

Research paper / Оригинальная статья
<https://doi.org/10.51176/1997-9967-2023-3-268-283>
MPHTI 06.39.27
JEL: C15, C52, C55, C81



Methodological Basis and Experience of Using Data Mining Methods in Trade

Diana T. Kaiyp^a, Makpal G. Zhartybayeva^a, Zhanar O. Oralbekova^{a*}

^aL.N. Gumilyov Eurasian national university, 2 K. Satpayev Str., Z19A0K6, Astana, Kazakhstan

For citation: Kaiyp, D.T., Zhartybayeva, M.G. & Oralbekova, Z.O. (2023). Methodological Basis and Experience of Using Data Mining methods in Trade. *Economics: the strategy and practice*, 18(3), 268-283, <https://doi.org/10.51176/1997-9967-2023-3-268-283>

ABSTRACT

The article explores data mining methods that allow us to get helpful information from the data. The possibility of using these methods in practice in the financial sector was considered. Since financial activity is closely related to our social life, the use of data mining methods plays an essential role in the analysis and forecasting of the financial market in the modern era of big data. However, due to differences in the experience of researchers in different disciplines, it is not easy to use data mining methods when analyzing financial data. Therefore, creating a methodological basis for the practical application of data mining methods in the analysis of financial data is an urgent issue. The purpose of this article is to create a methodological basis for using data mining methods for efficient trading. When processing product data, a priori methods and visualization methods were used, and their implementation in practice was described. As a result, scenarios of computer applications were created as a sample of the practical implementation of the algorithms of these methods. Building a quantitative trading strategy requires first statistical analysis of the information in the market and then testing the quantitative model on the collected data. This study developed a quantitative trading system based on data mining methods. The primary development tool used is the Jupyter web platform, and three cores have been developed: quantitative data selection, strategy testing on data, time series analysis, and visualization. The developed system supports modules for making simple trading decisions.

KEYWORDS: Economy, Trade, Strategy, Practice, Data Mining, Finance, Financial Sector, Kazakhstan

CONFLICT OF INTEREST: the authors declare that there is no conflict of interest.

FINANCIAL SUPPORT: the work was carried out within the framework of project AP09058557, funded by a grant from the MSHE of the Republic of Kazakhstan.

Article history:

Received 17 October 2022
Accepted 11 July 2023
Published 30 September 2023

*Corresponding author: Oralbekova Z.O. – PhD, L.N. Gumilyov Eurasian national university, 2 K. Satpayev Str., Z19A0K6, Astana, Kazakhstan, 87761205714, email: oralbekova@bk.ru

Data mining әдістерін саудада қолданудың әдіснамалық негізі мен тәжірибесі

Қайып Д.Т.^а, Жартыбаева М.Г.^а, Оралбекова Ж.О.^{а*}

^а Л.Н. Гумилев атындағы Еуразия ұлттық университеті, көш. Қ. Сәтбаев, 2, Z19A0K6, Астана, Қазақстан

Дәйексөз үшін: Қайып Д.Т., Жартыбаева М.Г., Оралбекова Ж.О. (2023). Data mining әдістерін саудада қолданудың әдіснамалық негізі мен тәжірибесі. Экономика: стратегия және практика, 18(3), 268-283, <https://doi.org/10.51176/1997-9967-2023-3-268-283>

ТҮЙІН

Мақалада data mining деректерден пайдалы ақпаратты алуға арналған үдеріс зерттелді. Бұл әдістерді қаржы секторында тәжірибеде қолдану мүмкіндігі қарастырылған. Қаржылық қызмет әлеуметтік өмірбен тығыз байланысты болғандықтан, қазіргі үлкен деректер дәуірінде қаржы нарығын талдау мен болжауда data mining әдістерін қолдану маңызды рөл атқарады. Дегенмен, әртүрлі пәндер бойынша зерттеушілердің тәжірибесінің айырмашылығына байланысты қаржылық деректерді талдауда data mining әдістерін қолдану оңай емес. Сондықтан қаржылық мәліметтерді талдауда data mining әдістерін тәжірибеде қолданудың әдістемелік базасын құру өзекті мәселе болып табылады. Бұл мақаланың мақсаты тиімді сауда үшін деректерді іздеу әдістерін пайдаланудың әдістемелік базасын құру болып табылады. Өнім туралы мәліметтерді өңдеу кезінде априорлық және визуализация әдістері қолданылды және олардың тәжірибеде орындалуы сипатталды. Нәтижесінде осы әдістердің алгоритмдерін іс жүзінде жүзеге асырудың мысалы ретінде компьютерлік қолданбалы сценарийлер жасалды. Сандық сауда стратегиясын құру алдымен нарықтағы ақпаратты статистикалық талдауды, содан кейін жиналған деректер бойынша сандық модельді сынауды талап етеді. Бұл зерттеуде data mining әдістеріне негізделген сандық сауда жүйесі әзірленді. Juruter веб-платформасы негізгі әзірлеу құралы ретінде пайдаланылды және 3 ядросы әзірленді: сандық деректерді іріктеу, деректер бойынша стратегияны тестілеу, уақыт серияларын талдау және визуализация. Әзірленген жүйе қарапайым сауда шешімдерін қабылдауға арналған модульдерді қолдайды.

ТҮЙІН СӨЗДЕР: экономика, сауда, стратегия, тәжірибе, data mining, қаржы, қаржы секторы, Қазақстан

МҮДДЕЛЕР ҚАҚТЫҒЫСЫ: авторлар мүдделер қақтығысының жоқтығын мәлімдейді.

ҚАРЖЫЛАНДЫРУ: жұмыс ҚР ҒЖБМ грантымен қаржыландырылған AP09058557 жобасы шеңберінде орындалды.

Мақала тарихы:

Редакцияға түсті 17 Қазан 2022

Жариялау туралы шешім қабылданды 11 Шілде 2023

Жарияланды 30 Қыркүйек 2023

* **Хат-хабаршы авторы:** Оралбекова Ж.О. – PhD, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, көш. Қ. Сәтбаев 2, Z19A0K6, Астана, Қазақстан, 87761205714, email: oralbekova@bk.ru

Методологическая основа и опыт применения data mining методов в торговле

Кайып Д.Т.^а, Жартыбаева М.Г.^а, Оралбекова Ж.О.^{а*}

^а Евразийский национальный университет Л.Н. Гумилева, ул. К. Сатпаева 2, Z19A0K6, Астана, Казахстан

Для цитирования: Кайып Д.Т., Жартыбаева М.Г., Оралбекова Ж.О. (2023). Методологическая основа и опыт применения data mining методов в торговле. Экономика: стратегия и практика, 18(3), 268-283, <https://doi.org/10.51176/1997-9967-2023-3-268-283>

