

РЕЦЕНЗИЯ

Дипломдық жұмыс
(жұмыс түрінің атауы)

Альжанова Динара Серікқызы
(білім алушының Т.А.Ә.)

6B06103 - Математикалық және компьютерлік модельдеу
(мамандық атауы мен шифрі)

Тақырыбы: Компания өнімдеріне сұранысты болжауда машиналық оқыту әдістерін қолдану

Орындалды:

- а) графикалық бөлім 19 парақ
- б) түсініктеме 23 бет

ЖҰМЫСҚА ЕСКЕРТУ

Дипломдық жұмыстың тақырыбы машиналық оқыту әдістерін тауарларға сұранысты болжауда қолдану сұрақтарын қамтиды және заманауи өзекті мәселе болып саналады. Жалпы жұмыс бойынша ешқандай ескертулер жоқ.

ЖҰМЫСТЫҢ БАҒАСЫ

Дипломдық жұмысты өнімдердің сатылу деңгейін арттыру мақсатында компания ресурстарын оңтайлы бөлу, болжау мәселесіндегі актуальді жұмыс ретінде санауға болады. Альжанова Динара жұмыс тақырыбы бойынша зерттеулер жүргізіп, өзгермелі заңдылықтарға бейімделетін және күрделі қатынастарды бекітетін деректерге негізделген тәсілді ұсынған. Алынған нәтижелерді ұқсас деректер жиынтығын пайдалана отырып сұранысты болжауда ұсынуға болатыны көрсетілген.

Дипломдық жұмыс 6B06103 – “Математикалық және компьютерлік модельдеу” мамандығына қойылатын талапқа және ережелерге сай. Жұмыс авторы толығымен 98% өте жоғары бағалануға лайықты.

Рецензент

Абай атындағы ҚазҰПУ

«Информатика және білімді ақпараттандыру»
кафедрасының қауымдастырылған профессоры,
педагогика ғылымдарының кандидаты



РАСТАЙМЫН: «Абай атындағы ҚазҰПУ» КЕАҚ
ЖЕТЕКШІСІ
Г.А. Асқарова РУКОВОДИТЕЛЬ
КАДРОВОЙ ПОЛИТИКИ
НАО «ҚазНПУ им. Абая»
КОЛЫ
ПОСЫЛСЫ
2024ж.

ҒЫЛЫМИ ЖЕТЕКШІНІҢ

ПІКІРІ

Дипломдық жұмыс

(жұмыс түрінің атауы)

Альжанова Динара Серікқызы

(білім алушының Т.А.Ә.)

6B06103 - Математикалық және компьютерлік модельдеу

(мамандық атауы мен шифрі)

Тақырыбы: Компания өнімдеріне сұранысты болжау кезінде машиналық оқыту әдістерін қолдану

Дипломдық жұмыста кәсіпорын өніміне сұранысты болжау теориясына, уақыт қатарына талдау жүргізілген, болжау әдістері, уақыт қатарларының модельдері қарастырылған.

Компания өнімдеріне сұранысты болжау кезінде машиналық оқыту әдістері қолданылған. Дипломдық жұмыстың программалық нәтижесі ARIMA, SARIMA модельдерін арқылы алынған. Жұмыста қойылған мақсат, міндеттер толық көлемде орындалған.

Альжанова Динара өзінің еңбекқорлығымен, зерттеу және талдау қабілеттілігімен, тапсырылған жұмысты уақытымен орындауымен ерекшеленеді. Тақырып бойынша анықталған тапсырламалар жоғары дәрежеде орындалған, 98 % бағалануға лайықты.

Ғылыми жетекші:

аға оқытушы

 Ергазина Р.А.

«31» ^(колы) 05 2024 ж.

Университеттің жүйе администраторы мен Академиялық мәселелер департаменті
директорының ұқсастық есебіне талдау хаттамасы

Жүйе администраторы мен Академиялық мәселелер департаментінің директоры көрсетілген еңбекке қатысты дайындалған Плагиаттың алдын алу және анықтау жүйесінің толық ұқсастық есебімен танысқанын мәлімдейді:

Автор: Альжанова Динара Серікқызы

Тақырыбы: Компания өнімдеріне сұранысты болжау кезінде машиналық оқыту әдістерін қолдану

Жетекшісі: Рыскуль Ергазина

1-ұқсастық коэффициенті (30): 2.2

2-ұқсастық коэффициенті (5): 0

Дәйексөз (35): 0.5

Әріптерді ауыстыру: 5

Аралықтар: 9

Шағын кеңістіктер: 1

Ақ белгілер: 0

Ұқсастық есебін талдай отырып, Жүйе администраторы мен Академиялық мәселелер департаментінің директоры келесі шешімдерді мәлімдейді :

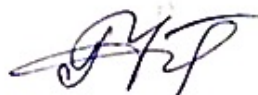
Ғылыми еңбекте табылған ұқсастықтар плагиат болып есептелмейді. Осыған байланысты жұмыс өз бетінше жазылған болып санала отырып, қорғауға жіберіледі.

Осы жұмыстағы ұқсастықтар плагиат болып есептелмейді, бірақ олардың шамадан тыс көптігі еңбектің құндылығына және автордың ғылыми жұмысты өзі жазғанына қатысты күмән тудырады. Осыған байланысты ұқсастықтарды шектеу мақсатында жұмыс қайта өңдеуге жіберілісін.

Еңбекте анықталған ұқсастықтар жосықсыз және плагиаттың белгілері болып саналады немесе мәтіндері қасақана бұрмаланып плагиат белгілері жасырылған. Осыған байланысты жұмыс қорғауға жіберілмейді.

Негіздеме:

Күні



Кафедра меңгерушісі

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ
МИНИСТРЛІГІ

«Қ. И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті»
коммерциялық емес акционерлік қоғамы

Автоматтандыру және ақпараттық технологиялар институты

Жоғары математика және модельдеу кафедрасы

Альжанова Динара Серікқызы

«Компания өнімдеріне сұранысты болжауда машиналық оқыту әдістерін
қолдану»

ДИПЛОМДЫҚ ЖҰМЫС

6B06103 – Математикалық және компьютерлік моделдеу

Алматы 2024

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

«К. И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті» коммерциялық
емес акционерлік қоғамы

Автоматтандыру және ақпараттық технологиялар институты

Жоғары математика және модельдеу кафедрасы

ҚОРҒАУҒА ЖІБЕРІЛДІ
«Жоғарғы Математика және
Модельдеу» кафедрасының
меңгерушісі физика-математика
ғылымдарының
кандидаты, қауымдастырылған
профессор Тулешева.Г.А.
(қолы)
«30» 05 2024 ж

ДИПЛОМДЫҚ ЖҰМЫС

Тақырыбы : «Компания өнімдеріне сұранысты болжауда машиналық оқыту әдістерін қолдану»

6B06103 – Математикалық және компьютерлік моделдеу

Орындаған

Альжанова Динара Серікқызы

Рецензент

Ғылыми жетекшісі

Физика-математика ғылымдарының

Ерғазина Р.А.,аға оқытушы

кандидаты, Абай атындағы Қазақ ұлттық

педагогикалық университеті «Информатика

және білім беруді ақпараттандыру бөлімі»

кафедрасының қауымдастырылған профессоры



Асқарова Г.А. РУКОВОДИТЕЛЬ
КАДРОВОЙ ПОЛИТИКИ
ИМПЛЕМЕНТАЦИИ
НАО «КазНТУ им. Абай»
ҚОЛЫ
ПОДПИСЬ
2024

(қолы)
Ерғазина Р.А.
«30» 05 2024

Алматы 2024

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
« Қ. И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті » коммерциялық
емес акционерлік қоғамы

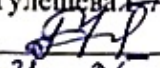
Автоматтандыру және ақпараттық технологиялар институты

Жоғары математика және модельдеу кафедрасы

6B06103 – Математикалық және компьютерлік моделдеу

БЕКІТЕМІН

«Жоғарғы Математика және
Модельдеу» кафедрасының
менгерушісі физика-математика
ғылымдарының кандидаты,
қауымдастырылған профессор
Тулешева Б.А.

 (қолы)
«31» 05 2024 ж.

**Дипломдық жұмысты орындауға арналған
ТАПСЫРМА**

Білім алушы Альжанова Динара Серікқызы

Тақырыбы: «Компания өнімдеріне сұранысты болжауда машиналық оқыту әдістерін қолдану»
Университет ректорының 2024 жылғы « 04 » желтоқсандағы №548-П/Ө бұйрығымен
бекітілген.

Аяқталған жұмысты тапсыру мерзімі: 2024 жылғы « 05 » 06 .

Дипломдық жобаның бастапқы деректері:

Дипломдық жұмыста әзірленуге жататын мәселелердің тізбесі немесе дипломдық жұмыстың
қысқаша мазмұны:

- а) Компания өнімдеріне сұранысты болжаудың теорисы
- б) Компания өнімдеріне сұранысты талдау
- в) Эксперименттер жүргізу және нәтижелерді талдау
- г) Уақыт қатарларының модельдері

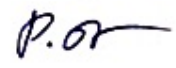




Ұсынылған негізгі әдебиеттер саны: 9

Дипломдық жобаны дайындау
КЕСТЕСІ

Бөлімдердің атаулары, зерттелген мәселелердің тізімі	Ғылыми жетекшіге және кеңесшілерге ұсыну мерзімі	Ескерту
Компания өнімдеріне сұранысты болжаудың теориясы	27.01.2024	орындалды
Компания өнімдеріне сұранысты талдау	25.02.2024	орындалды
Эксперименттер жүргізу және нәтижелерді талдау	21.03.2024	орындалды
Уақыт қатарларының модельдері	18.04.2024	орындалды

Қолтаңбалар


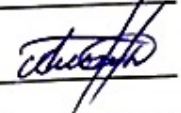
Аяқталған дипломдық жобаға консультанттар мен нормоконтролер оларға қатысты жұмыс бөлімдерін көрсете отырып

Бөлімдердің атаулары	Ғылыми жетекші, консультанттар, Т.А.Ә. (мұғ. дәрежесі, атағы)	Қол қойылған күні	Қолы
Компания өнімдеріне сұранысты болжаудың теориясы	Ерғазина Р.А.,аға оқытушы	30.05.24	
Кәсіпорын өнімдеріне сұранысты талдау	Ерғазина Р.А.,аға оқытушы	30.05.24	
Эксперименттер жүргізу және нәтижелерді талдау	Ерғазина Р.А.,аға оқытушы	30.05.24	
Уақыт қатарларының модельдері	Ерғазина Р.А.,аға оқытушы	30.05.24	
Норма бақылаушы	Шатманов Ж. Ж., физ-мат. ғылымдарының кандидаты, қауымдастырылған профессор	31.05.2024	

Ғылыми жетекшісі

Білім алушы тапсырманы орындауға алды

Күні



«30» 05 2024ж

Ерғазина Р.А.

Альжанова Д.С

МАЗМҰНЫ

	Кіріспе	7
1	Теориялық бөлім	8
1.1	Компания өнімдеріне сұранысты болжау	8
1.2	Уақыт қатарларын талдау	10
1.3	Статистикалық болжау әдістері	13
1.4	Болжау әдістері	15
2	Компания өнімдеріне сұранысты талдау	16
2.1	Эксперименттер жүргізу және нәтижелерді талдау	16
3	Жалпы анықтамалар	24
3.1	Уақыт қатарларының модельдері	24
3.1.1	ARIMA моделі	24
3.1.2	SARIMA (маусымдық ARIMA) моделі	25
3.1.3	ARIMA vs SARIMA ерекшеліктері	26
3.1.4	ARIMA мен SARIMA мәні	27
4	Нәтижелер	28
	Қорытынды	29
	Пайдаланылған әдебиеттер тізімі	31
	Қосымша А	
	Қосымша В	
	Қосымша С	

АНДАТПА

Бұл дипломдық жұмыс тауарларға сұранысты болжау мәселесін шешуде машиналық оқыту әдістерін қолдануға арналған. Өртүрлі модельдерді зерттеу және болжау сапасын жақсарту үшін жасалған пайдаланылған деректер жиынтығының талдауы берілген. Дипломдық жұмыста уақыттық қатарларды болжау үшін ең өзекті тәсілдер қолданылады. Модельдік болжамдар мен нақты мәндердің графиктері, сондай-ақ модельдердің құрылымы ұсынылған, соның арқасында осы жұмыста болжамның минималды қателігіне қол жеткізу мүмкін болды.

