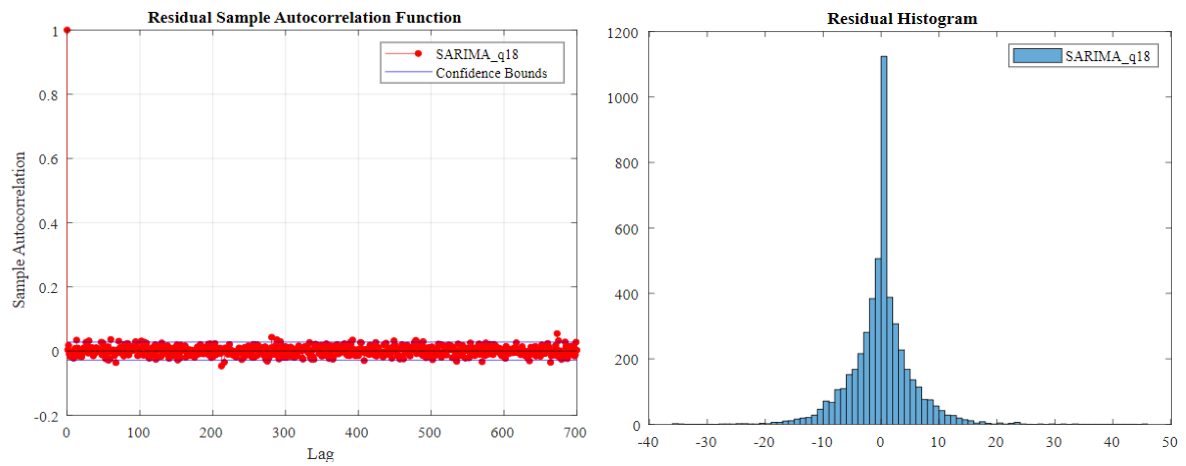
4 priedo 1 pav. SARIMA(0,0,5)(0,1,1) liekamųjų paklaidų autokoreliacijos funkcija ir histograma

2. SARIMA(2,1,2)(0,1,1)

$$(1 - \varphi_1 L - \varphi_1 L^2)(1 - L)(1 - L^{672})y_t = (1 + \theta_1 L + \theta_2 L^2)(1 + \theta_{672} L^{672})\varepsilon_t$$

4 priedo 2 lentelė. SARIMA(2,1,2)(0,1,1) regresijos koeficientai

Parametras	Vertė	t Statistikos reikšmė
Konstanta	0	
AR{1}	0.80374	25.8578
AR{2}	-0.0024174	-0.19099
MA{1}	-1.7022	-62.167
MA{2}	0.70218	25.8076
SMA{672}	-0.59432	-80.8852
Variance	28.9148	88.9436



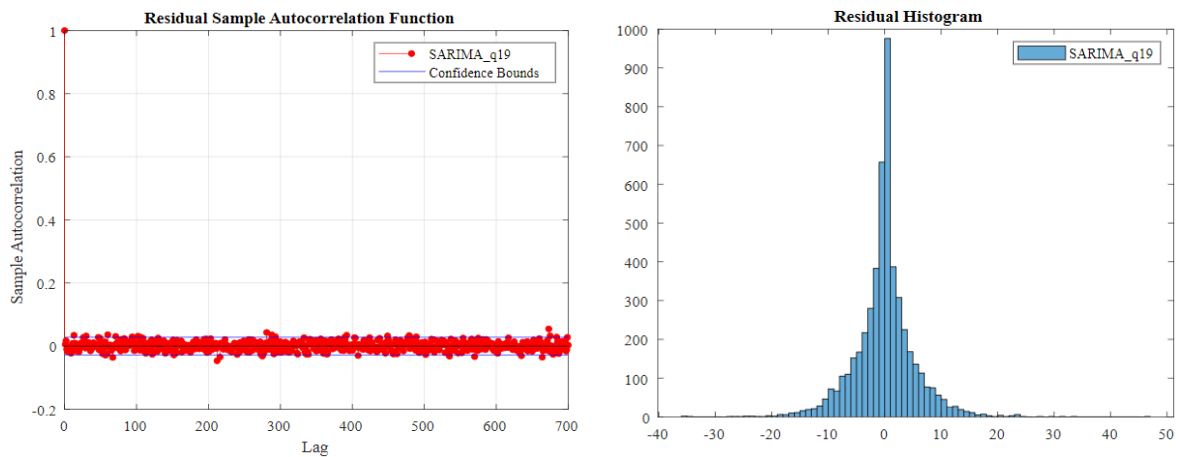
4 priedo 2 pav. SARIMA(2,1,2)(0,1,1) liekamųjų paklaidų autokoreliacijos funkcija ir histograma

3. SARIMA(1,1,2)(0,1,1)

$$(1 - \varphi_1 L)(1 - L)(1 - L^{672})y_t = (1 + \theta_1 L + \theta_2 L^2)(1 + \theta_{672} L^{672})\varepsilon_t$$