АННОТАЦИЯ

В статье исследован процесс data mining для получения полезной информации из данных. Рассмотрена возможность использования данных методов на практике в финансовой сфере. Поскольку финансовая деятельность тесно связана с социальной жизнью, использование data mining методов играет важную роль в анализе и прогнозировании финансового рынка в современную эпоху больших данных. Однако из-за различий в опыте исследователей разных дисциплин непросто использовать data mining методы при анализе финансовых данных. Поэтому создание методологической базы для практического применения data mining методов при анализе финансовых данных является актуальным вопросом. Цель данной статьи — создать методологическую базу для использования data mining методов для эффективной торговли. При обработке данных о продукте использовались априорные методы и методы визуализации, а также описывалась их реализация на практике. В результате были созданы сценарии компьютерных приложений как образец практической реализации алгоритмов этих методов. Построение количественной торговой стратегии требует сначала статистического анализа информации на рынке, а затем тестирования количественной модели на собранных данных. В этом исследовании была разработана количественная торговая система, основанная на data mining методах. В качестве основного инструмента разработки используется веб-платформа Jupyter, и были разработаны 3 ядра: отбор количественных данных, тестирование стратегии на данных, анализ временных рядов и визуализация. Разработанная система поддерживает модули для принятия простых торговых решений.

КЛЮЧЕВЫЕ СЛОВА: экономика, торговля, стратегия, практика, data mining, финансы, финансовая сфера, Казахстан

КОНФЛИКТ ИНТЕРЕСОВ: авторы заявляют об отсутствии конфликта интересов

ФИНАНСИРОВАНИЕ: Работа выполнена в рамках проекта AP09058557, финансируемого за счет гранта МНВО РК

История статьи:

Получено 17 октября 2022

Принято 11 июля 2023

Опубликовано 30 сентября 2023

* **Корреспондирующий автор:** Оралбекова Ж.О. – PhD, Евразийский национальный университет им. Л.Н. Гумилева, ул. К. Сатпаева 2, Z19A0K6, Астана, Казахстан, 87761205714, email: oralbekova@bk.ru

Кіріспе

Data mining деп деректердің үлкен көлемінен білімді алу немесе оларды өңдеу әдісін атайды. Деректерді өңдеу – бұл күтпеген байланыстарды табу және деректер иесіне түсінікті және пайдалы жаңа тәсілдермен деректерді қорытындылау үшін бақылау деректерінің жиынын талдау. Бұл деректер көбінесе үлкен көлемді болып келеді (Hand et.al. 2001). Data mining көптеген пәндер мен ғылым салаларын қамтитын зерттеудің қарқынды дамып келе жатқан бағыттарының бірі. Бұл пәндер ретінде деректер қоры, таратылған жүйелер, жасанды интеллект, визуализация, статистика, жоғары өнімді есептеулер және параллелді есептеулерді атап өтуге болады. Адамзаттың әр түрлі салаларында, мысалы қаржылық қызметтер саласында, бизнесте, ғылымда, медицинада, академиялық ортада өңделмеген деректерді адам жинап, деректерді өңдейтін және деректерді білімге айналдыруға мүмкіндік беретін компьютерлік бағдарламаларды қолданады. Деректерді өңдеудің мақсаты деректерді білімге айналдыру болып табылады. Data mining - бұл мәселелерді шешу үшін деректер қорында бұрын сақталған деректерді зерттеу үдерісі. Үлкен деректер қоймаларындағы үлгілерді табу әдісі data mining деп аталады. Деректерді талдау түрлеріне деректерді визуализациялау, жиынтық статистика, корреляциялық талдау, К-орташа мәндер кластерлеу мен сүзгілеу алгоритмдерін және машиналық оқытуды қолдану арқылы модельдеу кіреді (Vercellis, 2015). Қаржылық қызмет адамның әлеуметтік өмірімен тығыз байланыста болуына байланысты, қазіргі үлкен деректердің дәуірі жағдайында қаржы нарығын талдау мен болжауда data mining әдістерін қолдану маңызды рөл атқарады. Дегенмен де, әр түрлі пән зерттеушілерінің тәжірибесінің айырмашылығына байланысты қаржылық деректерді талдау барысында деректерді зияткерлік өңдеу әдістерін пайдалану оңайға соқпайды. Сол себептен қаржылық деректерді талдауда data mining әдістерін тәжірибеде қолданудың әдіснамалық негізін құру - өзекті мәселелердің бірі болып табылады.

Бұл мақаланың мақсаты - сауданы тиімді жүргізу үшін data mining әдістерін қолданудың әдіснамалық негізін құру болып табылады.

Мақала сауда және қаржы саласындағы мамандарға, сондай-ақ data mining әдістерін әзірлеумен және оларды тәжірибеде қолданумен айналысатын зерттеушілерге пайдалы болуы мүмкін. Сауда-саттықта data mining әдістерін қолдану компанияларға сатылымдар

мен тұтынушылардың мінез-құлқындағы маңызды заңдылықтарды анықтауға, ең тиімді маркетингтік стратегияларды таңдауға көмектеседі, сондай-ақ тауарлы-материалдық құндылықтар мен бағаларды басқаруды жақсартуға септігін тигізеді. Бұл компанияларға өнімдерге қолайлы бағаларды анықтауға және тауарлы-материалдық құндылықтар деңгейін оңтайландыруға мүмкіндік береді. Сонымен қатар, деректерді өңдеу әдістері де алаяқтық әрекетті анықтауға және сауда аймағында қауіпсіздікті қамтамасыз етуге көмектеседі.

Әдеби шолу

Data mining - бұл үлкен деректер жиынындағы қызықты және бұрын белгісіз үлгілер мен қатынастарды автоматты түрде алу процесі. Саудада data mining әдістері тұтынушылардың мінез-құлқын талдау, өнім трендтерін анықтау, сатуды болжау және тауарлық-материалдық қорларды басқару үшін қолданылады. Бұл бағытта ғалымдар бірқатар тәжірибелер мен зерттеулер жүргізді.