Жұмыстың негізгі бөлімі сұранысты болжау кезінде машиналық оқыту әдістерін пайдаланудан тұрды, модельдердің түрлеріне сәйкес сұраныс анықтамалары мен графиктері көрсетілген.

Зерттеудің практикалық маңыздылығы мынада: осы жұмыстың нәтижелері мен материалдары ұқсас деректер жиынтығын пайдалана отырып сұранысты болжауда пайдаланыға болады.

АННОТАЦИЯ

Данная диплом работа посвящена применению методов машинного обучения при решении задачи прогнозирования спроса на товары. Представлен анализ использованного набора данных для изучения различных моделей и повышения качества прогнозов. В дипломной работе используются наиболее актуальные методы прогнозирования временных рядов. Представлены графики модельных прогнозов и фактических значений, а также структура моделей, благодаря которой удалось добиться минимальной погрешности прогноза в данной работе.

Основная часть работы заключалась в использовании методов машинного обучения при прогнозировании спроса, были показаны определения спроса и графики по типам моделей.

Практическая значимость исследования заключается в том, что результаты и материалы данной работы могут быть использованы при прогнозировании спроса с использованием аналогичных наборов данных.

ANNOTATION

This thesis is devoted to the use of machine learning methods in solving the problem of forecasting demand for goods. An analysis of the used data set is presented to study various models and improve the quality of forecasts. The thesis uses the most current time series forecasting methods. Graphs of model forecasts and actual values are presented, as well as the structure of the models, thanks to which it was possible to achieve a minimum forecast error in this work.

The bulk of the work involved using machine learning techniques in demand forecasting, showing demand definitions and graphs by model type.

The practical significance of the study lies in the fact that the results and materials of this work can be used in forecasting demand using similar data sets.

КІРІСПЕ

Тауарларға сұранысты болжау үлкен рөл атқарады және технологиялық өнімдерді жасауда, пайдаланушы тәжірибесін жақсартуда және ұйымның тиімді жұмыс істеуінде негізгі процестердің бірі болып табылады. Болжаудың дәлдігін арттыру арқылы тәуекелдер мен шығындарды азайтуға, сондай-ақ қорларды оңтайландыру процестерінің тиімділігін арттыруға, тұтынушылар арасында ең өзекті болып табылатын компания мен өнімдерге арналған жоспар құруға болады.

Қазіргі нарық жағдайында сұранысты болжау кез-келген компания үшін стратегиялық жоспарлаудың негізгі аспектісі болып табылады. Болашақтағы өнімге деген сұранысты дәл болжау өндіріс процестерін оңтайландыруға, тауарлы-материалдық құндылықтарды басқаруға, тиімді маркетингтік стратегияларды жасауға және тұтынушыларға қызмет көрсетудің жоғары деңгейін қамтамасыз етуге мүмкіндік береді.

Соңғы онжылдықтарда машиналық оқыту әдістері сұранысты болжау мәселелерінде кеңінен қолданыла бастады. Бұл олардың үлкен көлемдегі деректерді талдау, күрделі тәуелділіктерді анықтау және өзгертін нарықтық жағдайларға бейімделу қабілетіне байланысты. Машиналық оқыту болжау процесін автоматтандыруға, субъективтілік дәрежесін төмендетуге және болжамдардың дәлдігін арттыруға мүмкіндік береді.

Бұл жұмыстың мақсаты компания өнімдеріне сұранысты болжау міндеттерінде машиналық оқыту әдістерін қолдануды зерттеу және талдау болып табылады. Зерттеу машиналық оқыту әдістерін қолдана отырып, сұранысты болжаудың бірнеше негізгі тәсілдерін, сондай-ақ олардың артықшылықтары мен шектеулерін қарастырады. Бұдан басқа, бизнестің түрлі салаларында осы тәсілдерді табысты іске асырудың практикалық мысалдарына талдау жүргізілетін болады.

Алынған нәтижелерге сүйене отырып, компанияның өніміне сұранысты болжау үшін машиналық оқытудың қандай әдістері тиімді екендігі және оларды қолдану кезінде қандай факторларды ескеру керектігі туралы қорытынды жасалады. Алынған нәтижелер стратегиялық шешім қабылдайтын компания басшылары үшін де, аналитика және машиналық оқыту мамандары үшін де пайдалы болуы мүмкін.

1 ТЕОРИЯЛЫҚ БӨЛІМ

1.1 Компания өнімдеріне сұранысты болжау

Машиналық оқыту - бұл сұранысты болжауды дәлірек және сенімді ету үшін қолданылатын технология.

Сұранысты дәл болжау кез-келген өндірістік компанияның жетістігі үшін өте маңызды. Бұл компаниялардың қажетті тауарлық-материалдық құндылықтарға ие болуын және сұраныстың өзгеруіне тез жауап беруін қамтамасыз етеді. Соңғы жылдары машиналық оқыту сұранысты сенімді болжаудың маңызды құрамдас бөлігі болды.

Сұранысты болжау кезінде машиналық оқыту алгоритмдері сатылымның тарихи үлгілерін талдап, болашақ тенденцияларды болжай алады. Бірінші қадам - өнім түрі, сатылған саны, сатып алу жиілігі, маусымдық, жеңілдіктер және т.б. сияқты өткен сатылым деректерін жинау. Бұл деректер уақыт бойынша сату үлгілерін анықтайтын үлгілерді жасау үшін алгоритмге беріледі. Бұл үлгілер анықталғаннан кейін оларды болашақ трендтердің дәл болжамын жасау үшін пайдалануға болады.

Кәсіпорын өнімдеріне сұранысты болжау

Компанияның өніміне сұранысты болжау стратегиялық жоспарлау мен қорларды басқарудың маңызды аспектісі болып табылады. Әдетте өнімге сұранысты болжау процесіне енгізілген бірнеше қадамдар:

Деректерді жинау: Біріншіден, компания өнімдерінің тарихи сатылым деректері жиналуы керек. Бұл деректер әртүрлі уақыт кезеңдері (апталар, айлар, тоқсандар) бойынша сатылымдар туралы ақпаратты, сондай-ақ бағалар, акциялар, жарнамалық науқандар, ауа-райы және т.б. сияқты сұранысқа әсер ететін факторларды қамтуы мүмкін.

Деректерді тазалау және алдын ала өңдеу: Деректер жиналғаннан кейін оны шектен тыс мәндерден тазарту, жетіспейтін мәндерді толтыру және талдау үшін пайдалы пішімге түрлендіру қажет.

Деректерді талдау: Барлау деректерін талдау тарихи сатылым деректеріндегі негізгі тенденцияларды, маусымдылықты, циклділікті және басқа үлгілерді анықтауға көмектеседі.

Болжау әдісін таңдау: Деректерді талдау негізінде белгілі бір жағдайға ең қолайлы болжау әдісі таңдалады. Бұл уақыт сериясы әдісі, машиналық оқыту, статистикалық модельдер және т.б. болуы мүмкін.

Болжау моделін әзірлеу: әдісті таңдағаннан кейін тарихи деректер негізінде болжау моделін әзірлеу және оқыту қажет. Модельді үйрету үшін сәйкес мүмкіндіктерді (мысалы, бағалар, жарнамалық шығындар, уақыт және т.б.) пайдалану маңызды.

Модельді бағалау және параметрлерді баптау: Үлгіні оқытқаннан кейін оны сынақ деректер жинағында бағалау қажет және болжамдардың сапасын жақсарту үшін модель параметрлері реттеледі.

Сұранысты болжау: Содан кейін модель белгілі бір уақыт кезеңінде компанияның өнімдеріне болашақ сұранысты болжау үшін пайдаланылуы мүмкін.

Болжамдарды бақылау және жаңарту: Болжамдар жаңа деректер мен сыртқы ортадағы өзгерістер (мысалы, нарық конъюнктурасының өзгеруі, тұтынушылық сұраныс тенденциялары және т.

Бұл процесс қайталанатын болуы мүмкін және оның тиімділігін тәжірибе негізіндегі болжау үлгілерін үздіксіз кері байланыс пен жетілдіру арқылы жақсартуға болады.

Машиналық оқыту әдістері сұранысты болжау

Машиналық оқыту әдістерін қолдана отырып, компанияның өнімдеріне сұранысты болжау әртүрлі тәсілдер мен алгоритмдер арқылы жүзеге асырылады.

Бұлар:

Сызықтық регрессия: Бұл әдіс сандық айнымалыларды болжау үшін қолданылады. Ол өнімге сұранысты болжауда пайдалы болуы мүмкін кіріс пен шығыс арасындағы сызықтық қатынасты іздейді.

Шешім ағаштары және кездейсоқ орман: Шешім ағаштары деректерді сипаттамаларға негізделген ішкі топтарға бөледі, ал кездейсоқ орман дәлірек болжау жасау үшін шешім ағаштарының ансамблін пайдаланады. Бұл әдістер деректердегі күрделі қатынастарды модельдеу үшін тиімді болуы мүмкін.

Қолдау векторлық машинасы (SVM): SVM әртүрлі сыныптардан деректерді бөлетін оңтайлы гипержазықтықты іздейді. Бұл әдіс өнімге сұранысты жіктеу мәселелері үшін пайдалы болуы мүмкін.

Нейрондық желілер: Терең нейрондық желілер кірістер мен шығыстар арасындағы күрделі сызықты емес қатынастарды анықтай алады, бұл өнім сұранысын болжауда пайдалы болуы мүмкін.

Уақыт қатарларының әдістері: ARIMA (Авто регрессивті біріктірілген жылжымалы орташа мән), экспоненциалды тегістеу және SARIMA (маусымдық ARIMA) сияқты уақыт қатарларының әдістері уақыт өте өзгеретін сұранысты болжау үшін пайдаланылуы мүмкін.

Градиентті күшейту(бустинг) : Градиентті күшейту(бустинг) - бірнеше әлсіз үлгілерді бір күшті үлгіге біріктіретін тиімді ансамбль әдісі. Бұл сұраныс болжамдарының сапасын жақсартуға көмектесуі мүмкін.

Кластерлік талдау: кластерлік талдауды пайдалана отырып, тұтынушыларды немесе нарық сегменттерін талдау және жіктеу арқылы сұранысты болжауды жақсартуға болады.

Нақты әдісті таңдау деректердің сипаттамаларына, қолда бар ақпаратқа және сіздің компанияңыздың өніміне сұранысты болжау тапсырмасының ерекшеліктеріне байланысты. Көбінесе тиімді шешім ең жақсы болжам сапасына жету үшін бірнеше әдістерді біріктіру болып табылады.

Бұл дипломдық жұмыста машиналық оқыту әдістерін қолдана отырып, уақыт қатарларының әдістерін пайдалану арқылы жүргізіміз. Соның ішінде уақыт қатарларының әдісі ARIMA, SARIMA (маусымдық ARIMA) әдістерін қолданамыз.

1.2 Уақыт қатарын талдау

Уақыттық қатар – зерттелетін процестің кез келген параметрлерінің тең уақыт кезеңіндегі мәндерінің қатары. Уақыт қатарлары әртүрлі ғылым салаларында кездеседі: медицинадағы кардиограммалар, астрономиядағы күн белсенділігінің графиктері, экономикадағы валюта бағамы. Мұндай қатарларды талдау қиын және маңызды міндет болып табылады, оны шешу белгілі бір сала үшін пайдалы нәтиже береді.