4 priedo 3 lentelė. SARIMA(1,1,2)(0,1,1) regresijos koeficientai

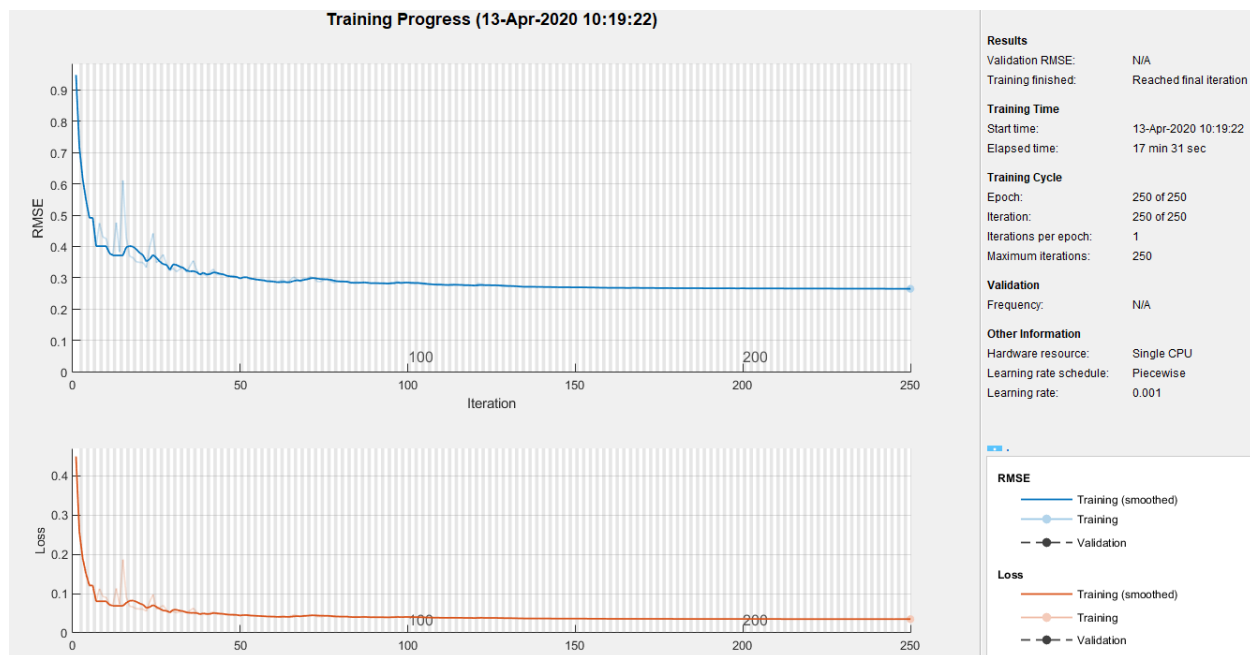
Parametras	Vertė	t Statistikos reikšmė
Konstanta	0	
AR{1}	0.79938	78.9433
MA{1}	-1.699	-122.4437
MA{2}	0.69904	51.0141
SMA{672}	-0.59425	-80.8763
Variance	28.9141	88.9142



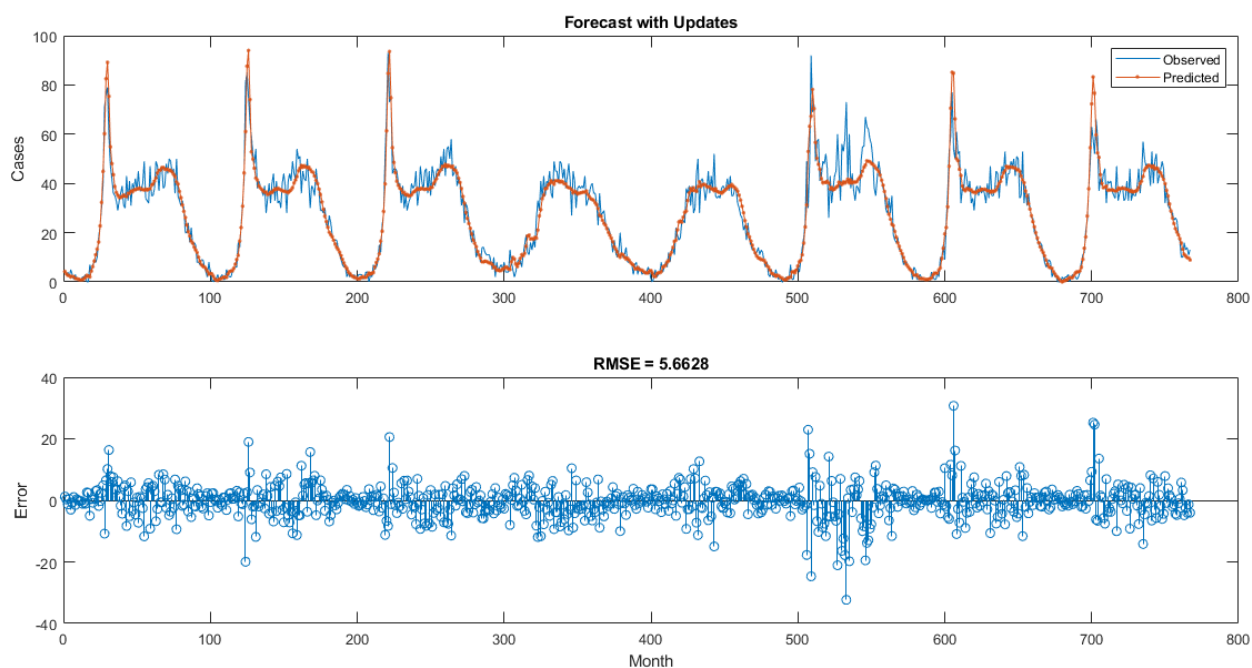
4 priedo 3 pav. SARIMA(1,1,2)(0,1,1) liekamųjų paklaidų autokoreliacijos funkcija ir histograma

