

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN  
ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ  
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН  
KAZAKHSTAN



**INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION  
TECHNOLOGIES**

Published since 2020.  
Volume 7. 2 (26). 2026  
April–June

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ  
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

2020 жылдан бері шығарылады  
Том 7. 2 (26). 2026  
Сәуір-Маусым

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И  
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Издается с 2020 г.  
Том 7. 2 (26). 2026  
Апрель-Июнь

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Зарегистрировано в Международном центре регистрации серийных изданий ISSN (ЮНЕСКО, Париж, Франция). ISSN 2708–2032 (print), ISSN 2708–2040 (online)

Журнал входит в Перечень научных изданий, рекомендуемых КОКНВО МНВО РК для публикации основных результатов научной деятельности.

#### EDITOR-IN-CHIEF:

**Kateryna Kolesnikova** — Doctor of Technical Sciences, professor, Vice-Rector for Research, International Information Technology University (Kazakhstan)

#### DEPUTY EDITOR-IN-CHIEF:

**Madina Ipalakova** — Candidate of Technical Sciences, associate professor, Director of the Research Department, International Information Technology University (Kazakhstan)

#### EDITORIAL BOARD:

**Abdul Razak** — PhD, professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

**Lucio Tommaso De Paolis** — Director of the R&D Department of the AVR Laboratory, Department of Engineering for Innovation, University of Salento (Italy)

**Liz Bacon** — Professor, Deputy Vice-Chancellor, Abertay University (United Kingdom)

**Michele Pagano** — PhD, Professor, University of Pisa (Italy)

**Mukhtarbay Otelbayev** — Doctor of Physical and Mathematical Sciences, professor, academician of the National Academy of Sciences of the Republic of Kazakhstan, professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

**Bolatbek Rysbauly** — Doctor of Physical and Mathematical Sciences, professor, professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

**Yevgeniya Daineko** — PhD, research professor, Department of Information Systems, International Information Technology University (Kazakhstan)

**Nurzhan Duzbayev** — PhD, associate professor, Vice-Rector for Digitalization and Innovation, International Information Technology University (Kazakhstan)

**Bakhtgerai Sinchev** — Doctor of Technical Sciences, professor, Department of Information Systems, International Information Technology University (Kazakhstan)

**Nurgul Seilova** — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)

**Ardak Mukhamediyeva** — Candidate of Economic Sciences, Dean of the Faculty of Business, Media and Management, International Information Technology University (Kazakhstan)

**Zamira Abdikalikova** — PhD, associate professor, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

**Yerlan Shildibekov** — PhD, associate professor, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)

**Damilya Yeskendirowa** — Candidate of Technical Sciences, associate professor, Head of the Department of Cybersecurity, International Information Technology University (Kazakhstan)

**Aigul Niyazgulova** — Candidate of Philological Sciences, Professor, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)

**Altai Aitmagambetov** — Candidate of Technical Sciences, Professor, Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

**Yelena Bakhtiyarova** — Candidate of Technical Sciences, associate professor, Head of the Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

**Kanibek Sansyzbay** — PhD, research professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

**Sakhybay Tynymbayev** — Candidate of Technical Sciences, Professor, Research Professor, Department of Computer Engineering, International Information Technology University (Kazakhstan)

**Ali Abd Almisreb** — PhD, associate professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

**Mohamed Ahmed Hamada** — PhD, associate professor, Department of Information Systems, International Information Technology University (Kazakhstan)

**Yang Im Chu** — PhD, Professor, Gachon University (South Korea)

**Tadeusz Wallas** — PhD, Vice-Rector, Adam Mickiewicz University (Poland)

**Orken Mamyrbayev** — PhD, Deputy Director for Science, RSE Institute of Information and Computational Technologies, Committee for Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Kazakhstan)

**Sergey Bushuyev** — Doctor of Technical Sciences, professor, Director of the Ukrainian Project Management Association "UKRNET," Head of the Department of Project Management, Kyiv National University of Construction and Architecture (Ukraine)

**Svetlana Beloshitskaya** — Doctor of Technical Sciences, professor, Department of Computing and Data Science, Astana IT University (Kazakhstan)

#### MANAGING EDITOR

**Raushan Mrzabayeva** — Master of Science, editor, International Information Technology University (Kazakhstan)

---

International Journal of Information and Communication Technologies

Periodicity: 4 times a year.

Languages: Kazakh, Russian, English

DOI prefix: 10.54309

ISSN 2708-2032 (print)

ISSN 2708-2040 (online)

Thematic focus: "Information technology"; "Digital technologies in the development of socio-economic systems"; "Information security and communication technologies".

Distribution: Materials are distributed under the Creative Commons Attribution 4.0

Journal website: <https://journal.iitu.edu.kz>

Owner: International Information Technology University JSC (Almaty).