Уақыт қатарын талдаудың негізгі мақсаттарының бірі оның әрекетін болжау болып табылады. Өткен бақылаулар негізінде болашақ құндылықтарды болжау қазіргі уақытта тиімдірек шешім қабылдауға мүмкіндік береді.

Уақыттық қатарларды талдау кезінде келесі компоненттерді ажырату әдеттегідей:

- тренд (T) – ұзақ мерзімді факторлардың әсерін сипаттайтын бірқалыпты өзгеретін құрамдас
- маусымдық компонент (S) – зерттелетін процестің циклдік ауытқуы
- кездейсоқ компонент (ε) – кездейсоқ факторлардың әсерін көрсететін компонент

Бұл компоненттерді оқшаулау талдаудың алғашқы кезеңдерінің бірі болып табылады. Осылайша, уақыт сериясының моделін $Y = T + S + \varepsilon$ – аддитивті модель және $Y = T * S * \varepsilon$ – мультипликативті модель ретінде сипаттауға болады. Екінші модель ең кең таралған болып саналады, ол өз кезегінде логарифм арқылы біріншіге дейін қысқарады.

Уақыттық қатарларды талдаудың негізгі түсініктері

Уақыттық қатар (динамикалық қатар) t уақытының дәйекті моменттеріндегі кейбір X атрибутының (кездейсоқ айнымалы) бақылаулар тізбегі ретінде түсініледі. Қатар деңгейлері жеке бақылаулар болып табылады, олар $x_t, t=1, \dots, n$ деп белгіленеді.

Уақыттық қатарды зерттеу кезінде бірнеше компоненттер бөлінеді:

$$x_t = u_t + \gamma_t + c_t + \varepsilon_t, t = 1 \dots n, \quad (1)$$

мұндағы

u_t – тренд, ұзақ мерзімді факторлардың таза әсерін сипаттайтын бірқалыпты өзгеретін компонент,

γ_t - қысқа мерзім ішінде (күн, апта және т.б.) процестердің қайталануын көрсететін маусымдық компонент;

c_t – бір жылдан астам ұзақ уақыт аралығындағы процестердің қайталануын көрсететін циклдік компонент;

ε_t – кездейсоқ құрамдас, есепке алынбайтын және тіркелмейтін кездейсоқ факторлардың әсерін көрсетеді.

Уақыт қатары $X(t)$ (t – уақыт) кездейсоқ процесінің жүзеге асырылуының бірі ретінде қарастырылады және оның мүшелері, әдетте, статистикалық тәуелсіз емес және бірдей таралады.

Уақыт қатарлары келесі сипаттамаларға сәйкес сипатталады:

- деңгейлерді көрсету формасы бойынша: абсолютті көрсеткіштер қатары, салыстырмалы көрсеткіштер, орташа мәндер;
- белгілі бір уақыт мезетіндегі деңгейлері көрсетілген көрсеткіштер саны бойынша: бір айнымалы және көп айнымалы уақыт қатарлары;
- уақыт параметрінің табиғаты бойынша: момент және интервал уақыт қатары;
- күндер мен уақыт аралықтары арасындағы қашықтық бойынша: бірдей қашықтықта, яғни кезеңдердің басталу және аяқталу даталары бір-бірін бірдей аралықпен орындағанда, толық емес – интервалдар әртүрлі.
- мәліметтердің толықтығы бойынша: толық және толық емес;
- детерминирленген және кездейсоқ. Біріншісі кейбір кездейсоқ емес функцияның мәндері негізінде алынады, екіншісі кейбір кездейсоқ шаманы жүзеге асырудың нәтижесі болып табылады;
- негізгі тенденцияның болуы бойынша: стационарлық және стационарлық емес қатарлар.

Уақыттық қатарларды талдау мен болжаудың келесі кезеңдерін бөлуге болады:

- қатардың құрамдас бөліктерін графикалық көрсету және талдау;
- қатардың стационарлықтығын тексеру;
- уақыт қатарын тегістеу және сүзу (төмен немесе жоғары жиілікті құрамдастарды жою);
- болжау моделін құру;
- бұрын жүргізілген зерттеулер негізінде уақыттық қатарды болжау.

Талдау мен болжаудың өзі басталатын маңызды кезең деректерді іздеу және оны дайындау болып табылады.

Уақыт қатарын талдау ең алдымен деректерді іздеуден және оны болжау үшін қолжетімді пішінге дейін азайтудан басталады. Зерттелетін уақыт қатарының стационарлық немесе стационарлық еместігін анықтау қажет.

Кез келген n , t , τ үшін n бақылаудың $x_1, x_2, \dots, x_{n+\tau}$ ықтималдығының бірлескен үлестірімі болса x_t , $t=1, \dots, n$ уақыт қатары қатаң стационар деп аталады. Яғни стационарлық уақыт қатарлары үшін ықтималдық сипаттамалар t моментіне тәуелді емес. Сондықтан математикалық күту $M(x_t)=a$ және стандартты ауытқуды мына формулалар арқылы x_t мәндерінен бағалауға болады:

$$a = \bar{x}_t = \sum_{t=1}^n x_t / n \quad (2)$$

$$\sigma = \sqrt{\frac{\sum_{t=1}^n (x_t - \bar{x}_t)^2}{n}} \quad (3)$$

Уақыт қатарының стационарлығынан дисперсия, күтілетін мән және ковариация сияқты қасиеттер уақыт бойынша тұрақты болатыны шығады.

Уақыттық қатарларды болжау – әрі қарай болжау және стратегиялық шешімдер қабылдау үшін статистика мен модельдеуді пайдалана отырып, уақыттық қатарлар деректерін талдау процесі. Бұл әрқашан дәл болжам бола бермейді және болжамдардың ықтималдығы, әсіресе уақыттық қатарлар деректеріндегі жиі өзгеретін айнымалылармен, сондай-ақ біздің бақылауымыздан тыс факторлармен жұмыс істегенде айтарлықтай өзгеруі мүмкін. Дегенмен, болжау басқа ықтимал нәтижелерге қарағанда қандай нәтижелердің ықтималдығы жоғары немесе азырақ болатынын түсінуге мүмкіндік береді. Көбінесе бізде деректер неғұрлым толық болса, соғұрлым болжам дәлірек болуы мүмкін.

Айнымалының болашақтағы әрекетін болжай білу, яғни уақытша қатарларды болжау үшін қандай болжау әдісін қолдануға болатынын білу үшін алдымен уақытша қатарлардың сипаттамаларын анықтау керек. Уақыт қатарын талдау осы мақсатқа қызмет етеді. Уақыт серияларын талдаудың басқа да мақсаттары бар, мысалы, уақыттық қатарлар деректерін көрсету, деректердің әрекетін түсіндіру және біз уақыттық қатарды алған процесті бақылау

Уақыт қатарларын болжау тәсілдерін статистикалық және адаптивті тәсілдерге бөлуге болады. Болжау әдісін таңдау көптеген факторларға байланысты – тарихи деректердің өзектілігі мен қолжетімділігі, қажетті дәлдік дәрежесі, болжамды уақыт кезеңі (қысқа мерзімді, орта мерзімді немесе ұзақ мерзімді болжау), болжамның құны/пайдасы. компанияға болжам және талдау үшін қолжетімді уақыт.

Кәсіпорындар өндірісті жоспарлау мен бақылауды, қысқа мерзімді ақша қаражаттарына қажеттілікті және сатудың маусымдық ауытқуларына түзетулер енгізуді қамтитын операциялық себептерге байланысты бірнеше сағаттан бірнеше күнге дейінгі уақыт аралығын болжау жасау үшін қысқа мерзімді болжауды пайдаланады.

Орта мерзімді болжамдар бірнеше айға дейінгі кезеңге бағытталған және дұрыс тактикалық шешімдер қабылдау үшін жасалады. Олар, мысалы, бюджеттеу саласында маңызды. Дұрыс емес болжау ұйым үшін ауыр зардаптарға әкелуі мүмкін, себебі ол тым оптимистік болып шықса, ұйымда, мысалы, сатылмаған өнімдер қалады және артық өндіріс болады.

Ұзақ мерзімді болжамдар ұйымда қаржылық қарқынды ресурстармен қамтамасыз етумен байланысты маңызды стратегиялық шешімдер қабылдауды қамтиды. Олар арнайы емес, жалпы мәселелермен айналысады және есептеулер үкімет саясаты, әлеуметтік өзгерістер және технологиялық өзгерістер сияқты факторларға көбірек сүйенеді.

Статистикалық болжау әдістері – қолда бар деректер негізінде болашақ мәндерді болжау үшін қолданылатын математикалық тәсілдер. Бұл әдістер уақыттық қатарларды, регрессиялық талдауды, машиналық оқытуды және т.б. қоса алғанда, әртүрлі әдістерді қамтуы мүмкін. Олар экономика, қаржы, маркетинг, деректер туралы ғылым және басқалар сияқты әртүрлі салаларда

маңызды рөл атқарады, мамандарға болжамдар мен деректерді талдау негізінде негізделген шешімдер қабылдауға көмектеседі.

Статистикалық әдістерге негізделген болжау көбінесе үлгіні дұрыс таңдауды, деректерді алдын ала өңдеуді және талдауды, нәтижелерді тексеруді және интерпретациялауды талап етеді. Бұл әдістер уақыттық қатарларды, тұтынушылардың мінез-құлқын, қаржылық нәтижелерді, нарықтық үрдістерді және т.б. болжау үшін пайдаланылады.

1.3 Статистикалық болжау әдістері

Статистикалық болжау әдістері ең объективті, сенімді және ғылыми дәлелденген әдіс. Бұл жұмыста статистикалық талдау және болжаудың статистикалық әдістері жасалады. Зерттеулер уақыт қатарлары арқылы жүзеге асырылады.

Статистикалық жүйелер арқылы уақыт қатарларын талдау және болжау процесі келесі қадамдарды қамтиды:

- жүйеге деректерді енгізу;
- әр түрлі графиктердің көмегімен деректерді визуализациялау;
- таңдалған статистикалық әдістерге сәйкес келетін деректерді түрлендіру;
- статистикалық әдістердің алгоритмдерін іске асыру;
- талдау нәтижелерін сандық және мәтіндік ақпараты бар графиктер мен кестелер түрінде шығару;
- алынған нәтижелерді түсіндіру.

СМА-қарапайым жылжымалы орташа әдіс

Мұнда барлық болашақ мәндердің болжамдары тарихи деректердің орташа мәніне тең. Егер біз тарихи деректерді y_1, \dots, y_T деп белгілесек, онда болжамдарды деп жаза аламыз.

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T) / T \quad (4)$$

Жылжымалы орташа мәндерді сатудың жылжымалы орташа көрсеткішпен түсірілген үлгіге байланысты өсу трендінде немесе төмендеу трендінде екенін жылдам анықтау үшін пайдалануға болады, яғни трендтерді оңай тануға болатындай, ауытқуларды (шындар мен ең төмен нүктелер) тегістеу үшін жылжымалы орташа мән қолданылады.

Статистикалық болжау әдістері - негізгі міндеттеріне объективті деректер негізінде болжаудың қазіргі заманғы математикалық-статистикалық әдістерін әзірлеу, зерделеу және қолдану; болжаудың сараптамалық әдістерін ықтималдық-статистикалық модельдеу теориясы мен практикасын дамыту; тәуекел жағдайында болжау әдістері және экономикалық-математикалық және эконометрикалық әдістерді бірлесіп пайдалана отырып, болжаудың аралас әдістері кіретін ғылыми және оқу пәні (математикалық-статистикалық және

сараптамалық) модельдер. Болжаудың статистикалық әдістерінің ғылыми базасы қолданбалы статистика және шешім қабылдау теориясы болып табылады.