Krishnamoorthy мен Karthikeyan, Sun еңбектерінде деректерді зияткерлік өңдеудің әдістері арқылы деректерде жасырылған кейбір пайдалы ақпаратты анықтау мақсатында k-орташа мәндер кластерлеу алгоритмін оңтайландырып, класстарға жіктеу негізі ретінде Мурковский және Марков қашықтықтары пайдаланылды (Krishnamoorthy & Karthikeyan, 2022; Sun, 2022). Сонымен бірге мемлекеттің экспорттау өнімдеріне қатысты деректерді талдауда кластерлеу әдісін қолданудың артықшылықтарын атап өтуге болады. Екі шетелдік елдер арасындағы экспорттау өнімдерінің деректерін өңдеу барысында k-орташа мәндер кластерлеу әдісі арқылы экспорттық өнімдерді жақсарту мақсатында қандай өнімдерді жақсарту керек екенін көру үшін data mining әдісін қолдану тәжірибеде жүзеге асырылған (Pandin & Fahrudin, 2019). Кәсіпорынның қаржылық стратегиясын іске асырудың тиімділігіне бизнес органың әсер ету ерекшеліктерін анықтауға және оны қаржылық жағдайға бейімдеу жүйесін құруға қатысты әдістемелік ережелер кешені А.Бородин, Г.Панаедова, С.Фрумина, А.Қаирбекұлы, Н.Щеголеватыхтың жұмысында сипатталған (Borodin et.al, 2021). Тиімді нарық гипотезасының жорамалдарына қатысты алаңдаушылық трейдерлердің әртүрлі топтарының мінез-құлқы қаржы нарығының эволюциясына қалай әсер етуі мүмкін екендігі зерттелген (Ma et.al., 2022). Олар 6 жыл ішінде 5164 жеке тұлға сауда-саттық қызметі деректеріне

талдау жасаған. Сондай-ақ әртүрлі сауда стратегияларын қолданатын екі агенттің негізінде нарықтың микроқұрылымын модельдеуде data mining әдістерін қолдануға болады (Li et.al., 2021). Технологиялық жетістіктер үлкен деректер жиынтығының кең таралғанын білдіреді. Мұндай деректер жиынтығы бізге болашақ оқиғалардың ықтимал нәтижесін модельдеу және болжау үшін бұрын-соңды болмаған мүмкіндіктер береді. Дегенмен, мұндай деректер жиынтығы өзімен бірге жаңа қиындықтарды да әкелуі мүмкін. Пайдалы болжамдар жасау үшін осы кемшіліктерді, сондай-ақ осы үлкен деректер жиынтығының мүмкіндіктерін білу өте маңызды. Hand әртүрлі салалардағы программалық қамтамасыздандыруларды пайдалану арқылы осы қиындықтардың кейбірін зерттеді (Hand, 2009). Білімді іздеу мен деректерді өңдеудің пәнаралық саласы үлкен деректер қажеттілігінен туындады, бұл деректерді өңдеуден жаңа білімді алу үшін дәстүрлі статистикалық тәсілдерден тыс жаңа аналитикалық әдістерді қажет ететіндігін Shu мен Ye атап өтті (Shu & Ye, 2022). (Liu et.al., 2022) жұмысының авторлары data mining әдістерінің көмегімен Қытайдағы кәсіпорындардың маркетингтік тиімділігін арттыруға және маркетингтің инвестициялары мен шығынын азайтуға болатынын көрсеткен.

Десай мен Рамагеридің «Бөлшек сектордағы деректерді өңдеудің рөлі» мақаласы бөлшек сауда саласындағы деректерді өңдеу әдістерінің маңыздылығына жан-жақты шолу жасайды. Мақалада деректердің үлкен көлемінің қолжетімділігі бөлшек саудагерлерге осы деректер жиынынан құнды түсініктерді алу үшін деректерді өңдеу әдістерін пайдалану қажеттілігі көрсетілген. Авторлар деректерді өңдеу тұжырымдамасын және оны бөлшек сауда секторында қолдануды енгізуден бастайды. Олар тұтынушыларды сегменттеу, нарық себетін талдау, тұтынушылардың шығынын талдау және алаяқтықты анықтау сияқты әртүрлі бөлшек сауда мәселелерін шешу үшін деректерді өңдеу әдістерін қалай қолдануға болатынын түсіндіреді. Олар бөлшек саудагерлерге деректер сапасына инвестиция салуы және осы қиындықтарды шешу үшін деректердің құпиялылық саясатын әзірлеуі керек деп ұсынады. Мақала бөлшек сауда секторындағы деректерді өңдеудің кірісті арттыру, шығындарды азайту және тұтынушылардың қанағаттануын жақсарту сияқты артықшылықтарын қорытындылаумен аяқталады. Авторлар сонымен қатар бөлшек

саудагерлер нарықта бәсекеге қабілетті болып қалу үшін деректерді өңдеудің жаңа әдістері мен технологияларын зерттеуді жалғастыру керек деп болжайды. Тұтастай алғанда, Десай мен Рамагеридің мақаласы бөлшек сауда саласындағы деректерді өңдеудің қосымшаларына пайдалы кіріспе береді. Авторлар әртүрлі бөлшек сауда мәселелерін шешу және бизнес өнімділігін жақсарту үшін деректерді өңдеу әдістерін қалай пайдалануға болатынын тәжірибелік мысалдармен ұсынды. Олар сондай-ақ бөлшек сауда секторында деректерді өңдеуді енгізуге қатысты қиындықтарды атап өтіп, осы мәселелерді шешу үшін шешімдер ұсынды (Desai & Ramageri, 2013).

Кашван мен Велудың «Кластерлеу және деректерді өңдеу әдістерін қолдану арқылы тұтынушыларды сегменттеу» мақаласы бөлшек сауда саласында тұтынушыларды сегменттеу үшін кластерлеу және деректерді іздеу әдістерін пайдалану туралы зерттеуді ұсынды. Мақалада ұсынылған зерттеу бөлшек сауда дүкенінен тұтынушылар транзакцияларының деректер жинағына бағытталған. Авторлар тұтынушыларды сатып алу тәртібіне қарай сегменттеу үшін k-орталары, анық емес c-орталары және иерархиялық кластерлеу сияқты кластерлеу әдістерін пайдаланды. Олар сондай-ақ әртүрлі тұтынушы сегменттерінің сатып алу әрекетіндегі үлгілерді анықтау үшін қауымдастық ережелерін өндіруді пайдаланды. Зерттеу нәтижелері кластерлеу және деректерді өңдеу әдістерін бөлшек сауда саласында тұтынушыларды сегменттеу үшін тиімді пайдалануға болатынын көрсетті. Авторлар осы әдістерді пайдалану бөлшек саудагерлерге сатып алу тәртібі мен демографиялық сипаттамалары әртүрлі тұтынушылар сегменттерін анықтауға қалай көмектесетінін анықтады (Kashwan & Velu, 2013).