Copyright: © International Journal of Information and Communication Technologies, 2026

---

## РЕДАКЦИЯ

## БАС РЕДАКТОР:

**Колесникова Катерина Викторовна** — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі проректор (Қазақстан)

## БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

**Ипалакова Мадина Түлегеновна** — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі департамент директоры (Қазақстан)

## РЕДАКЦИЯЛЫҚ АЛҚА:

**Разак Абдул** — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессоры (Қазақстан)

**Лучино Томмазо де Паолис** — Саленто Университеті (Италия) инновация және технологиялық инжиниринг департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры

**Лиз Бэкон** — профессор, Абертей Университеті (Ұлыбритания) вице-канцлерінің орынбасары

**Микеле Пагано** — PhD, Пиза Университетінің (Италия) профессоры

**Өтелбаев Мұхтарбай Өтелбайұлы** — физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының профессоры (Қазақстан)

**Рысбайұлы Болатбек** — физика-математика ғылымдарының докторы, профессор, Есептеу және деректер ғылымдары департаментінің профессоры, Astana IT University (Қазақстан)

**Дайнеко Евгения Александровна** — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессор-зерттеушісі (Қазақстан)

**Дузаев Нуржан Тоқсуажасвич** — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті цифрландыру және инновациялар жөніндегі проректор (Қазақстан)

**Синчев Бахтгерей Кусанович** — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессоры (Қазақстан)

**Сейлова Нургуль Абадуллаевна** — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті компьютерлік технологиялар және киберқауіпсіздік факультетінің деканы (Қазақстан)

**Мухамедиева Ардак Габитовна** — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті бизнес медиа және басқару факультетінің деканы (Қазақстан)

**Абикаликова Замира Тұрсынбаевна** — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының меңгерушісі (Қазақстан)

**Шильдибеков Ерлан Жаржанович** — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті экономика және бизнес кафедрасының меңгерушісі (Қазақстан)

**Дамеля Максұтовна Ескендірова** — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының меңгерушісі (Қазақстан)

**Ниязгулова Айгуль Аскарбековна** — филология ғылымдарының кандидаты, доцент, профессор, Халықаралық ақпараттық технологиялар университеті медиакоммуникация және Қазақстан тарихы кафедрасының меңгерушісі (Қазақстан)

**Айтмағамбетов Алтай Зуфарович** — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының профессоры (Қазақстан)

**Бахтиярова Елена Ажибековна** — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының меңгерушісі (Қазақстан)

**Канибек Сансызбай** — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессор-зерттеушісі (Қазақстан)

**Тынымбаев Сахибай** — техника ғылымдарының кандидаты, профессор, Халықаралық ақпараттық технологиялар университеті компьютерлік инженерия кафедрасының профессор-зерттеушісі (Қазақстан)

**Алмисреб Али Абд** — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының қауымдастырылған профессоры (Қазақстан)

**Мохамед Ахмед Хамада** — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының қауымдастырылған профессоры (Қазақстан)

**Янг Им Чу** — PhD, Гачон университетінің профессоры (Оңтүстік Корея)

**Тадеш Валлас** — PhD, Адам Мицкевич атындағы (Польша) университеттің проректоры

**Мамырбаев Оркен Жумажанович** — PhD, ҚР ҒЖБМ Ғылым комитеті ақпараттық және есептеу технологиялары институты ӨМК директорының ғылым жөніндегі орынбасары (Қазақстан)

**Бушув Сергей Дмитриевич** — техника ғылымдарының докторы, профессор, Украинаның "УКРНЕТ" жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сулест университеті жобаларды басқару кафедрасының меңгерушісі (Украина)

**Белюшицкая Светлана Васильевна** — техника ғылымдарының докторы, доцент, Astana IT University есептеу және деректер ғылымы кафедрасының профессоры (Қазақстан)

## ЖАУАПТЫ РЕДАКТОР:

**Мрзабаева Раушан Жалиевна** — магистр, Халықаралық ақпараттық технологиялар университетінің редакторы (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Мерзімділігі: жылына 4 рет.

Басылым тілі: қазақ, орыс, ағылшын.

Тақырып бағыты: "Ақпараттық технологиялар"; "Ақпараттық қауіпсіздік және коммуникациялық технологиялар"; "Әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология".

Журнал сайты: <https://journal.iitu.edu.kz>

Тарату: материалдар Creative Commons Attribution 4.0 лицензиясы бойынша таратылады

Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).

Авторлық құқық: © Халықаралық ақпараттық және коммуникациялық технологиялар журналы, 2026

РЕДАКЦИЯ

ГЛАВНЫЙ РЕДАКТОР:

**Колесникова Катерина Викторовна** — доктор технических наук, профессор, проректор по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

**Ипалакова Мадина Тулегеновна** — кандидат технических наук, ассоциированный профессор, директор департамента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

**Разак Абдул** — PhD, профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

**Лучио Томмазо де Паолис** — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

**Лиз Бэкон** — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

**Микеле Пагано** — PhD, профессор Университета Пизы (Италия)

**Отелбаев Мухтарбай Отелбайулы** — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

**Рысбайулы Болатбек** — доктор физико-математических наук, профессор, профессор Astana IT University (Казахстан)

**Дайнеко Евгения Александровна** — PhD, профессор-исследователь кафедры информационных систем Международного университета информационных технологий (Казахстан)