Болжамның дәлдігін бағалау болжау процедурасының қажетті бөлігі болып табылады. Әдетте тәуелділікті қалпына келтірудің ықтималдық-статистикалық модельдері қолданылады, мысалы, максималды ықтималдылық әдісі бойынша ең жақсы болжам жасау. Параметрлік (әдетте қалыпты қателіктер моделі негізінде) және параметрлік емес болжамның дәлдігін бағалау және оған сенімділік шекаралары (ықтималдықтар теориясының орталық шекті теоремасы негізінде) әзірленді.

Көпөлшемді регрессия, оның ішінде таралу тығыздығының параметрлік емес бағаларын қолдана отырып, қазіргі кездегі негізгі статистикалық болжау аппараты болып табылады. Қалыпты болжамды қабылдамау үшін ықтималдық теориясының көп өлшемді Орталық шекті теоремасына, сызықтық технологияға және конвергенция мұрагерлігіне негізделген басқа математикалық аппаратқа сүйену керек. Бұл параметрлерді нүктелік және интервалдық бағалауға, олардың параметрлік емес қойылымдағы 0-ден айырмашылығының маңыздылығын тексеруге, болжам үшін сенімділік шекараларын құруға мүмкіндік береді.

Модельдің сәйкестігін тексеру мәселесі, сондай-ақ факторларды таңдау мәселесі өте маңызды. Реакцияға әсер ететін факторлардың априорлық тізімі әдетте өте кең, оны қысқартқан жөн және қазіргі заманғы зерттеулердің негізгі бағыт «ақпараттық белгілердің» таңдау әдістеріне бағытталған. Алайда, бұл мәселе әлі шешілген жоқ. Ерекше әсерлер пайда болады. Сонымен, көпмүшелік дәрежесінің жиі қолданылатын бағалары асимптотикада геометриялық үлестірімге ие екендігі анықталды. Ықтималдық тығыздығын бағалаудың параметрлік емес әдістері және оларды ерікті түрдің регрессиялық тәуелділігін қалпына келтіру үшін қолдану перспективалы. Бұл саладағы ең көп таралған нәтижелер сандық емес статистикалық тәсілдер арқылы алынады.

Компьютерлерді қарқынды қолданатын болжау жүйелері болжаушының бірыңғай автоматтандырылған жұмыс орны шеңберінде әртүрлі болжау әдістерін біріктіреді.

Сандық емес сипаттағы мәліметтер негізінде болжау, атап айтқанда сапалық белгілерді болжау сандық емес деректер статистикасының нәтижелеріне негізделген. Интервалдық деректерге негізделген регрессиялық талдау, оның ішінде ноталық және ұтымды іріктеу көлемін анықтау және есептеу, сондай - ақ бұлыңғыр деректерді регрессиялық талдау сандық емес деректер статистикасы шеңберінде регрессиялық талдаудың жалпы тұжырымы және оның ерекше жағдайлары-дисперсиялық талдау және дискриминантты талдау, соның ішінде интервалдық деректер негізінде өте перспективалы болып көрінеді. формальды түрде әртүрлі әдістерге көзқарас болжаудың заманауи статистикалық әдістерін бағдарламалық қамтамасыз етуде пайдалы.

Заманауи компьютерлік болжау технологиялары эконометрикалық мәліметтер базасын, имитациялық және сараптамалық, математикалық-статистикалық және модельдеу блоктарын біріктіретін Экономикалық-

математикалық динамикалық модельдерді қолдана отырып, болжаудың интерактивті статистикалық әдістеріне негізделген.

Болжаудың қазіргі статистикалық әдістеріне авторегрессия модельдері, Бокс-Дженкинс моделі, нейрондық желі моделі, параметрлік және параметрлік емес тәсілдерге негізделген эконометрикалық теңдеулер жүйесі жатады.

1.4 Болжау әдістері

Уақыт қатарларын болжау әдістерінің ішінде бірнеше негізгі топтар бар:

- регрессия әдістері
- нейрондық желі әдістері
- Бокс-Дженкинс әдістері

Регрессия әдістері

Болжалды айнымалы бірнеше тәуелсіз айнымалыларға тәуелді, олардың сипаты әртүрлі болуы мүмкін. Жалпы, көп регрессия моделі келесідей пішінге ие:

$$Y = F(x_1, x_2, \dots, x_n) + \varepsilon \quad (5)$$

мұндағы F – n тәуелсіз айнымалының функциясы, ε – қате компоненті. Ерекше жағдай кеңінен қолданылатын сызықтық регрессия моделі болып табылады:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (6)$$

Мұндағы β_0, \dots, β_n – таңдалған регрессия коэффициенттері.

Нейрондық желі әдістері

Болжау үшін нейрондық желілерді пайдаланатын әдістердің ең заманауи және барған сайын дамып келе жатқан тобы.

Бокс-Дженкинс әдістері.

Бокс-Дженкинстің классикалық әдістерінде тәуелсіз айнымалылар қолданылмайды. Болашақ мәндерді болжау уақыттық қатардың алдыңғы мәндерінің тарихына ғана сүйенеді.

2 Компания өнімдеріне сұранысты талдау

Сатуды болжауға арналған мәліметтер жиынтығы машиналық оқыту әдістерін оқыту және тестілеу үшін Kaggle сайтынан алынған «Store Item Demand Forecasting Challenge» деректер жиынтығы пайдаланылды. Бұл 5 жыл ішінде дүкендерде тауарларды сату туралы ақпаратты қамтитын салыстырмалы түрде қарапайым және таза деректер жиынтығында уақыт қатарларының әртүрлі әдістерін зерттеуге мүмкіндік береді.

Деректер жиынтығы 3 кестеден тұрады:

« train.csv» - жаттығу деректері

« test.csv »- сынақ деректері

« sample_submission.csv » - нәтижелерді ұсыну файлының үлгісі.

Деректер келесі бағандардан тұрады:

« date » – тауарды сату күні,

« store » – дүкен идентификаторы,

« item» – тауар идентификаторы,

«sales» – белгілі бір дүкенде белгілі бір күні сатылған тауарлар саны.

Жаттығу деректерінде 913000 жол және 4 баған бар. Сынақ деректерінде 45000 жол және 4 баған бар. Сынақ деректеріндегі "id" бағаны нәтижелерді ұсыну файлының үлгісімен байланысу үшін қажет.

2.1 Эксперименттер жүргізу және нәтижелерді талдау

	date	store	item	sales	id	date	store	item	
0	2013-01-01	1	1	13	0	0	2018-01-01	1	1
1	2013-01-02	1	1	11	1	1	2018-01-02	1	1
2	2013-01-03	1	1	14	2	2	2018-01-03	1	1
3	2013-01-04	1	1	13	3	3	2018-01-04	1	1
4	2013-01-05	1	1	10	4	4	2018-01-05	1	1
...
912995	2017-12-27	10	50	63	44995	44995	2018-03-27	10	50
912996	2017-12-28	10	50	59	44996	44996	2018-03-28	10	50
912997	2017-12-29	10	50	74	44997	44997	2018-03-29	10	50
912998	2017-12-30	10	50	62	44998	44998	2018-03-30	10	50
912999	2017-12-31	10	50	82	44999	44999	2018-03-31	10	50
913000 rows x 4 columns					45000 rows x 4 columns				

Сурет - 1 – Жаттығу мен сынақ деректеріндегі жолдар мен бағандары өлшемі

Практикалық тәжірибе нәтижелері

Жаттығу деректері және сынақ деректерінің жалпы көрсеткіштері көрсетілген, алынған нәтижелерге сәйкес жолдар мен бағандар саны, жалпылама ақпарат алуға болады.

```
train.describe()
```

	store	item	sales
count	913000.000000	913000.000000	913000.000000
mean	5.500000	25.500000	52.250287
std	2.872283	14.430878	28.801144
min	1.000000	1.000000	0.000000
25%	3.000000	13.000000	30.000000
50%	5.500000	25.500000	47.000000
75%	8.000000	38.000000	70.000000
max	10.000000	50.000000	231.000000

Сурет - 2 – Train жаттығу деректері

```
print('Start: ', train['date'].min().date())  
print('End: ', train['date'].max().date())
```

Start: 2013-01-01

End: 2017-12-31

Сурет - 3 – Train бастапқы және соңғы мәліметтер тізімі

Сынақ деректері мәліметтерінің көрсеткіштері.

```
test.head()
```

	id	date	store	item
0	0	2018-01-01	1	1
1	1	2018-01-02	1	1
2	2	2018-01-03	1	1
3	3	2018-01-04	1	1
4	4	2018-01-05	1	1

Сурет – 4 – Сынақ деректері

```
test.describe()
```

	id	store	item
count	45000.000000	45000.000000	45000.000000
mean	22499.500000	5.500000	25.500000
std	12990.525394	2.872313	14.43103
min	0.000000	1.000000	1.000000
25%	11249.750000	3.000000	13.000000
50%	22499.500000	5.500000	25.500000
75%	33749.250000	8.000000	38.000000
max	44999.000000	10.000000	50.000000

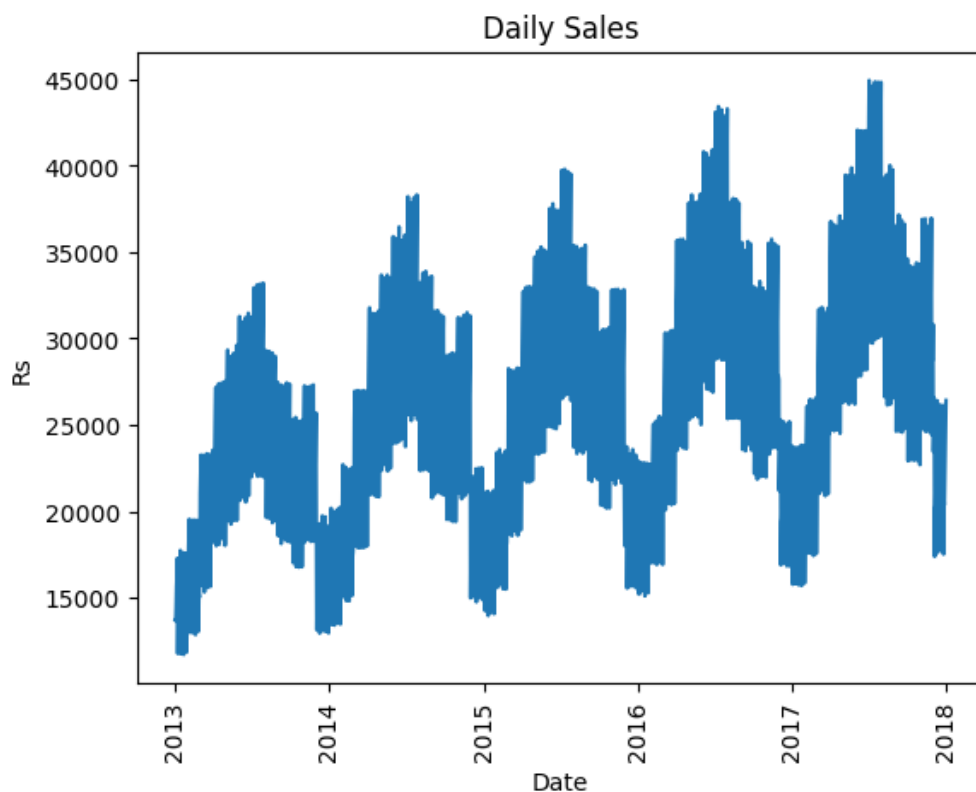
Сурет - 5– Сынақ деректерінің көрсеткіші