5 Priedas

LSTM modelio mokymo kreivė



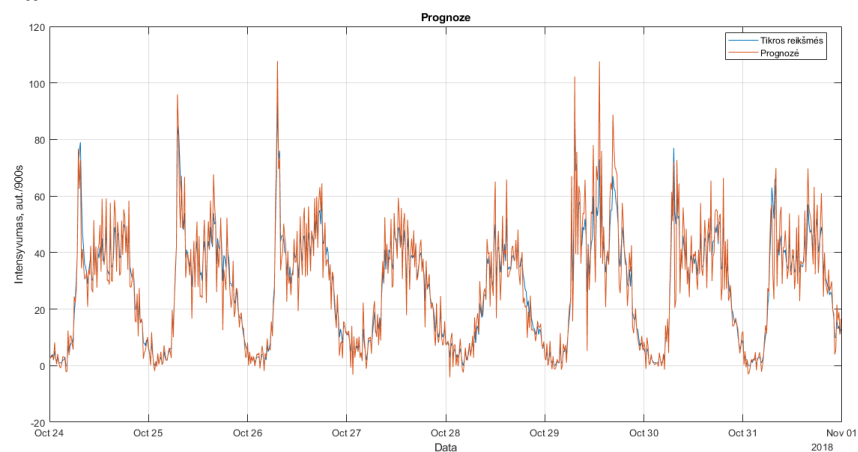
5 priedo 1 pav. LSTM modelio paklaidų optimizavimas mokymosi proceso metu



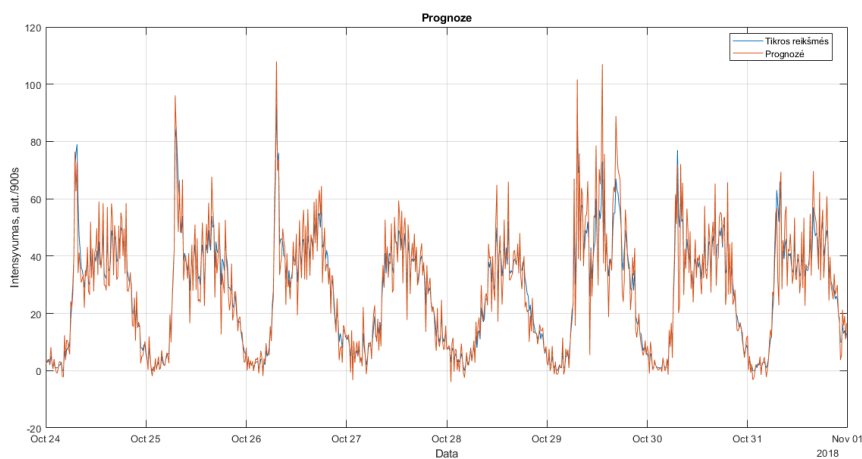
4 priedo 2 pav. LSTM modelio prognozė ir liekamosios paklaidos

6 priedas

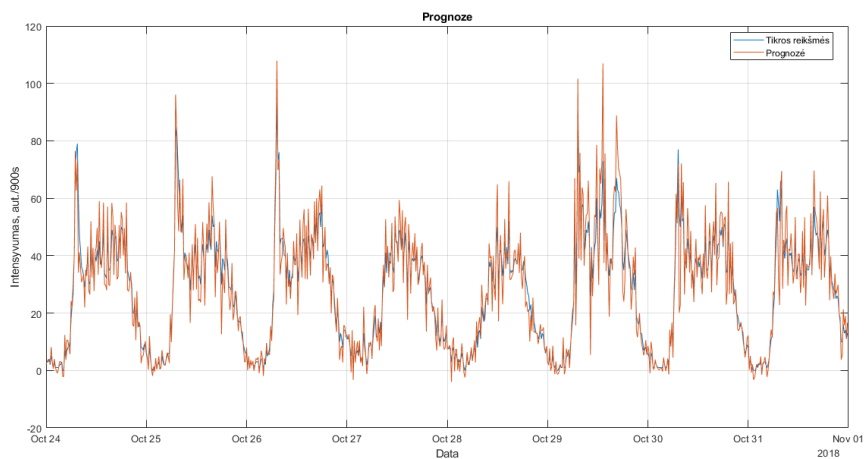
Prognozių grafikai



6 priedo 1 pav. SARIMA(0,0,5)(0,1,1) modelio prognozė



6 priedo 2 pav. SARIMA(2,1,2)(0,1,1) modelio prognozė



6 priedo 3 pav. SARIMA(1,1,2)(0,1,1) modelio prognozė

7 priedas

Skaitinės prognozių vertės 1 paros reikšmės (2018-10-24)