Kashwan мен Mikalef еңбектері үлкен деректерді талдау мүмкіндіктері туралы әдебиеттерге жан-жақты шолуды ұсынды. Авторлар негізгі зерттеу тақырыптарын анықтап, осы саладағы болашақ зерттеулерге ұсыныстар берді. Авторлар 2012 және 2017 жылдар аралығында жарияланған зерттеулерге назар аудара отырып, үлкен деректерді талдау мүмкіндіктері туралы әдебиеттерге жүйелі шолу жасады. Шолу ақпараттық жүйелер саласындағы жетекші академиялық журналдардың 102 зерттеуін қамтиды. Авторлармен қарастырылған әдебиеттерден алты негізгі зерттеу тақырыбын атап өтуге болады: (1) үлкен деректерді талдау мүмкіндік-

терін тұжырымдау, (2) үлкен деректерді талдау мүмкіндіктерінің алдыңғыларын анықтау, (3) үлкен деректерді талдау мүмкіндіктерін өлшеу, (4) үлкен деректердің әсерін бағалау бизнес өнімділігі бойынша аналитикалық мүмкіндіктер, (5) үлкен деректерді талдау мүмкіндіктерін жеңілдететін немесе кедергі келтіретін ұйымдық факторларды зерттеу және (6) үлкен деректерді талдау мүмкіндіктерінің этикалық және әлеуметтік салдарын зерттеу. Мақалада сонымен қатар алты зерттеу тақырыбының әрқайсысы бойынша болашақ зерттеулерге ұсыныстар берілген. Мысалы, авторлар үлкен деректерді талдау мүмкіндіктерін концептуализациялау бойынша болашақ зерттеулер осы мүмкіндіктердің көп өлшемді сипатын қамтитын жан-жақты және біріктірілген құрылымды әзірлеуге бағытталуы керек деп ұсынды (Kashwan & Velu, 2013; Mikalef et.al., 2018).

Ван Нгуен және т.б. Еуропалық операциялық зерттеулер журналында «Қайта өндірілген өнімдерге тұтынушылардың сұранысын болжау: деректерді өңдеу тәсілі» атты зерттеуін ұсынды (van Nguyen et.al., 2020). Бұл мақала қайта өңделген өнімдерге тұтынушылардың сұранысын болжау үшін деректерді іздеу тәсілін әзірлеуге бағытталған, бұл қайта өндіруші фирмаларға өндіріс пен қорларды басқару стратегияларын жақсырақ жоспарлауға көмектеседі. Авторлар қайта өңдеу өнеркәсібіне шолу жасаудан және тұтынушы сұранысын дәл болжау маңыздылығын ұсынудан бастайды. Олар қайта өңдеуде қолданылған өнімді бастапқы күйіне келтіру және оларды жаңа өнім ретінде сату процесін қамтитынын түсіндіреді. Дегенмен, қайта өңделген өнімге сұраныс көбінесе белгісіз болып келеді және оған тұтынушылардың қалауы, өнім сапасы және баға белгілеу сияқты әртүрлі факторлар әсер етуі мүмкін. Содан кейін мақалада тұтынушылардың қайта өндірілген өнімдерге сұранысын болжау үшін деректерді іздеу тәсілі ұсынылған. Авторлар қытайлық қайта өңдеу фирмасының қайта өңделген өнімдері үшін тарихи сағылым деректерінің жинағын пайдаланды. Олар тұтынушы сұранысы үшін болжамды модельдерді әзірлеу үшін шешім ағаштарын, нейрондық желілерді және қолдау векторлық машиналарын қоса алғанда, деректерді өңдеудің бірнеше әдістерін қолданды. Тұтастай алғанда, мақала тұтынушылар сұранысын болжау үшін деректерді өңдеу әдістерінің әлеуетін көрсету арқылы қайта өңдеу өнеркәсібіне құнды үлес қосады. Зерттеу қайта өндіруші фирмалар

үшін сұранысты дәл болжау маңыздылығын көрсетеді және тұтынушылардың қайта өндірілген өнімге сұранысына әсер ететін факторлар туралы түсінік береді. Бұл мақаладағы ұсынылған деректерді іздеу тәсілі басқа қайта өңдеуші фирмалар үшін өндіріс пен қорларды басқару стратегияларын әзірлеуде пайдалы болуы мүмкін.

Кәсіпкерлік операцияларда уақыттың маңыздылығы да кеңінен мойындалған (Li et.al., 2019; Srividya et.al., 2017; Singh & Yassine, 2019). Дегенмен, тұтынушылардың деректерін зерттеу бойынша бар зерттеулер әлі де жан-жақты өлшемдерді бірінші орынға шығарған жоқ. Біздің мақалада тұтынушылардың толық бейнесін алу үшін үлкен деректерді өңдеу әдістері мен бұл әдістерді әзірлеу қажеттілігінен туындаған зерттеулер олқылығы анықталған.

Бұл әдеби шолу data mining әдістерін қаржылық деректерді өңдеуде қолданудың артықшылықтарын көрсетеді. Қорытындылай келе, әдістемелік база мен саудадағы деректерді өңдеу тәжірибесі компанияларға тұтынушылар, нарықтық үрдістер және ішкі процестер туралы құнды білім алуға мүмкіндік береді. Бұл маркетингтік стратегияларды жақсартуға, операцияларды оңтайландыруға және қауіпсіздікті арттыруға әкеледі. Дегенмен, деректер көлеміне, есептеу ресурстарына және деректерді қорғауға қатысты шектеулер мен мәселелер ескерілуі керек.

Әдістер

Априори алгоритмі ассоциация ережелерін шығару үшін қолданылатын негізгі алгоритм болып табылады. Оның екі қадамдық үдерісі бар:

1. Белгіленген ең аз жиілікте орын алатын транзакциялар жинағындағы элементтердің барлық комбинацияларын табу. Бұл комбинациялар жиі элементтер жиынтығы деп аталады.

2. Жиі элементтер жиынындағы элементтердің ықтимал бірлескен кездесуін білдіретін ережелерді құру. Априори алгоритмі жиі элементтер жиынында элементтің бар болуы ықтималдығын есептейді.

Априори алгоритмі жиі элементтер жиынындағы элементтер арасындағы ықтималдық қатынастарды білдіретін ассоциация ережелерін жасайды. Ассоциация ережесі «егер алдыңғы болса, содан кейін нәтиже» түрінде болады. Онда белгілі бір позиция немесе заттар тобы, яғни бір тауар себетінде басқа заттың болуын білдіреді, демек, белгілі бір ықтималдықпен іске асады.

Ассоциацияның кластерлеу мен ауытқуды анықтаудан айырмашылығы, деректер жиынындағы даналар (жолдар) арасындағы ұқсастықтарды немесе айырмашылықтарды анықтауға бағытталады. Ассоциация ережесін анықтау деректер жиынындағы атрибуттар (бағандар) арасындағы қатынастарды байқауға бағытталады. Жалпы мағынада, ол өнімдер арасындағы корреляцияларды іздейді. Искерлік ассоциация ережелерін пайдалана отырып, деректерде болуы мүмкін үлгілерді іздеу арқылы тұтынушыларының мінез-құлқы туралы сұрақтарға жауап бере алады. Маркетингтік науқан жұмысының нәтижесі, тұтынушының сатып алу үлгілерінің өзгеруіне әсері, өнімнің орналасқан жері сатып алу әрекетіне әсері, жаңа өнімнің мақсатты тұтынушыларына қатысты мәселелерге нарық-себет талдауын қолдануға болады.