Жаттығу деректерінде 913000 жол және 4 баған бар. Сынақ деректерінде 45000 жол және 4 баған бар. Сынақ деректеріндегі "id" бағаны нәтижелерді ұсыну файлының үлгісімен байланысу үшін қажет.

```
store_daily_sales.head()
```

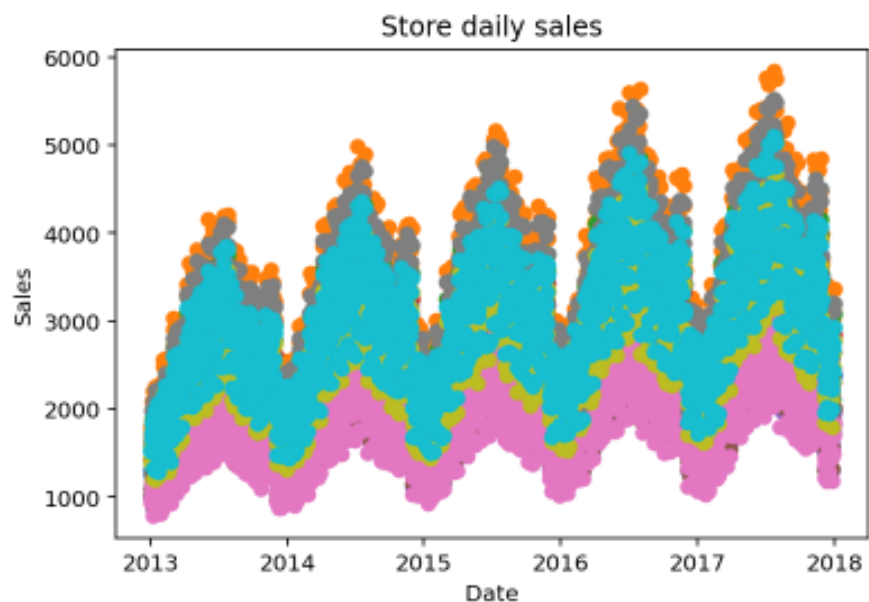
	store	date	sales
0	1	2013-01-01	1316
1	1	2013-01-02	1264
2	1	2013-01-03	1305
3	1	2013-01-04	1452
4	1	2013-01-05	1499

Сурет - 6 –Мәліметтер жиынтығы

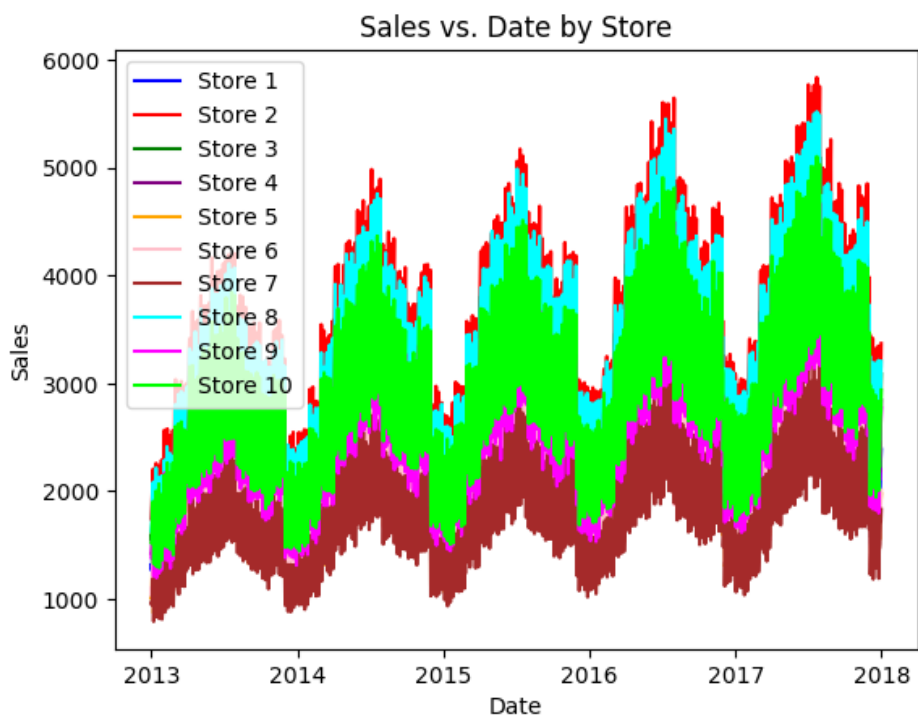


Сурет - 7 –Жылдық көрсетілімдер

Оқыту «train.csv», онда 01.01.2013 – ден 09.30.2017 - ға дейінгі сату күндері болжау модельдерін үйрету үшін, ал 10.01.2018 -ден 12. 31.2018-ге дейін тестілеу үшін пайдаланылады.

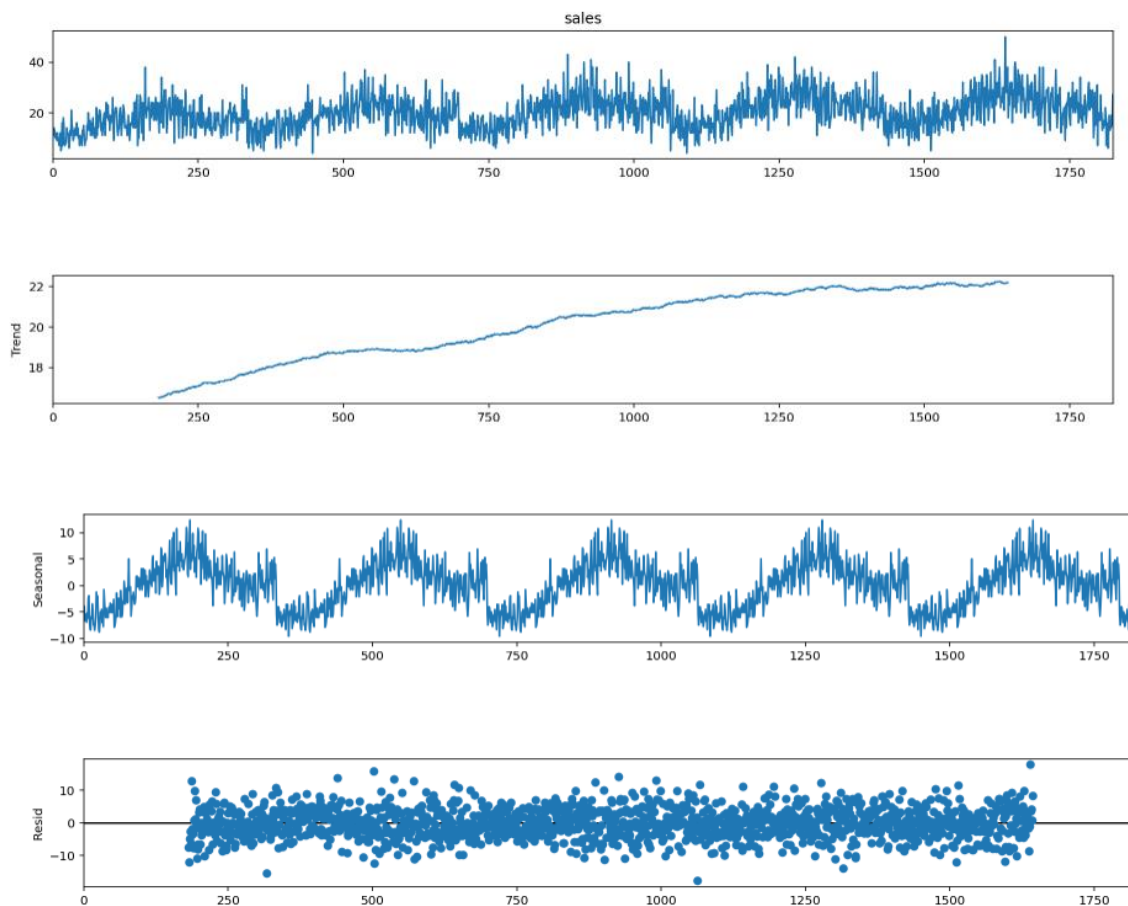


Сурет - 8 – Сатылымдар көрсеткіші



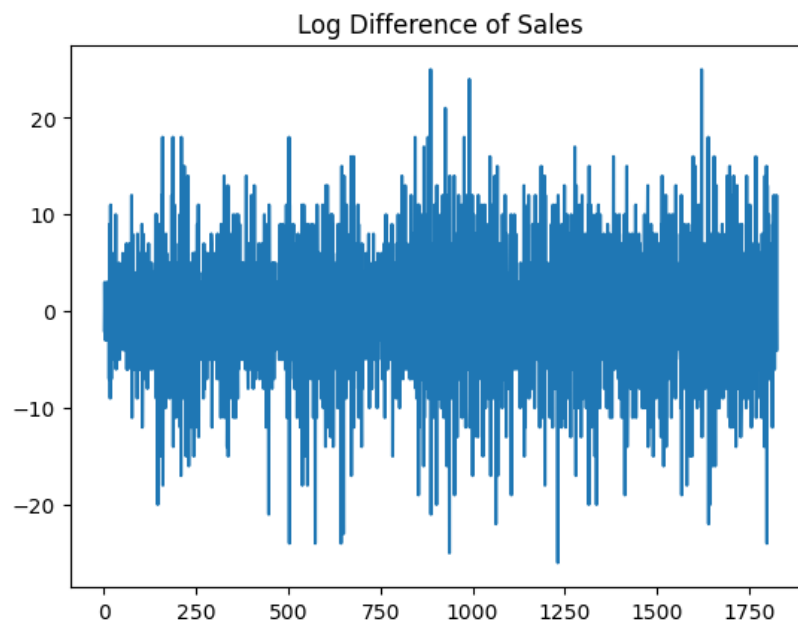
Сурет - 9 – Бұл 10 түрлі дүкенде 50 түрлі тауарлардың салыстырмалы деректері

Көрсетілген уақыт қатары тұрақсыз және маусымдық сипатқа ие, сатылымдар әрқашан жылдың басында төмен, жылдың ортасында жоғары, ал жылдың соңында сатылымдар қайтадан төмен. Бұл деректер стационарлық емес дегенді білдіреді, өйткені орташа уақыт өткен сайын артады.



Сурет - 10 – Уақыт қатарларының ыдырауының нәтижесі

Белгілі бір дүкенде белгілі бір күні сатылған тауарлар санының журналдық айырмашылығы көрсетілген.