Data	Realus intensyvumas	1) SARIMA(0,0,5)(0,1,1)672	2) SARIMA(1,0,1)(0,1,1)672	3) SARIMA(2,1,2)(0,1,1)672	4) SARIMA(1,1,2)(0,1,1)672	LSTM
24/10/2018 00:00	3	2	2	2	2	4
24/10/2018 00:15	3	3	3	3	3	3
24/10/2018 00:30	3	4	4	4	4	3
24/10/2018 00:45	4	3	3	3	3	2
24/10/2018 01:00	2	2	2	2	2	2
24/10/2018 01:15	5	8	8	8	8	2
24/10/2018 01:30	3	3	3	3	3	2
24/10/2018 01:45	1	0	1	1	1	1
24/10/2018 02:00	3	4	4	4	4	1
24/10/2018 02:15	1	1	1	1	1	1
24/10/2018 02:30	1	1	1	1	1	1
24/10/2018 02:45	1	1	1	1	1	1
24/10/2018 03:00	1	1	1	1	1	1
24/10/2018 03:15	2	3	3	3	3	1
24/10/2018 03:30	2	3	3	3	3	2
24/10/2018 03:45	3	2	2	2	2	3
24/10/2018 04:00	0	1	1	1	1	2
24/10/2018 04:15	0	1	1	1	1	2
24/10/2018 04:30	7	12	12	12	12	4
24/10/2018 04:45	4	4	4	4	4	6
24/10/2018 05:00	7	11	11	11	11	8
24/10/2018 05:15	8	10	10	10	10	11
24/10/2018 05:30	8	8	8	8	8	13
24/10/2018 05:45	10	6	6	6	6	16
24/10/2018 06:00	18	24	24	24	24	23
24/10/2018 06:15	23	22	22	22	22	32
24/10/2018 06:30	30	27	27	27	27	45
24/10/2018 06:45	40	40	40	40	40	60
24/10/2018 07:00	71	77	77	76	76	83
24/10/2018 07:15	76	63	63	63	63	89
24/10/2018 07:30	79	73	73	73	73	75
24/10/2018 07:45	59	34	34	34	34	55
24/10/2018 08:00	47	41	41	41	41	48
24/10/2018 08:15	42	36	36	36	36	44
24/10/2018 08:30	36	31	31	31	31	39
24/10/2018 08:45	35	31	32	32	32	36

Lentelės tęsinys (I)

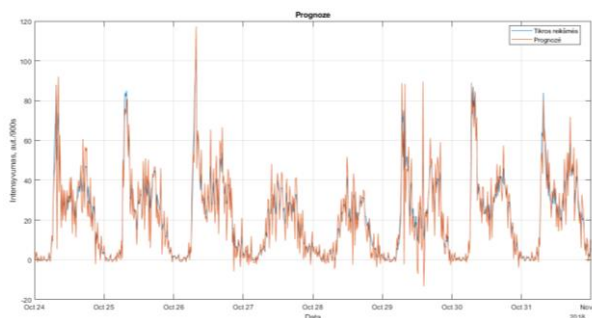
24/10/2018 09:00	33	33	33	33	33	35
24/10/2018 09:15	29	21	21	22	22	34
24/10/2018 09:30	33	34	35	35	35	34
24/10/2018 09:45	36	32	33	33	33	35
24/10/2018 10:00	39	42	43	43	43	35
24/10/2018 10:15	31	27	27	28	28	35
24/10/2018 10:30	30	26	26	26	26	35
24/10/2018 10:45	43	52	52	52	52	36
24/10/2018 11:00	33	32	32	32	32	36
24/10/2018 11:15	41	42	43	43	43	37
24/10/2018 11:30	31	28	28	28	28	37
24/10/2018 11:45	40	41	41	41	41	37
24/10/2018 12:00	38	43	43	43	43	37
24/10/2018 12:15	42	50	50	50	50	38
24/10/2018 12:30	37	33	33	33	33	38
24/10/2018 12:45	45	59	59	59	59	38
24/10/2018 13:00	37	41	40	40	40	38
24/10/2018 13:15	37	36	36	36	36	37
24/10/2018 13:30	40	41	41	41	41	37
24/10/2018 13:45	49	59	59	59	59	38
24/10/2018 14:00	34	30	30	30	30	37
24/10/2018 14:15	33	30	30	30	30	38
24/10/2018 14:30	32	29	28	28	28	38
24/10/2018 14:45	47	58	57	57	57	39
24/10/2018 15:00	35	30	30	30	30	40
24/10/2018 15:15	35	30	30	30	30	41
24/10/2018 15:30	38	42	42	42	42	42
24/10/2018 15:45	49	59	58	58	58	44
24/10/2018 16:00	47	54	54	54	54	45
24/10/2018 16:15	37	33	33	33	33	45
24/10/2018 16:30	46	51	51	51	51	46
24/10/2018 16:45	47	46	46	46	46	46
24/10/2018 17:00	38	32	32	32	32	46
24/10/2018 17:15	39	33	33	33	33	46
24/10/2018 17:30	48	51	51	51	51	46
24/10/2018 17:45	44	39	39	39	39	45
24/10/2018 18:00	50	55	55	55	55	45
24/10/2018 18:15	49	53	53	53	53	45
24/10/2018 18:30	44	40	40	40	40	44