Бұл мақалада ассоциация ережелері дүкендегі күнделікті тауарларды сатып алудың нақты мысалына қолданылуы сипатталды. Статистикадағы “график” термині екі мағынаға ие болуы мүмкін. Бұл, әсіресе танымал қолданыста, деректерді ұсыну үшін көптеген сызбалардың кез келгеніне қатысты болуы мүмкін. Мысалы, сызықтық диаграмма, гистограмма және т.с.с. Техникалық тұрғыдан алғанда, бұл деректер құрылымы мен желілердегі визуализацияға жатады. Априори алгоритмі – бұл элементтер комбинациясының ассоциациялық байланысын анықтауға мүмкіндік беретін ассоциация ережелері бар деректерді өңдеу алгоритмі (Han et.al, 2011). Априори алгоритмінде элементтер жиынтығын жиі іздеу ассоциация ережесінің техникасын қолдану арқылы орындалады. Бұл алгоритм қосымша ақпаратты табу үшін белгілі жиілік атрибуты білімін пайдаланады. Априори алгоритмінде ең аз қолдау мен минималды сенімділік ескеріле отырып, пайда болуы мүмкін үміткер анықталады. Қолдау элемент комбинациясының мәнін немесе пайызын білдіреді (Vercellis, 2015). Мұнда қолдау мәнінің формуласы келесідей:

$$supp(X) = \frac{|\{t \in T; X \in t\}|}{|T|},$$

мұндағы $supp(X)$ - А элементі үшін қолдау мәні.

Төменде 2 элементтің қолдау мәнінің формуласы берілген:

$$supp(x_1 \cup x_2) = \frac{\sigma(x_1 \cup x_2)}{|T|},$$

мұндағы $supp(x_1 \cup x_2)$ - А және В тармақтары үшін қолдаудың мәні.

Мысалы, әдетте ұн мен сүтті сатып алатын тұтынушылар бар, олар төменде көрсетілген:

$$support = 5\%, confidence = 65\%$$

Жоғарыда келтірілген мысалда 5% қолдау мәні ұн мен сүтті бір уақытта сатып алу бойынша жалпы тұтынушылық транзакциялардың жалпы сомасы 5% екенін көрсетеді деп түсіндіруге болады. 65 пайыздық сенімділік тұтынушылар ұн сатып алатын болса, міндетті түрде сүтті 65 пайызға сатып алатынын көрсетеді.

Нәтижелер мен талқылау

Априори алдын ала белгілі мәліметтермен жұмыс істейді. Оны сатуда қолдану үшін белгілі бір дүкеннің сатылым кестесін білу керек. Бұл кестені қолдана отырып, jupyter notebook сайтында python арқылы тәжірибе түрінде жүзеге асырылды. Jupyter notebook жүйесінің data mining-те қолданудың басты тиімділігі берілген тапсырмаларды реттілігін бұзбай, жылдам орындалуын атап өтуге болады (Nielsen, 2015; Nussbaumer, 2015; Aggarwal & Han, 2014).

Ыңғайлы болу үшін ең алдымен тауарлар деректері төмендегі кестеде бейнеленгендей MS Excel бағдарламасында құрастырылды.

Бұл мәліметтер жиынына сатып алынған тауардың счет-фактурасы кіреді. Тауардың сипаттамасы, атауы, сатып алынған күн счет ұсыну күні деп беріліп тұр. Тауардың бағасы мен клиент ID және тауардың қай елде өндірілгені берілген. Ең соңғы бағандарда тауардың келген күні мен жарамдылық мерзімі көрсетілген.

Көрсетілген 1-кестеде счет-фактура бағаны бойында Е арқылы берілген счет-фактуралар тізбегі бар. Е арқылы берілген счет-фактуралар бізде жарамдылығы өткен тауарды білдіреді.

Кесте 1 - Құрылған кесте
Table 1 - Created table

Счет-фактура	Сипаттамасы	Саны	Счет ұсыну күні	Өлшем бірлігінің бағасы	Клиент ID	Ел	Келген күні	Жарамдылық мерзімі өтетін күні	Категория
540113	Цесна ұны 2кг	74	05.01.2022 9:11	755	12961	Қазақстан	23.08.2021	23.02.2022	ұн өнімдері
E540113	Сүт Lactel 2,5%	48	05.01.2022 9:11	430	11659	Қазақстан	23.08.2021	07.09.2021	сүт өнімдері
E540113	Сары май 82%	27	05.01.2022 9:11	770	13660	Қазақстан	02.10.2021	02.11.2021	сүт өнімдері
540113	Сұлтан макарондары слагетти 1,6кг	22	05.01.2022 9:11	1640	16109	Қазақстан	18.09.2021	21.03.2022	ұн өнімдері
540113	Йогурт Растишка алма-алмұрт	61	05.01.2022 9:11	610	11697	Қазақстан	20.12.2021	22.06.2022	сүт өнімдері
540113	Биг Ланч кеспесі тауық	4	05.01.2022 9:11	410	13517	Қазақстан	27.11.2021	30.05.2022	ұн өнімдері
540113	Lays Чипсы	14	05.01.2022 9:11	464	13254	Ресей	04.01.2022	07.07.2022	снэк өнімдері
540113	Бижан шұжығы мұсылмандық	53	05.01.2022 9:11	2530	12759	Қазақстан	01.01.2022	04.07.2022	ет өнімдері
540114	Балмұздақ Oreo 263 г	26	05.01.2022 9:44	3170	12679	Қазақстан	14.10.2021	16.04.2022	сүт өнімдері
540114	Йогурт Эрмигурт құлпынай-банан	33	05.01.2022 9:44	455	13080	Қазақстан	08.12.2021	10.06.2022	сүт өнімдері
540114	Boszhan сиыр коллеттері бур-герге арналған 600 г	25	05.01.2022 9:44	1815	16047	Қазақстан	21.10.2021	23.04.2022	ет өнімдері
540115	Батон Nap Aletmi	18	05.01.2022 9:51	190	14945	Ресей	01.08.2021	01.02.2022	ұн өнімдері
540115	Алма Алорт	68	05.01.2022 9:51	765	12422	Қазақстан	18.12.2021	20.05.2022	жеміс-көкөніс өнімдері

540120	Жұмыртқа Казгер-Кус 10 дана	61	05.01.2022 10:08	620	15830	Қазақстан	26.12.2021	26.01.2022	жұмыртқа өнімдері
540120	Қаймақ President 20% 400 г	79	05.01.2022 10:08	680	11699	Қазақстан	03.09.2021	06.03.2022	сүт өнімдері
540120	Қаймақты сүт Простоквашино 10% 350 г	96	05.01.2022 10:08	540	12593	Ресей	06.10.2021	08.04.2022	сүт өнімдері
540120	Нан	36	05.01.2022 10:08	140	11515	Қазақстан	12.09.2021	15.03.2022	үн өнімдері
....
540257	Сүт Lactel 2,5%	68	10.01.2022 17:33	495	15691	Қазақстан	03.09.2021	06.03.2022	сүт өнімдері
540257	Шырын Сады Придонья алма	66	10.01.2022 17:33	635	15025	Қазақстан	12.09.2021	15.03.2022	сусындар

Ескерту - автормен құрастырылған

Құрылған кестені пайдалана отырып, Python тілінде априори мен визуализация жасалды. Түсінікті болу үшін априори мен визуализацияның жасалу жолын толығымен сипаттап берейік.