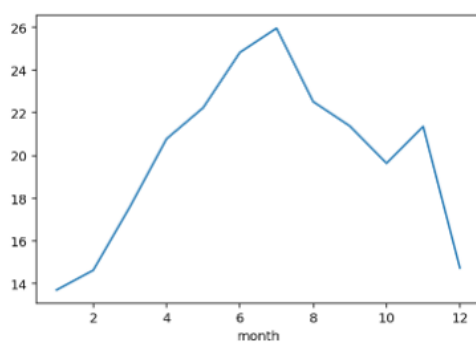
Сурет -11 –Тауарлар санының журналдық айырмашылығы

ARIMA моделі практикалық тәжірибе нәтижелері

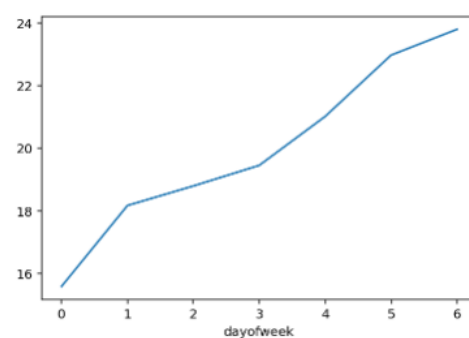
	(p, d, q)	AIC
0	(6, 1, 7)	10902.795288
1	(7, 1, 7)	10916.636960
2	(5, 1, 5)	10994.688059
3	(6, 1, 5)	11015.168593
4	(4, 1, 7)	11020.827036
...
59	(4, 1, 0)	11616.440522
60	(3, 1, 0)	11701.431885
61	(2, 1, 0)	11805.944671
62	(1, 1, 0)	11963.116936
63	(0, 1, 0)	12327.036154

64 rows × 2 columns

Сурет -12 – Arima моделінің көрсеткіші



Сурет -13 – Күндік көрсеткіш

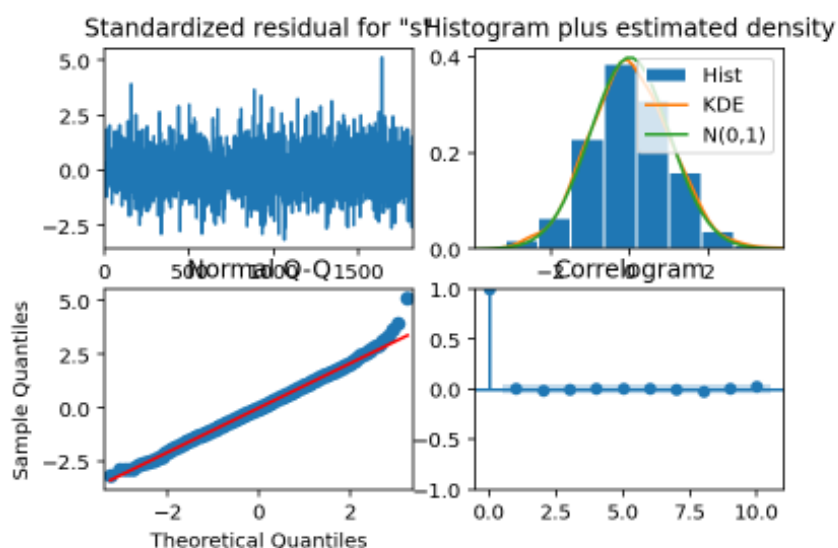


Сурет -14 – Апталық көрсеткіш

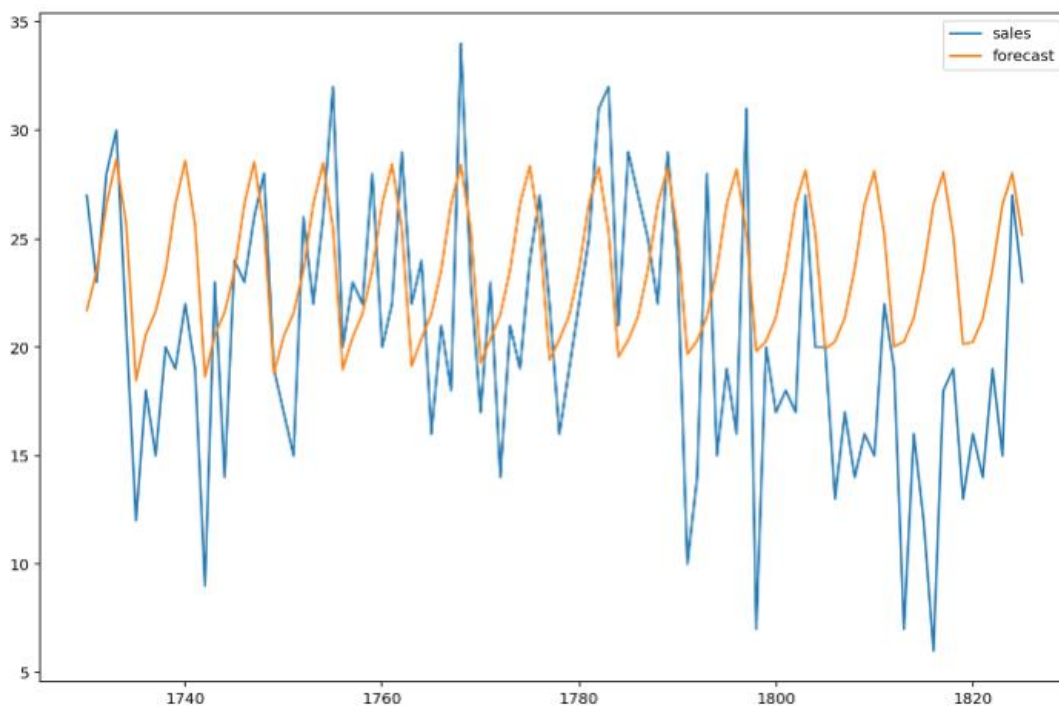
SARIMA моделі практикалық тәжірибе нәтижелері

SARIMAX Results						
Dep. Variable:	sales	No. Observations:	1826			
Model:	SARIMAX(6, 1, 7)	Log Likelihood	-5437.398			
Date:	Tue, 28 May 2024	AIC	10902.795			
Time:	11:55:28	BIC	10979.926			
Sample:	0	HQIC	10931.247			
			- 1826			
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.9791	0.012	-84.028	0.000	-1.002	-0.956
ar.L2	-0.9757	0.010	-100.387	0.000	-0.995	-0.957
ar.L3	-0.9794	0.009	-105.835	0.000	-0.998	-0.961
ar.L4	-0.9809	0.009	-110.086	0.000	-0.998	-0.963
ar.L5	-0.9849	0.009	-106.731	0.000	-1.003	-0.967
ar.L6	-0.9620	0.012	-82.452	0.000	-0.985	-0.939
ma.L1	0.0709	0.023	3.107	0.002	0.026	0.116
ma.L2	0.0760	0.024	3.216	0.001	0.030	0.122
ma.L3	0.0892	0.023	3.928	0.000	0.045	0.134
ma.L4	0.1006	0.022	4.533	0.000	0.057	0.144
ma.L5	0.0989	0.023	4.223	0.000	0.053	0.145
ma.L6	0.0509	0.023	2.244	0.025	0.006	0.095
ma.L7	-0.7803	0.023	-34.431	0.000	-0.825	-0.736
sigma2	21.1299	0.619	34.153	0.000	19.917	22.342
Ljung-Box (L1) (Q):		0.27	Jarque-Bera (JB):		15.67	
Prob(Q):		0.60	Prob(JB):		0.00	
Heteroskedasticity (H):		1.34	Skew:		0.14	
Prob(H) (two-sided):		0.00	Kurtosis:		3.35	

Сурет - 15– SARIMA статистикалық мәліметтер жиынтығы



Сурет - 16 – SARIMA Диагностика



Сурет - 17 – SARIMA сатуды болжау

	date	store	item	sales	day	month	year	dayofweek	model
0	2013-01-01	1	1	13	1	1	2013	1	0.000000
1	2013-01-02	1	1	11	2	1	2013	2	12.999749
2	2013-01-03	1	1	14	3	1	2013	3	11.867910
3	2013-01-04	1	1	13	4	1	2013	4	12.888250
4	2013-01-05	1	1	10	5	1	2013	5	12.875858
5	2013-01-06	1	1	12	6	1	2013	6	11.837706
6	2013-01-07	1	1	10	7	1	2013	0	12.106872
7	2013-01-08	1	1	9	8	1	2013	1	11.984759
8	2013-01-09	1	1	12	9	1	2013	2	11.399865
9	2013-01-10	1	1	9	10	1	2013	3	12.292798
10	2013-01-11	1	1	9	11	1	2013	4	11.566164
11	2013-01-12	1	1	7	12	1	2013	5	10.529704
12	2013-01-13	1	1	10	13	1	2013	6	10.369958
13	2013-01-14	1	1	12	14	1	2013	0	10.245113
14	2013-01-15	1	1	5	15	1	2013	1	10.732498

Сурет -18– Мәліметтер көрсетілімі

3 ЖАЛПЫ АНЫҚТАМАЛАР

3.1 Уақыт қатарларының модельдері

Тұрақты аралықтармен өлшенетін бақылаулардан тұратын уақыттық қатарлар деректері әртүрлі салаларда кеңінен қолданылады. Осы деректерге негізделген болашақ құндылықтарды дәл болжау негізделген шешімдер қабылдау үшін өте маңызды. Уақыттық қатарларды болжауда екі қуатты статистикалық модель, ARIMA және SARIMA кеңінен қолданылады.

3.1.1 ARIMA моделі

ARIMA (авторегрессивті интеграцияланған жылжымалы орташа) танымал уақыттық қатарларды болжау үлгісі. Ол әртүрлі салаларда уақыттық қатар деректер жинағындағы өткен бақылауларға негізделген болашақ мәндерді талдау және болжау үшін кеңінен қолданылады.

ARIMA моделі үш негізгі компонентті біріктіреді: авторегрессия (AR), Интегралды (I) және жылжымалы орташа (MA).

ARIMA өзегін ашу: құрылыс блоктары мен механизмдері

Авторегрессивті (AR) компонент: өткенмен байланыс.

ARIMA моделі уақыттық қатарларды болжауда шешуші рөл атқаратын авторегрессивті компонентке негізделген. Тарихи баға деректеріне сүйене отырып, ертеңгі акция бағасын болжауға тырысып көріңіз. AR ARIMA құрамдас бөлігі ертеңгі бағаға өзінің бұрынғы мәндері әсер ететінін мойындайды. Негізінде ол ағымдағы бақылау мен артта қалған бақылаулар арасындағы корреляцияны зерттейді. Өткен тенденциялардың болашақ құндылықтарға қалай әсер ететінін түсіну арқылы AR құрамдас бөлігі ARIMA моделінің болжамдық күшінің негізін құрайды.

$$y_t = c + \sum_{n=1}^p a_n y_{t-n} + \varepsilon_t \quad (7)$$

Интегралды (I) компонент: стационарлы емес стационарға түрлендіру

Тұрақтылық – уақыттық қатарларды талдаудағы негізгі ұғым. Көптеген статистикалық әдістер деректердің орташа мәні мен дисперсиясы уақыт өте тұрақты болып қалады деп болжайды, бұл көбінесе нақты деректерде болмайды. Біріктірілген ARIMA компоненті бұл мәселені айырмашылықтарды қосу арқылы шешеді¹. Дифференциация алдыңғы бақылауды ағымдағыдан алып тастауды, трендтерді тиімді жоюды және деректерді стационарлық етуді қамтиды. Бұл түрлендіру ARIMA-ға стационарлық деректер болжамында жұмыс істеуге мүмкіндік береді, дәлірек болжам жасауға мүмкіндік береді.

Жылжымалы орташа (MA) құрамдас бөлігі: Қысқа мерзімді ауытқуларды түсіру

Авторегрессияның, интеграцияның және жылжымалы орташаның құрамдас бөліктері арасындағы синергия ARIMA моделінің пайда болуына әкеледі. Математикалық түрде ARIMA(p, d, q) ретінде ұсынылған, ол үш маңызды параметрді қамтиды: p (AR реті), d (дифференциалдау реті) және q (MA реті). Бұл параметрлер үлгілерді анықтау және болжамдар жасау кезінде ARIMA моделінің әрекеті мен тиімділігін анықтайды. Осы құрамдастарды тиімді біріктіру арқылы ARIMA уақыттық қатарлар деректерін талдау және болжамдарды жасау үшін жан-жақты құрылымды қамтамасыз етеді.

$$y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \dots + \omega_q \varepsilon_{t-q} + \varepsilon_t \quad (8)$$

Маусымдық навигацияны пайдаланып парадигманы дамыту

ARIMA уақыттық қатарларды болжауға арналған қуатты құрал болғанымен, ол маусымдылықты көрсететін деректермен жұмыс істеу кезінде қиындықтарға тап болады — белгіленген уақыт аралықтарында үлгілерді қайталайды. Болжаудың дәлдігін жақсарту үшін SARIMA қолдаған жөн.