Lentelės tęsinys (II)

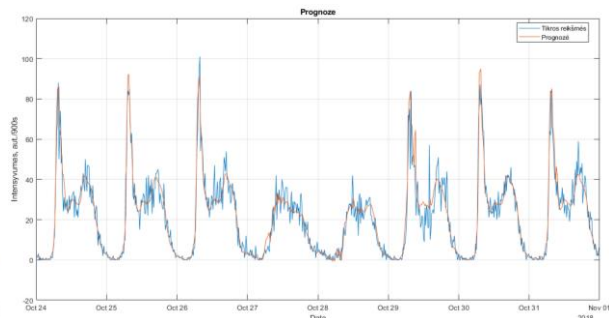
24/10/2018 18:45	46	49	50	50	50	43
24/10/2018 19:00	36	34	34	34	34	41
24/10/2018 19:15	50	58	59	58	58	40
24/10/2018 19:30	35	28	28	28	28	37
24/10/2018 19:45	33	28	28	28	28	34
24/10/2018 20:00	31	30	30	30	30	31
24/10/2018 20:15	33	35	34	34	34	29
24/10/2018 20:30	30	28	28	28	28	26
24/10/2018 20:45	20	16	16	16	16	23
24/10/2018 21:00	20	15	15	15	15	20
24/10/2018 21:15	24	25	26	25	25	19
24/10/2018 21:30	16	10	10	11	11	17
24/10/2018 21:45	22	28	28	28	28	16
24/10/2018 22:00	15	15	15	15	15	14
24/10/2018 22:15	16	17	17	17	17	13
24/10/2018 22:30	15	15	16	16	16	12
24/10/2018 22:45	8	2	2	3	3	10
24/10/2018 23:00	8	5	5	5	5	8
24/10/2018 23:15	7	5	5	5	5	7
24/10/2018 23:30	9	8	8	8	8	6
24/10/2018 23:45	9	10	10	10	10	6

8 priedas

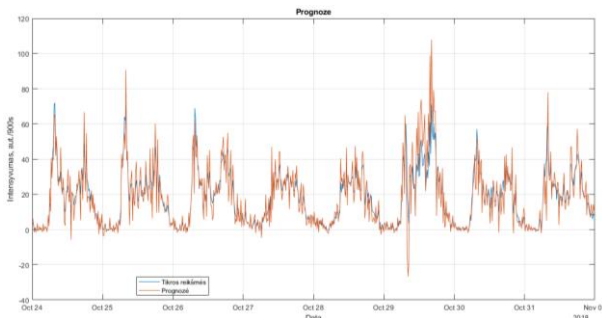
Kitų eismo juostų prognozės



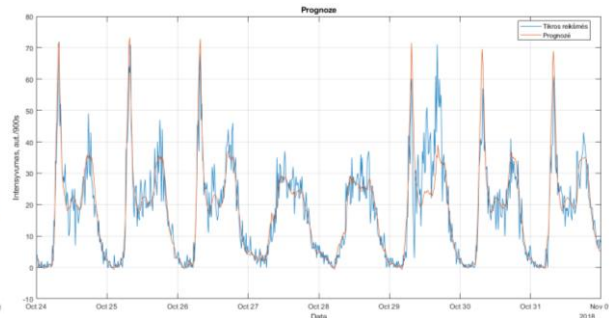
8 priedo 1 pav. VD2.2, SARIMA(1,0,1)(0,1,1)