1. Ең бірінші қадам болып mlxtend кітапханасы `pip install mlxtend` командасы арқылы орнатылды.

Бұл кітапхананы қолданудың себебі жасанды зерде түрі машиналық оқытуға арналған біршама көмекші функциялар бар. Жинақтау және дауыс беру жіктеуіштері, модельді бағалау, сипаттамаларды алу, сонымен қатар графиканы әзірлеу және құру сияқты мүмкіндіктерді қамтиды. Құжаттамадан басқа, бұл қағаз пакетті егжей-тегжейлі түсіну үшін жақсы ресурс болып табылады.

2. Numpy, pandas, Априори, association rules пакеттері жүктелді:

```
import numpy as np
import pandas as pd
from mlxtend.frequent_patterns import
apriori
```

```
from mlxtend.frequent_patterns import
association_rules
```

3. Деректер базадан - Excel-ден оқылды, ол үшін келесі листинг пайдаланылды:

```
#Деректерді тазарту
dataB['Сипаттамасы'] = dataB['Сипатта-
масы'].str.strip() #бос орындарды басынан және
соңынан жояды
dataB.dropna(axis=0, subset=['Счет-факту-
ра'], inplace=True) #қайталанатын счет-факту-
раны жояды
dataB['Счет-фактура'] = dataB['Счет-фак-
тура'].astype('str') #счет-фактураны түрлендіру
dataB = dataB[~dataB['Счет-фактура'].str.
contains(' E')] #E-дегеніміз жарамдылық, өт-
кен тауардың счет-фактурасы
dataB.head()
```

4. Келесі қадам априори тәсіліне деректерді қолдану үшін дайындадық. Бұл алдын-ала дайындау жұмыстары деректер ішіндегі бос орындарды алып тастаудан, қайталанатын счет-фактураларды жоюдан, несиелік транзакцияларды жоюдан тұрады (2-кесте).

Кесте 2 - Деректерді алдын-ала өңдеу нәтижесі

Table 2 – Result of data preprocessing

үн өнімдері	232
сүт өнімдері	222
сусындар	196
жеміс- көкөніс өнімдері	174
снэк өнімдері	104
ет өнімдері	72
жұмыртқа өнімдері	34

Ескерту – Автормен Jupyter ортасында Python программалау тілінде жазылған программаның көмегімен құрастырылды.

5. Деректерді санагтар бойынша бөлуді жүзеге асырылуы 3-кестеде келтірілген.

Кесте 3 - Деректерде жарамдылық өткен тауарларды алып тастау

Table 3 - Exclusion of expired goods in the data

Сипаттамасы Ел	Йогурт Растишка алма- алмұрт	Йогурт Эрмигурт құлпынай-банан	Айран Фуд Мастер 2,5%	Қымыз Алтай	Кұрт Шоро классикалық	Маргарин 3 Желания Пампушка	Петропавл майы сыыр 82,5%	Сары май 82%	Сүт Lactel 2.5%	Балмұздақ Oreo 263 г	Бурёнка қойылтылған сүті қантпен Масло- Дел 8,5%
Ресей	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Ескерту – Автормен Jupyter ортасында Python программалау тілінде жазылған программаның көмегімен құрастырылды.

6. Деректер көзінде қанша мемлекет барын анықтау 4-кестеде сипатталғандай жүзеге асырылды.

Кесте 4 - Деректерде мемлекеттер санын анықтау
Table 4 - Determine the number of states in the data

Қазақстан	1198
Ресей	133
Өзбекстан	22

Ескерту – Автормен Jupyter ортасында Python программалау тілінде жазылған программаның көмегімен құрастырылды.

7. Белгілі бір ел үшін транзакцияларды бақылау мүмкіндігі қарастырылған. Мысал ретінде кодта көрсетілгендей Ресей мемлекеті алынды. Нәтижесі 5-кестеде келтірілген.

Ресей бойынша транзакцияларды бөлу
`myshop = (dataB[dataB['Ел'] ==»Россия»].groupby(['Счет-фактура', Сипаттамасы'])['Саны'] .sum().unstack().reset_index().fillna(0) .set_index('Счет-фактура'))`

Кесте 5 - Ресей бойынша транзакцияларды көру нәтижесі

Table 5 - View transactions

Сипаттамасы Счет-фактура	Батон Nap Alemi туралған	Пияз	Қияр Миринда	Қаймақ Просток- вашино 10% 350г	Қаймақ Просток- вашино 15%	Ламбер сыры 55%	Greenfield Jasmine Dream көк шайы	Lays картоп қытыр- лақтары
510113	0.0	0.0	0.0	0.0	0.0	0.0	0.0	77.0
510115	35.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
510120	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
510146	0.0	0.0	0.0	0.0	0.0	0.0	99.0	0.0
510169	35.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Ескерту – Автормен Jupyter ортасында Python программалау тілінде жазылған программаның көмегімен құрастырылды.

8. Ыңғайлы болу үшін барлық оң мән-дерді 1-ге, ал қалғандарын 0-ге түрлендірдік. Нәтижесі 6-кестеде көрсетілген:

#барлық оң мән-дерді 1-ге, ал қалғандарын 0-ге түрлендеру

```
def my_encode(x): if x <= 0: return 0 if x
>= 1: return 1
sets= myshop.applymap(my_encode) sets.
head()
```

Кесте 6 - Бинарлы тәсілді қолдану

Table 6 - Using a binary approach

Сипаттамасы Счет- фактура	Батон Nap Alemi туралған	Қияр Миринда	Қаймақ Просток- вашино 10% 350г	Қаймақ Просток- вашино 15%	Greenfield Jasmine Dream көк шайы	Lays картоп қытырлақ- тары
510113	0	0	0	0	0	1
510115	1	0	0	0	0	0
510120	0	0	1	0	0	0
510146	0	0	0	0	1	0
510169	1	0	0	0	0	0

Ескерту – Автормен Jupyter ортасында Python программалау тілінде жазылған программаның көмегімен құрастырылды.