3.1.2 SARIMA (маусымдық ARIMA) моделі

SARIMA(p, d, q)(P, D, Q)_s моделі айқын маусымдық компоненті бар уақыттық қатарларға ARIMA моделінің жалпылауы болып табылады. Сонымен қатар, мұндай модельде процестің циклдік ауытқуларын есепке алу үшін маусымдық параметрлер (P, D, Q, s) енгізіледі.

SARIMA (маусымдық ARIMA) енгізіңіз. Бұл модель ARIMA моделіне өте ұқсас, тек авторегрессивті және қозғалмалы орташа құрамдастардың қосымша жиынтығы бар. Қосымша кешігулер маусымдық жиілікпен өтеледі (мысалы, 12 – ай сайын, 24 – сағат сайын).

$$y_t = c + \sum_{n=1}^p a_n y_{t-n} + \sum_{n=1}^q \theta_n \varepsilon_{t-n} + \sum_{n=1}^P \varphi_n y_{t-sn} + \sum_{n=1}^Q \eta_n \varepsilon_{t-sn} + \varepsilon_t \quad (9)$$

SARIMA (маусымдық авторегрессивті біріктірілген жылжымалы орташа көрсеткіш) маусымдық емес құрамдастарға қосымша маусымдықты қамтитын ARIMA (авторегрессивті біріктірілген жылжымалы орташа) үлгісінің кеңейтімі болып табылады. ARIMA үлгілері уақыттық қатарларды талдау және болжау үшін кеңінен қолданылады, ал SARIMA үлгілері маусымдық үлгілері бар деректерді өңдеу үшін арнайы әзірленген.

SARIMA моделі келесідей ұсынылған:

m = жылына бақылаулар саны; P = маусымдық AR терминдерінің саны;

D = маусымдық айырмашылықтар саны; Q = маусымдық терминдер саны

MA

Үлгінің маусымдық бөліктері үшін бас әріптерді, ал маусымдық емес бөліктері үшін кіші әріптерді қолданамыз.

SARIMA үлгілерінің маусымдық құрамдас бөлігі келесі үш компонентті қосады:

Маусымдық авторегрессия (P): Бұл құрамдас серияның ағымдағы мәні мен оның өткен мәндері арасындағы, әсіресе маусымдық кешігулермен байланысты анықтайды.

Маусымдық интеграция (D): Маусымдық емес айырмашылыққа ұқсас, бұл құрамдас сериядан маусымдықты жою үшін қажетті айырмашылықты есептейді.

Маусымдық жылжымалы орташа (Q): Бұл құрамдас ағымдағы мән мен маусымдық кешігулермен алдыңғы болжамдардың қалдық қателері арасындағы қатынасты модельдейді.

ARIMA(1,0,1)(2,1,0)12 моделі математикалық түрде өрнектеледі:

$$y_t - y_{t-12} = W_t \quad (10)$$

$$W_t = \mu + \varphi_1 W_{t-1} + \varphi_2 W_{t-12} + \varphi_3 W_{t-24} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (11)$$

3.1.3 ARIMA vs SARIMA ерекшеліктері

ARIMA vs SARIMA: маусымдық

ARIMA және SARIMA үлгілерінің арасындағы негізгі айырмашылық олардың уақыттық қатар деректерінің маусымдылығын өңдеу әдісі болып табылады.

ARIMA (Авторегрессивті біріктірілген жылжымалы орташа мән): Бұл модель деректердегі трендтер мен қысқа мерзімді тәуелділіктерді анықтау үшін тиімді, бірақ стационарлықты болжайды, яғни статистикалық сипаттар (орташа және дисперсия сияқты) уақыт бойынша тұрақты болады. Деректер белгілі бір аралықтарда қайталанатын үлгілерді көрсеткенде, ARIMA күреседі, мысалы, мереке күндерінің айналасындағы шыңдары бар ай сайынғы сату көрсеткіштері.

SARIMA (маусымдық ARIMA): Бұл маусымдықты есепке алу үшін арнайы жасалған ARIMA кеңейтімі. SARIMA маусымдық әсерлерді есепке алу үшін қосымша параметрлерді қамтиды. Негізінде ол маусымдық емес үлгілер үшін ARIMA-ның күшті жақтарын қайталанатын маусымдық циклдарды модельдеу мүмкіндігімен біріктіреді. Маусымдық қосымша параметрлерді (P, D, Q) және маусымдық кезеңді (m) қолдану арқылы есепке алынады. Бұл параметрлер өткен маусымдық мәндердің болашақ болжамдарға әсерін модельдейді.

ARIMA vs SARIMA: үлгі икемділігі

Модельдің икемділігіне келетін болсақ, ARIMA және SARIMA ымыраға келуді ұсынады:

ARIMA: ARIMA оңайырақ, бағалау үшін параметрлері аз. Бұл оны әртүрлі маусымдық емес деректер үлгілеріне көбірек бейімдейді. Ол үрдістерді,

циклдарды және кездейсоқ ауытқуларды негізгі процестің егжей-тегжейіне кірместен басқара алады.

SARIMA: SARIMA қосымша маусымдық параметрлерді енгізеді, бұл оны ARIMA-мен салыстырғанда маусымдық емес деректерге икемді етеді. Бұл қарапайым болжау тапсырмалары үшін тым күрделі болуы мүмкін. SARIMA-ның шынайы күші оның қайталанатын маусымдық циклдерді модельдеу қабілетінде жатыр. Бұл маусымдықты көрсететін деректер үшін оны әлдеқайда икемді етеді. Ол өткен маусымдық мәндердің болашақ болжамдарға әсерін басып, дәлірек болжам жасауға мүмкіндік береді.

ARIMA vs SARIMA: Болжам дәлдігі

Болжаудың дәлдігіне келетін болсақ, SARIMA әдетте ARIMA-дан асып түседі, бірақ есте сақтау керек кейбір жағдайлар бар:

Маусымдық күші. Деректеріңіздегі маусымдық үлгілер неғұрлым күшті болса, SARIMA ұсынатын пайдасы соғұрлым көп болады. Нәзік маусымдылыққа байланысты ARIMA мен SARIMA арасындағы айырмашылық аз болуы мүмкін.

Деректер ұзындығы. Үлкенірек деректер жиынтығымен ARIMA және SARIMA екеуі де дәлірек болжамдар жасай алады. Дегенмен, SARIMA маусымдық өңдеуінің артықшылығы бірнеше маусымдық циклдарды қамтитын ұзағырақ деректер жиынын пайдаланған кезде айқынырақ болады.

Үлгіні таңдау және параметрлерді орнату. Үлгі параметрлерін оңтайландыру (ARIMA үшін p, d, q және SARIMA үшін P, D, Q, m) екі үлгі үшін де маңызды. Тор іздеу немесе статистикалық сынақтар сияқты әдістерді пайдалану нақты деректеріңіз үшін ең жақсы конфигурацияны анықтауға көмектеседі.

SARIMA әдетте маусымдықты есепке алу мүмкіндігіне байланысты дәлірек болжамдарды ұсынса да, бұл әрқашан ең жақсы таңдау бола бермейді.

Деректеріңіз күшті маусымдық үлгілерді көрсетсе, SARIMA-ға басымдық беріңіз.

Таза маусымдық емес деректер үшін ARIMA-ны қарастырған жөн, әсіресе үлгіні түсіндіру маңызды болса.

3.1.4 ARIMA мен SARIMA мәні

Деректерді талдаудың үнемі дамып келе жатқан әлемінде ARIMA және SARIMA үлгілері болжамды аналитиканың негізі болып қала береді. Олардың тарихи үрдістерді түсіру, маусымдықты есепке алу және болашақ құндылықтарды болжау қабілеті шешім қабылдаушыларға құнды ақпарат береді. Уақыттық қатарлар деректерінің көлемі өскен сайын, бұл модельдердің маңыздылығы тек артады. ARIMA және SARIMA принциптері мен әдістерін меңгере отырып, талдаушылар мен деректер ғалымдары өткеннің баға жетпес ақпаратын ашу және болашақтың күрделілігін түсіну үшін құралдармен жабдықталған.

4 Нәтижелер

Бұл 5 жыл ішінде дүкендерде тауарларды сату туралы ақпаратты қамтитын салыстырмалы түрде қарапайым және таза деректер жиынтығында уақыт қатарларының әртүрлі әдістерін зерттеуде шыққан нәтижелер жиынтығы. 10 түрлі дүкенде 50 түрлі тауарлардың салыстырмалы деректерінің қорытынды мәліметтері.

```
item: 1 store: 1 Finished.    item: 50 store: 1 Finished.
item: 1 store: 2 Finished.    item: 50 store: 2 Finished.
item: 1 store: 3 Finished.    item: 50 store: 3 Finished.
item: 1 store: 4 Finished.    item: 50 store: 4 Finished.
item: 1 store: 5 Finished.    item: 50 store: 5 Finished.
item: 1 store: 6 Finished.    item: 50 store: 6 Finished.
item: 1 store: 7 Finished.    item: 50 store: 7 Finished.
item: 1 store: 8 Finished.    item: 50 store: 8 Finished.
item: 1 store: 9 Finished.    item: 50 store: 9 Finished.
item: 1 store: 10 Finished.   item: 50 store: 10 Finished.
```

Сурет -19 – Мәліметтер көрсетілімі

	id	sales
0	0	12.715270
1	1	15.172579
2	2	16.224379
3	3	16.294104
4	4	18.691853
...
44995	44995	86.819947
44996	44996	87.213439
44997	44997	93.933543
44998	44998	98.422761
44999	44999	102.964935

45000 rows × 2 columns

Сурет - 20 – Sample деректері

ARIMA және SARIMA үлгілері уақыттық қатар деректеріне енгізілген үлгілерді, трендтерді және циклдік ырғақтарды ашып, өткен мен болашақ арасындағы көпір ретінде қызмет етеді.

Тұтастай алғанда, ARIMA және SARIMA әдістері уақыт серияларын, соның ішінде компания өнімдеріне сұранысты болжауға арналған қуатты құралдар болып табылады және деректердегі үлгілерді анықтау және модельдеу үшін сәтті қолданылуы мүмкін.

ҚОРЫТЫНДЫ

Бұл дипломдық жұмыс тауарларға сұранысты болжау мәселесін шешуде машиналық оқыту әдістерін қолдануға арналған. Әр түрлі модельдерді зерттеуге және болжау сапасын жақсартуға арналған мәліметтер жиынтығына талдау жасалады. Бұл дипломдық жұмыста уақыт қатарларын болжаудың ең өзекті тәсілі қолданылды, модельдер мен нақты мәндердің болжамды графиктері ұсынылған.

Сұранысты болжау кәсіпорындарға тауарлы-материалдық құндылықтарды басқарудан бастап жеткізілім тізбегін оңтайландыруға дейінгі көптеген факторларға әсер ететін дұрыс шешім қабылдауға көмектеседі.

Компания өнімдеріне сұранысты болжау кезінде Машиналық оқыту әдістерін қолдану болжамдардың дәлдігі мен сенімділігін арттыру бағытындағы маңызды қадам болып табылады. Машиналық оқыту әртүрлі айнымалылар арасындағы күрделі қатынастарды ескеруге және модельдерді өзгеретін нарық жағдайларына бейімдеуге мүмкіндік береді.

Машиналық оқыту әдістерін қолдану арқылы компаниялар өндіріс процестерін оңтайландыруға, тауарлы-материалдық құндылықтарды басқаруға және тұтынушыларға қызмет көрсетуді жақсартуға мүмкіндік беретін сұраныстың дәлірек және өзекті болжамдарын ала алады. Бұл өз кезегінде бизнестің тиімділігін арттыруға және шығындарды азайтуға ықпал етеді.