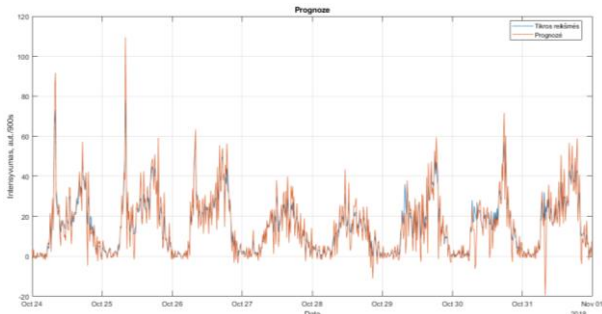
8 priedo 2 pav. VD2.2, LSTM



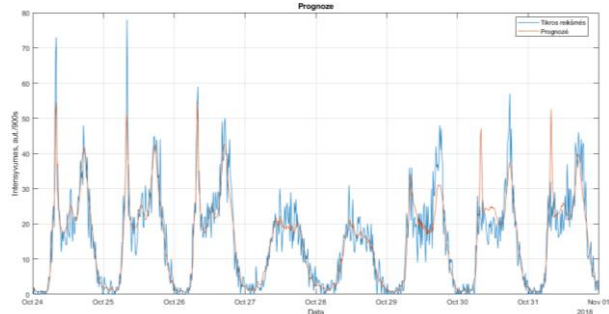
8 priedo 3 pav. VD2.3, SARIMA(1,0,1)(0,1,1)



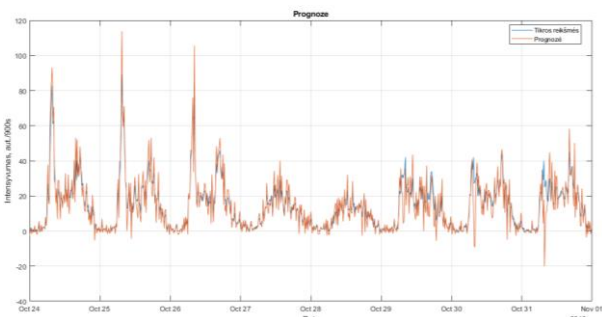
8 priedo 4 pav. VD2.3, LSTM



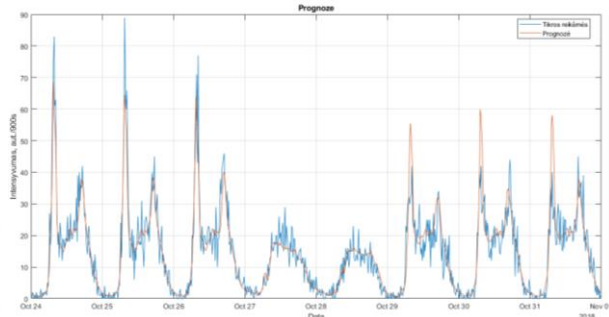
8 priedo 5 pav. VD4.1, SARIMA(1,0,1)(0,1,1)



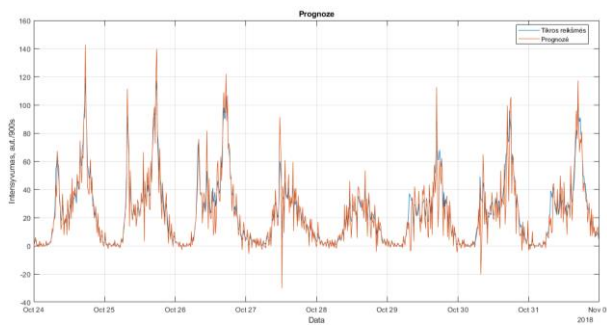
8 priedo 6 pav. VD4.1, LSTM



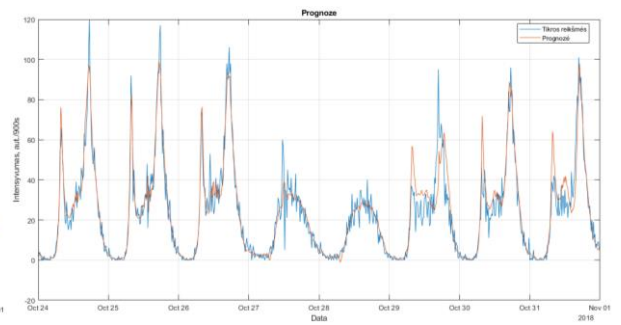
8 priedo 7 pav. VD4.2, SARIMA(1,0,1)(0,1,1)



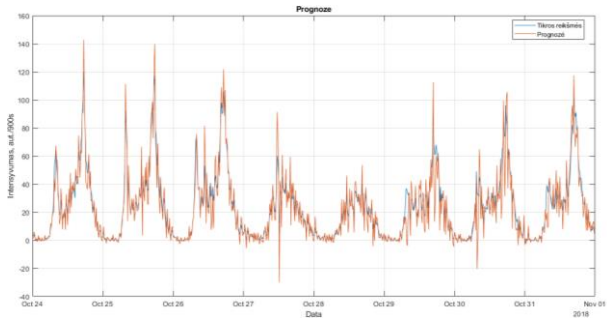
8 priedo 8 pav. VD4.2, LSTM



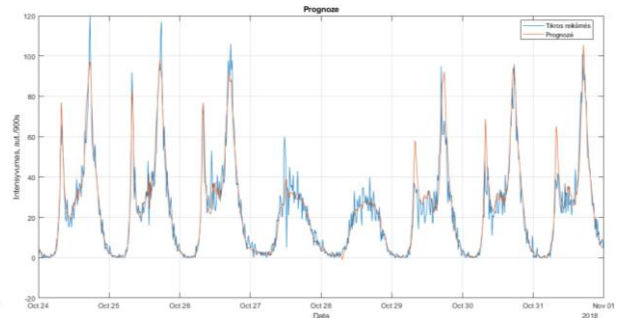
8 priedo 9 pav. VD6.1, SARIMA(1,0,1)(0,1,1)