9. Априори арқылы 7-кестеде көрсетілгендей элементтердің жиілік жиынтығы құрылды:

Кесте 7 - Тауарлардың жиілік жиынтығын құру
Table 7 - Creating a frequency set of goods

№	antecedents	consequents	Antecedents support	Consequents support	support	confidence	lift	leverage	conviction
0	(Қияр Миринда)	(Lays картоп қытырлақтары)	0.263158	0.192982	0.122807	0.466667	2.418182	0.072022	1.513158
1	(Lays картоп қытырлақтары)	(Қияр Миринда)	0.192982	0.263158	0.122807	0.636364	2.418182	0.072022	2.02616
2	(Батон Nan Alemi туралған, Қияр Миринда)	(Lays картоп қытырлақтары)	0.043860	0.192982	0.043860	1.000000	5.181818	0.035396	inf
3	(Батон Nan Alemi туралған, Lays картоп қытырлақтары)	(Қияр Миринда)	0.043860	0.263158	0.043860	1.000000	3.800000	0.032318	inf
4	(Қияр Миринда)	(Батон Nan Alemi туралған, Lays картоп қытырлақтары)	0.263158	0.043860	0.043860	0.166667	3.800000	0.032318	1.147368

Ескерту – Автормен Jupyter ортасында Python программалау тілінде жазылған программаның көмегімен құрастырылды.

10. Ұсынымдарды әзірлеу нәтижесі 8-кестеде көрсетілген.

Кесте 8 - Ұсынымдарды даярлау көрінісі
Table 8 - Proposal preparation view

№	antecedents	consequents	Antecedents support	Consequents support	support	confidence	lift	leverage	conviction
1	(Батон Nan Alemi туралған, Қияр Миринда)	(Lays картоп қытырлақтары)	0.04386	0.192982	0.043860	1.0	5.181818	0.035396	inf
2	(Батон Nan Alemi туралған, Lays картоп қытырлақтары)	(Қияр Миринда)	0.04386	0.263158	0.043860	1.0	3.800000	0.032318	inf

Ескерту – Автормен Jupyter ортасында Python программалау тілінде жазылған программаның көмегімен құрастырылды.

Танымал тауарлардың жиілігі арқылы қай өнімнің сұранысқа ие екенін, қандай өнімдер үшін сауда көлемін арттыруға болатындығын анықтауға болады. Және танымалдығы аз тауарлар санын аз көлемде алуға немесе сол тауарларды қандай бағытта ұсыныс жасау арқылы сауда көлемін арттыруға септігін тигізеді.

2-суретте бейнеленгендей ең танымал тауарлар қатарына қарапайым нан мен туралған

нандар жатады. Ал танымалдығы аз өнім ретінде цесна спагетти мен киви жемісін атап өтуге болады.

Жасалған жұмысымыздан байқай алатынымыз, біз ең алдымен деректерді жинау және өңдеу арқылы сауда көлемін арттыруға жұмыс жасадық (3-сурет).



Сурет 3 - Жасалған жұмыстың диаграммасы

Figure 3 - Diagram of the work done

Ескерту - автормен құрастырылған

Қорытынды

Деректерді өңдеу міндетін орындау үшін әрқашан тиімді зерттеу стратегиялары қажет, ол қолданылатын ресурстар көлемі мен зерттеу міндеттерімен анықталады. Ассоциация ережелері, априори және визуализация, кластерлеу - data mining әдістерінің бірнеше мысалдары ғана. Бұл зерттеудің мақсатына, мақалада көрсетілгендей, ассоциация ережелерін талдау, априори және визуализация әдістерін қолдану арқылы қол жеткізілді. Мақалада data mining әдістері, соның ішінде априори және визуализация әдістерін тауар өнімдері туралы шынайы деректерді өңдеуде қолдану арқылы пайдалы жаңа ақпарат алуға болатыны көрсетілді, дүкендегі транзакция деректерінің мәселесін зерттеу барысында ассоциация ережелері тұтынушылардың сатып алу әдеттерін түсіну үшін пайдаланылды. Бұл әдістердің алгоритмдерін тәжірибеде жүзеге асырылу мысалында компьютерлік қолданба-

лы сценарийлер ұсынылды. Сол сценарийлер негізінде мәселен тауар сатушылар мен тауар сатып алушылар бизнес жүргізу стратегиясын құра алады.

Әдебиеттерді талдау және шолу негізінде біз болашақта осы әдістерді саудада дамыту және қолдану бойынша бірнеше ұсыныстар жасай аламыз:

1. Деректерді визуализациялауды жақсарту: Деректерді визуализациялаудың заманауи технологиялары сауда деректерін көбірек көрнекі және интерактивті графикалық ұсынуға мүмкіндік береді. Бұл трейдерлер мен талдаушыларға ақпаратты жақсы түсінуге, жасырын үлгілерді ашуға және көрнекі талдау негізінде дұрыс шешімдер қабылдауға көмектеседі.

2. Деректерді біріктіру әдістерін пайдалану: Деректерді біріктіру әдістерін пайдалану сауда көрсеткіштері арасындағы жасырын байланыстар мен тәуелділіктерді

анықтауға мүмкіндік береді. Болашақта бұл әдістерді жетілдіру және оларды басқа аналитикалық құралдармен біріктіру ұсынылады.

3. Машиналық оқытуды және жасанды интеллектті дамыту: машиналық оқыту мен жасанды интеллект технологияларының дамуымен деректерді іздеу әдістері саудада одан да күшті құрал бола алады. Машиналық оқыту алгоритмдерін пайдалану деректерді талдау, үлгілерді анықтау және шешімдер қабылдау процестерін автоматтандырады. Бұл болжау дәлдігін жақсартуға, сауда стратегияларын оңтайландыруға және операциялардың табыстылығын арттыруға көмектеседі.

4. Үлкен деректерді біріктіру: Заманауи саудада деректер көлемі күн сайын артып келеді. Data mining әдістерін пайдалану үлкен көлемдегі деректерден құнды ақпаратты тиімді талдауға және шығаруға мүмкіндік береді. Дегенмен, болашақта үлкен деректердің әлеуетін толық пайдалану үшін оны біріктіру мен өңдеуге көбірек көңіл бөлу қажет. Үлкен деректермен жұмыс істеуге арналған технологиялар мен құралдарды енгізу саудада талдау мен шешім қабылдау сапасын арттыруға көмектеседі. Алайда бұл көп есептеу ресурстарын қажет ететіндігін ескерген жөн. Сондықтан қол жетімді бұлтты технологияларды, есептеуді оңтайландыру құралдарын, мысалы Браузердегі Google Python кодымен жұмыс істеуге арналған GoogleCollab тегін интерактивті бұлттық ортасын пайдалануға болады.

5. Болжамды аналитиканы дамыту: Сауда-саттықта data mining әдістерін пайдалану болжамдарды құруға және болашақ трендтер мен оқиғаларды болжауға мүмкіндік береді. Дегенмен, болашақта болжамдық аналитика саласындағы зерттеулерді жалғастырып, дәлірек және сенімді болжау үшін жаңа әдістер мен модельдерді әзірлеу қажет. Бұл трейдерлер мен инвесторларға деректер негізінде неғұрлым негізделген шешімдер қабылдауға және қаржылық көрсеткіштерін жақсартуға мүмкіндік береді.