Осылайша, компанияның өніміне сұранысты болжау кезінде Машиналық оқыту әдістерін қолдану қазіргі заманғы нарықтың серпінді және бәсекеге қабілетті ортасында неғұрлым негізделген және табысты бизнес шешімдерін жасауға көмектесетін маңызды стратегиялық инвестиция болып табылады. Arima (Autoregressive Integrated Moving Average) және SARIMA (Seasonal Autoregressive Integrated Moving Average) әдістері уақыт қатарын модельдеу мен болжаудың қуатты құралдары болып табылады.

ARIMA және SARIMA әдістері туралы қорытындылар:

ARIMA-бұл авторегрессияны (AR), интеграцияны (I) және жылжымалы орташа мәнді (MA) ескеретін модель. Бұл әдіс деректердегі күрделі заңдылықтарды модельдеуге және уақыт қатарының ағымдағы және өткен мәндерін ескеруге мүмкіндік береді.

SARIMA -модельге маусымдық (s) есебін қосатын ARIMA әдісінің кеңейтімі. Бұл маусымдық ауытқулармен уақыт қатарларын тиімді модельдеуге және болжауға мүмкіндік береді.

Екі әдіс те компанияның өніміне сұранысты болжауда пайдалы болуы мүмкін, өйткені олар деректердегі уақыт заңдылықтарын, маусымдылықты және трендтерді ескеруі мүмкін.

ARIMA мен SARIMA-ның артықшылықтарының бірі-олардың экзогендік айнымалыларды нақты орнатуды қажет етпестен стационарлық уақыт қатарларын автоматты түрде өңдеу мүмкіндігі.

Дегенмен, ARIMA және SARIMA-ны пайдалану кезінде деректердің стационарлығын болжау, модельдің оңтайлы параметрлерін анықтау

қажеттілігі және коэффициенттерді түсіндіру мәселелері сияқты әдістердің шектеулерін ескеру қажет.

ARIMA және SARIMA үлгілері уақыттық қатар деректеріне енгізілген үлгілерді, трендтерді және циклдік ырғақтарды ашып, өткен мен болашақ арасындағы көпір ретінде қызмет етеді.

Тұтастай алғанда, ARIMA және SARIMA әдістері уақыт серияларын, соның ішінде компания өнімдеріне сұранысты болжауға арналған қуатты құралдар болып табылады және деректердегі үлгілерді анықтау және модельдеу үшін сәтті қолданылуы мүмкін.

ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

- 1 Орлов А.И. Прикладная статистика, 2006.
- 2 Савинская Д. Н., Кочкарова П.А. , Зейн В., Шуняев А.,А.Современные методы прогнозирования временных рядов.
- 3 Новиков А.А., Будзинская Е.О., Канаева О.Н. Анализ методов прогнозирования временных рядов: қол жеткізу режимі: file:///C:/Users/Dinara/Downloads/elibrary_44063546_89476144.pdf .
- 4 Осин А.А.,Фомин А.К.,Сологуб Г.Б.,Виноградов В.И. Моделирование и анализ данных ,2020. Т.10,№ 4. -С.41–50.
- 5 Канторович Г.Г., Лекционные и методические материалы. Анализ временных рядов.
- 6 How to Decompose Time Series Data into Trend and Seasonality [Электронды ресурс].–Қол жеткізу режимі: <https://machinelearningmastery.com/decomposetime-series-data-trend-seasonality/>.
- 7 Store Item Demand Forecasting Challenge | Kaggle [Электронды ресурс]. –Қол жеткізу режимі: <https://www.kaggle.com/c/demand-forecasting-kernelonly/overview>.
- 8 Введение в модель SARIMA: <https://medium.com/@ritusantra/introduction-to-sarima-model-cbb885ceabe8>.
- 9 Arima vs Sarima model .Қол жеткізу режимі: <https://www.geeksforgeeks.org/arima-vs-sarima-model/> .

А Қосымша

```
import os
import pandas as pd
import numpy as np
import datetime
```

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_pacf
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_squared_error
import statsmodels.formula.api as smf
import statsmodels.tsa.api as smt
import statsmodels.api as sm
import scipy.stats as scs
```

```
import warnings
from warnings import simplefilter
simplefilter('ignore')
warnings.filterwarnings('ignore')
```

```
train = pd.read_csv('train.csv', parse_dates=['date'])
test = pd.read_csv('test.csv', parse_dates=['date'])
train.head()
```

```
%matplotlib inline
%config InlineBackend.figure_formats = 'retina'
```

```
import tensorflow as tf
tf.random.set_seed(1)
```

```
train.head()
```

```
test.head()
```

```
train.describe()
```

```
test.describe()
```

```
train.shape
```

```
test.shape
```

А қосымшасының жалғасы

```
train.isnull().sum()
```

```
train.info()
```

```
print('Minimum date from train set: %s' % train['date'].min().date())  
print('Maximum date from train set: %s' % train['date'].max().date())
```

```
lag_size = (test['date'].max().date() - train['date'].max().date())  
print('Maximum date from train set: %s', train['date'].max().date())  
print('Maximum date from test set: %s', test['date'].max().date())  
print("Forecast lag size", lag_size)
```

```
daily_sales = train.groupby('date', as_index=False)['sales'].sum()  
store_daily_sales = train.groupby(['store', 'date'], as_index=False)['sales'].sum()  
item_daily_sales = train.groupby(['item', 'date'], as_index=False)['sales'].sum()
```

```
daily_sales_sc = plt.scatter(x=daily_sales['date'], y=daily_sales['sales'])  
plt.xlabel('Date')  
plt.ylabel('Sales')  
plt.title('Daily sales')  
plt.show()
```

```
store_daily_sales_sc = []  
for store in store_daily_sales['store'].unique():  
    current_store_daily_sales = store_daily_sales[(store_daily_sales['store']==store)]  
    store_daily_sales_sc.append(plt.scatter(x=current_store_daily_sales['date'], y=current_store_daily_sales['sales']))  
  
plt.xlabel('Date')  
plt.ylabel('Sales')  
plt.title('Store daily sales')  
plt.show()
```

```
item_daily_sales_sc = []  
for item in item_daily_sales['item'].unique():  
    current_item_daily_sales = item_daily_sales[(item_daily_sales['item'] == item)]  
    item_daily_sales_sc.append(plt.scatter(x=current_item_daily_sales['date'], y=current_item_daily_sales['sales']))  
  
plt.xlabel('Date')  
plt.ylabel('Sales')  
plt.title('Store daily sales')  
plt.show()
```

А қосымшасының жалғасы

```
def expand_df(df):
    data = df.copy()

    data['day'] = data.date.dt.day
    data['month'] = data.date.dt.month
    data['year'] = data.date.dt.year
    data['dayofweek'] = data.date.dt.dayofweek
    return data
```

```
train = expand_df(train)
train.head()
```

```
store_daily_sales.head()
```

```
plt.plot(daily_sales['date'], daily_sales['sales'])
plt.title('Daily Sales')
plt.ylabel('Rs')
plt.xlabel('Date')
plt.xticks(rotation=90)
plt.show()
```

```
store_colors = {
    1: 'blue',
    2: 'red',
    3: 'green',
    4: 'purple',
    5: 'orange',
    6: 'pink',
    7: 'brown',
    8: 'cyan',
    9: 'magenta',
    10: 'lime'
}
```

```
fig, ax = plt.subplots()
for store, color in store_colors.items():
    store_data = store_daily_sales[store_daily_sales['store'] == store]
    ax.plot(store_data['date'], store_data['sales'], label=f'Store {store}', color=color)
```

```
plt.show()
```

```
item_sales = train.groupby(['item', 'date'], as_index=False)['sales'].sum()
day_sales = train.groupby(['day', 'date'], as_index=False)['sales'].sum()
month_sales = train.groupby(['month', 'date'], as_index=False)['sales'].sum()
year_sales = train.groupby(['year', 'date'], as_index=False)['sales'].sum()
day_of_week_sales = train.groupby(['dayofweek', 'date'], as_index=False)['sales'].sum()
```

```
item_sales.head()
```

А қосымшасының жалғасы

```
for item in unique_items:
    item_data = item_sales[item_sales['item'] == item]
    ax.plot(item_data['date'], item_data['sales'], label=f'Item {item}')
```

```
ax.set_xlabel('Date')
ax.set_ylabel('Sales')
ax.set_title('Sales vs. Date by Item')
```

```
Text(0.5, 1.0, 'Sales vs. Date by Item')
```

```
train_item1 = train[train['item']==1]
train_final = train_item1[train_item1['store']==1]
```

```
result = sm.tsa.seasonal_decompose(train_final['sales'], model='additive', period=365)
```

```
fig = plt.figure()
fig = result.plot()
fig.set_size_inches(14, 12)
```

```
plt.plot(first_diff)
plt.title("Log Difference of Sales")
plt.show()
```

Қосымша В

```
#ARIMA
def optimize_ARIMA(endog, order_list):

    results = []

    for order in order_list:
        try:
            model = SARIMAX(endog, order=order, simple_differencing=False).fit(dispatch=False)
        except:
            continue

        aic = model.aic
        results.append([order, model.aic])

    result_df = pd.DataFrame(results)
    result_df.columns = ['(p, d, q)', 'AIC']

    result_df = result_df.sort_values(by='AIC', ascending=True).reset_index(drop=True)

    return result_df
```

```
from itertools import product
```

```
ps = range(0, 8, 1)
d = 1
qs = range(0, 8, 1)
```

```
# Create a list with all possible combination of parameters
parameters = product(ps, qs)
parameters_list = list(parameters)
```

```
order_list = []
```

```
for each in parameters_list:
    each = list(each)
    each.insert(1, 1)
    each = tuple(each)
    order_list.append(each)
```

```
order_list[:10]
```

```
result_df = optimize_ARIMA(train_final.sales, order_list)
result_df
```

```
import statsmodels.api as sm
resDiff = sm.tsa.arma_order_select_ic(train_final.sales, max_ar=8, max_ma=8, ic='aic', trend='c')
print('ARMA(p,q) = ', resDiff['aic_min_order'], 'is the best.')
```

Қосымша С

```
arima_model = SARIMAX(train_final.sales, order=(6,1,7), simple_differencing=False)
res = arima_model.fit(dispatch=False)
print(res.summary())
```

```
res.plot_diagnostics();
```

```
start_index = 1730
end_index = 1826
train['forecast'] = res.predict(start = start_index, end = end_index, dynamic = True)
train[start_index:end_index][['sales', 'forecast']].plot(figsize=(12, 8))
```

```
ax.set(title='Forecast of Store 1 Item 1')
plt.show()
```

```
train_final['model'] = predict.predicted_mean
train_final.head(15)
```

```
#SARIMA
len(train_final.sales)
```

```
ax.set(title='Forecast of Store 1 Item 1')
plt.show()
```

```
results = []
tr_start, tr_end = '2015-01-01', '2017-09-30'
te_start, te_end = '2017-10-01', '2017-12-31'
for i in range(1,51):
    for s in range(1,11):
        buf = df[(df.item==i)&(df.store==s)].copy()
        target_exog = buf[te_start:].drop(['id', 'store', 'item', 'sales'], axis = 1) # exog for predict.

        tra = buf['sales'][tr_start:tr_end].dropna()
        tes = buf['sales'][te_start:te_end].dropna()
        exog_train = buf.drop(['id', 'store', 'item', 'sales'], axis = 1)[tr_start:tr_end].dropna()
        exog_train = exog_train.astype(int)

        mod = SARIMAX(tra, order=(6,1,7), seasonal_order=(0,0,0,0), exog = exog_train, freq='D',
                      enforce_stationarity=False, enforce_invertibility=False).fit()

        target_exog = target_exog.astype(int)
        pred = mod.get_prediction(tr_end, '2018-03-31', exog = target_exog)
        results.extend(pred.predicted_mean['2018-01-01':])
        print('item:', i, 'store:', s, 'Finished.')
```

```
sample['sales'] = results
sample.to_csv('submission.csv', index=False)
```

```
sample
```