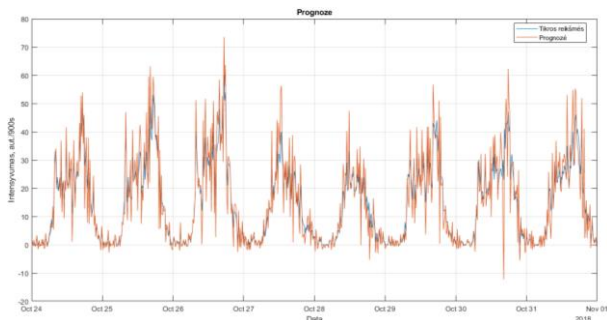
8 priedo 10 pav. VD6.1, LSTM



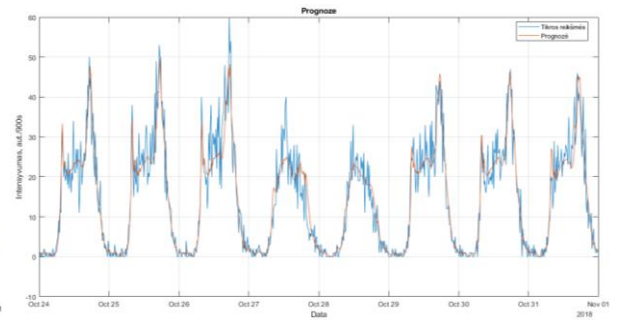
8 priedo 11 pav. VD6.2, SARIMA(1,0,1)(0,1,1)



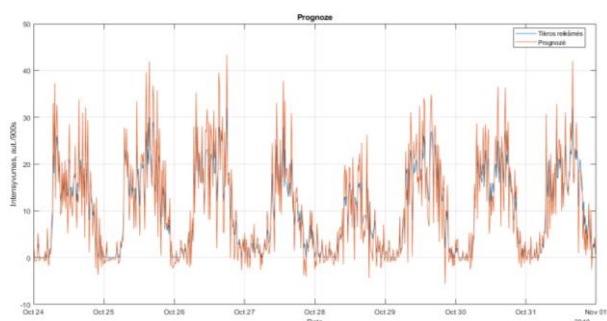
8 priedo 12 pav. VD6.2, LSTM



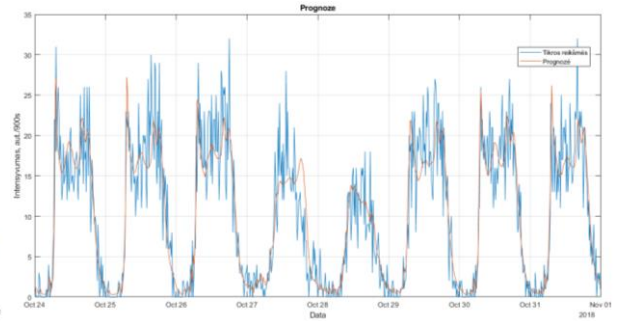
8 priedo 13 pav. VD6.3, SARIMA(1,0,1)(0,1,1)



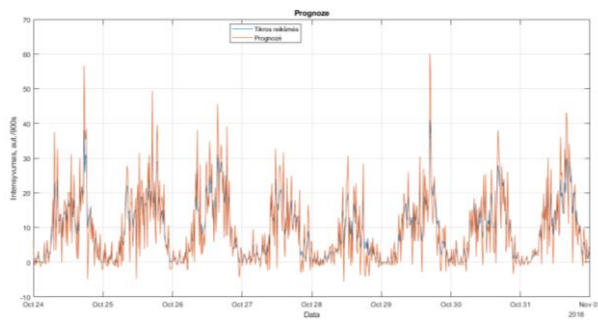
8 priedo 14 pav. VD6.3, LSTM



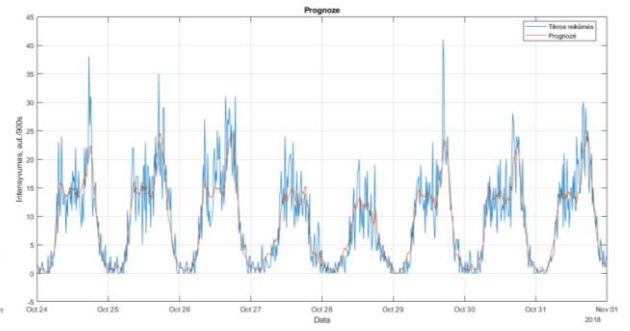
8 priedo 15 pav. VD8.1, SARIMA(1,0,1)(0,1,1)