Жалпы, сауда-саттықта data mining әдістерін пайдалану операциялардың тиімділігі мен табыстылығын арттыру үшін үлкен әлеуетке ие. Жұмыс ҚР ҒЖБМ грантымен қаржыландырылған АР09058557 жобасы шеңберінде орындалды.

References

1. Aggarwal, C. C., & Han, J. (2014). *Frequent Pattern Mining*. International Publishing Switzerland. Springer.
2. Ahmed, S. R. (2004). Applications of data mining in retail business. In *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004*. (Vol. 2, pp. 455-459). IEEE. <https://doi.org/10.1109/ITCC.2004.1286695>
3. Borodin, A., Panaedova, G., Frumina, S., Kairbekuly, A., & Shchegolevatykh, N. (2021). Modeling the business environment of an energy holding in the formation of a financial strategy. *Energies*, 14(23), 8107-8125. <https://doi.org/10.3390/en14238107>
4. Desai, B., & Ramageri, B. (2013). Role of data mining in retail sector. *International Journal on Computer Science and Engineering*, 5(1), 47-50.
5. Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Third edition, Elsevier.
6. Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. The MIT Press, Massachusetts Institute of Technology.
7. Hand, D. J. (2009). Mining the past to determine the future: Problems and possibilities. *International Journal of Forecasting*, 25(3), 441-451. <https://doi.org/10.1016/j.ijforecast.2008.09.004>
8. Kashwan, K. R., & Velu, C. M. (2013). Customer Segmentation Using Clustering and Data Mining Techniques. *International Journal of Computer Theory and Engineering*, 5(6), 856-861. <https://doi.org/10.7763/IJCTE.2013.V5.811>
9. Krishnamoorthy, M., & Karthikeyan, R. (2022). Pattern mining algorithms for data streams using itemset. *Measurement: Sensors*, 24, 100421-100430. <https://doi.org/10.1016/j.measen.2022.100421>
10. Li, B., Ma, Z., & Wang, T. (2021). Find trade patterns in China's stock markets using data mining agents. *Studies in Computational Intelligence*, 897, 171-179. https://doi.org/10.1007/978-3-030-49728-6_11
11. Li, Q., Li, S., Zhang, S., Hu, J., & Hu, J. (2019). A Review of Text Corpus-Based Tourism Big Data Mining. *Applied Sciences*, 9(16), 3300. <https://doi.org/10.3390/app9163300>
12. Liu, Y., Cao, J., & Zhang, Q. (2022). The product marketing model of the economic zone by the sensor big data mining algorithm. *Sustainable Computing: Informatics and Systems*, 36, 100820. <https://doi.org/10.1016/j.suscom.2022.100820>
13. Ma, T., Fraser-Mackenzie, P.A.F., Sung, M., Kansara, A.P., & Johnson J.E.V. (2022). Are the least successful traders those most likely to exit the market? A survival analysis contribution to the efficient market debate. *European Journal of Operational Research*, 299 (1), 330-345. <https://doi.org/10.1016/j.ejor.2021.08.050>
14. Mikalef, P., Pappas, I.O., Krogstie, J., & Giannakos, M. (2018). Big Data Analytics Capabilities:

A Systematic Literature Review and Research Agenda. *Information Systems and e-Business Management*, 16, 547–578. <https://doi.org/10.1007/s10257-017-0362-y>

15. Nielsen, F. A. (2015). *Data Mining with Python (Working draft)*. [cited March 30, 2023]. Available at: <https://www.webpages.uidaho.edu/~stevel/504/Data%20Mining%20with%20Python.pdf>

16. Nussbaumer, K. C. (2015). *Storytelling with data: a data visualization guide for business professionals*. Wiley Publishing inc.

17. Pandin, C. R. M., & Fahrudin, R. (2019). Application of data mining for Indonesian products export in South Korea using clustering: Indonesia Trade Promotion Center Busan. *IOP Conference Series: Materials Science and Engineering*, 662(2), 022053. <https://doi.org/10.1088/1757-899X/662/2/022053>

18. Shu, X., & Ye, Y. (2022). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. <https://doi.org/10.1016/j.ssresearch.2022.102817>

19. Srividya, K., Sowjanya, A. M., & Kumar, T. A. (2017). Sentiment Analysis of Facebook Data Using Naïve Bayes Classifier. *International Journal of Computer Science and Information Security*, 15 (1), 179–186.

20. Singh, S., & Yassine, A. (2018). Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting. *Energies*, 11(2), 452–478. <https://doi.org/10.3390/en11020452>

21. Sun, L. (2022). Research on Mining Balanced Competition Strategy in Financial Market Based on Computer Data Mining Method. *Mobile Information Systems*, 6202890. <https://doi.org/10.1155/2022/6202890>

22. Vercellis, C. (2015). *Business intelligence: data mining and optimization for decision making*. Wiley Publishing inc.

23. van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2020). Predicting customer demand for remanufactured products: A data-mining approach. *European Journal of Operational Research*, 281(3), 543–558. <https://doi.org/10.1016/j.ejor.2019.08.015>

24. Wang, F., Li, M., Mei, Y., & Li, W. (2020). Time Series Data Mining: A Case Study with Big Data Analytics Approach. *IEEE Access*, 8, 14322–14328. <https://doi.org/10.1109/ACCESS.2020.2966553>

Information about the authors

Diana T. Kaiyp – L.N. Gumilyov Eurasian national university, Astana, Kazakhstan, email: kayp.di@bk.ru

Макпал G. Zhartybayeva – Ph.D., Gumilyov Eurasian national university, Astana, Kazakhstan, email: makkenskii@mail.ru, ORCID ID: <https://orcid.org/0000-0002-3987-9353>

***Zhanar O. Oralbekova** – Ph.D., L.N. Gumilyov Eurasian national university, Astana, Kazakhstan, email: oralbekova@bk.ru, ORCID ID: <https://orcid.org/0000-0002-3726-8882>

Авторлар туралы мәліметтер

Қайып Д.Т. – Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан, email: kayp.di@bk.ru

Жартыбаева М.Г. – PhD, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан, email: makkenskii@mail.ru, ORCID ID: <https://orcid.org/0000-0002-3987-9353>

***Оралбекова Ж.О.** – PhD, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан, email: oralbekova@bk.ru, ORCID ID: <https://orcid.org/0000-0002-3726-8882>

Сведения об авторах

Кайып Д.Т. – Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, email: kayp.di@bk.ru

Жартыбаева М.Г. – PhD, Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, email: makkenskii@mail.ru, ORCID ID: <https://orcid.org/0000-0002-3987-9353>

***Оралбекова Ж.О.** – PhD, Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, email: oralbekova@bk.ru, ORCID ID: <https://orcid.org/0000-0002-3726-8882>