8 priedo 16 pav. VD8.1, LSTM



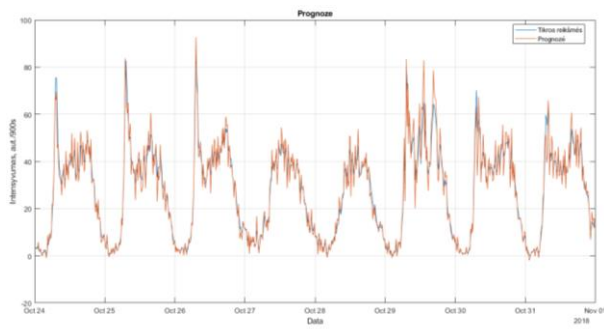
8 priedo 17 pav. VD8.2, SARIMA(1,0,1)(0,1,1)



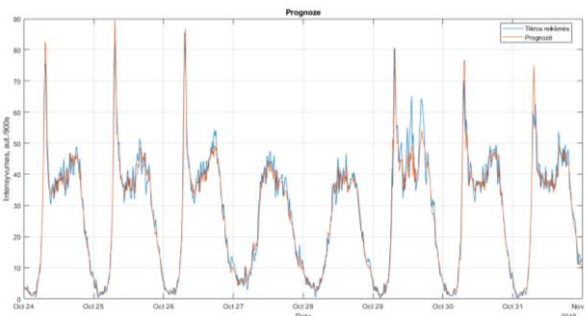
8 priedo 18 pav. VD8.2, LSTM

9 Priedas

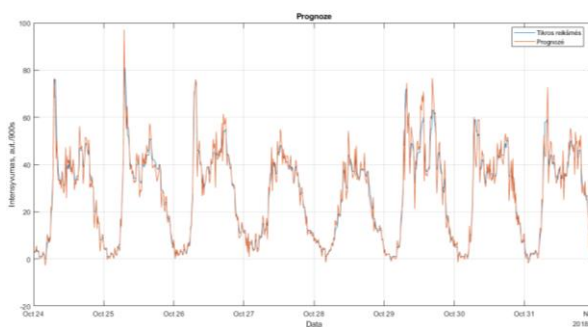
Glodintų duomenų prognozės



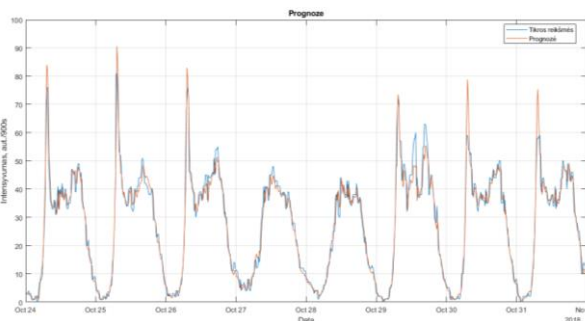
9 priedo 1 pav. Gauso filtras, 3 reikšmių plotis
SARIMA(1,0,1)(0,1,1)



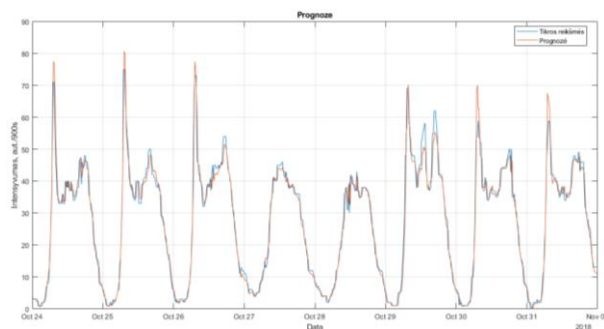
9 priedo 2 pav. Gauso filtras, 3 reikšmių plotis
LSTM



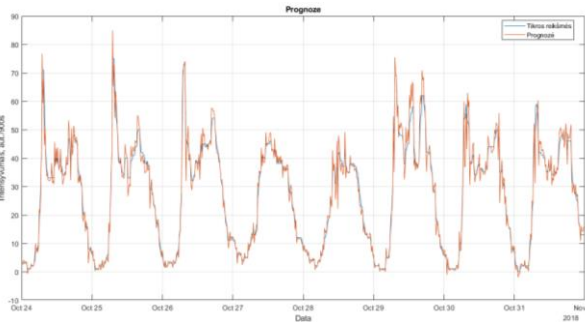
9 priedo 3 pav. Slenkanti mediana, 3 reikšmių plotis
SARIMA(1,0,1)(0,1,1)



9 priedo 4 pav. Slenkanti mediana, 3 reikšmių plotis
LSTM



9 priedo 5 pav. Slenkanti mediana, 5 reikšmių plotis
SARIMA(1,0,1)(0,1,1)



9 priedo 6 pav. Slenkanti mediana, 5 reikšmių plotis
LSTM