**Дузбаев Нуржан Токкужаевич** — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

**Синчев Бахтгерей Куспанович** — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

**Сейлова Нургуль Абадуллаевна** — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

**Мухамедиева Ардак Габитовна** — кандидат экономических наук, декан факультета бизнеса медиа и управления Международного университета информационных технологий (Казахстан)

**Абдикаликова Замира Турсынбаевна** — PhD, ассоциированный профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

**Шильдибеков Ерлан Жаржанович** — PhD, ассоциированный профессор, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

**Дамеуш Максустовна Ескендирова** — кандидат технических наук, ассоциированный профессор, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

**Ниязгулова Айгуль Аскарбековна** — кандидат филологических наук, доцент, профессор, заведующая кафедрой медиакоммуникации и истории Казахстана Международного университета информационных технологий (Казахстан)

**Айтмаганбетов Алтай Зуфарович** — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

**Бахтиярова Елена Ажибековна** — кандидат технических наук, ассоциированный профессор, заведующая кафедрой радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

**Канибек Сансызбай** — PhD, ассоциированный профессор, профессор-исследователь кафедры кибербезопасности, Международного университета информационных технологий (Казахстан)

**Тынымбаев Сахиябай** — кандидат технических наук, профессор, профессор-исследователь кафедры компьютерной инженерии, Международного университета информационных технологий (Казахстан)

**Алимсерб Али Абд** — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

**Мохамед Ахмед Хамада** — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

**Янг Им Чу** — PhD, профессор университета Гачон (Южная Корея)

**Тадеуш Валлас** — PhD, проректор университета имен Адама Мицкевича (Польша)

**Мамырбаев Оркен Жумажанович** — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

**Бушуев Сергей Дмитриевич** — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

**Белошицкая Светлана Васильевна** — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

**Мрзабаева Раушан Жалиевна** — магистр, редактор Международного университета информационных технологий (Казахстан)

Международный журнал информационных и коммуникационных технологий

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Периодичность: 4 выпусков в год.

Язык издания: казахский, русский, английский.

Тематическая направленность: "Информационные технологии"; "Информационная безопасность и коммуникационные технологии"; "Цифровые технологии в развитии социо-экономических систем".

Сайт журнала: <https://journal.iitu.edu.kz>

Распространение: материалы распространяются по лицензии Creative Commons Attribution 4.0

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Авторские права: © Международный журнал информационных и коммуникационных технологий, 2026

## CONTENTS

### DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

**D. Abzhanova, A. Biloshchytskyi**

A MODEL AND METHOD FOR MANAGING DATA ON EMISSIONS FROM STATIONARY SOURCES OF POLLUTION IN AN INTELLIGENT ENVIRONMENTAL MONITORING SYSTEM .....11

**A. Slanbekova, M. Rakhimzhanova, A. Zhanibekova, A. Alimagambetova, M. Xudoyberganov**

EARLY DETECTION OF HYDROLOGICAL HAZARDS BASED ON SPATIOTEMPORAL ANALYSIS .....27

**A. Sakhipov, R. Seitbek, D. Yedilkhan**

REAL-TIME PASSENGER FLOW PREDICTION USING AUTOMATED DATA PREPARATION .....38

**K. Tokayev, A. Moldagulova**

A REVIEW OF PREDICTIVE MODELS APPROACHES TO STOCK MARKET FORECASTING IN KAZAKHSTAN .....60

### INFORMATION TECHNOLOGY

**F.N. Abdraimova, A.A. Kereibayeva, D.S. Dyussenova, D.A. Aliyeva, T.Zh. Toktarova**

AI TECHNOLOGIES IN LANGUAGE EDUCATION: PRACTICAL ASPECTS AND CHALLENGES OF STUDENT USAGE .....83

**G. Azieva, M. Yessenova, A. Abzhapparova, G. Abdikerimova, P. Schmidt**

HYBRID STACKING FRAMEWORK FOR CROP CLASSIFICATION USING UAV DATA .....97

**A.K. Aitim**

JOINT MORPHOLOGICAL DISAMBIGUATION AND POS TAGGING FOR AGGLUTINATIVE LANGUAGES .....109

**A. Auyezova, Aigerim Aitim, B. Sinchev, W. Wójcik**

NP-COMPLETE ALGORITHMS IN LEGALTECH: HYBRID ARCHITECTURE FOR SEMANTIC RETRIEVAL IN KAZAKHSTAN'S LEGAL ACTS .....125

**Y. Hozak, S. Paliy, T. Honcharenko**

EVALUATING OPTIMAL CONFIGURATIONS OF SHALLOW CNNs ON LOW-COMPLEXITY DATASETS .....138

**S.A. Yesniyazova, S.T. Kaimov**

PREDICTIVE MAINTENANCE OF HEAVY-DUTY TRUCKS USING EXPLAINABLE MACHINE LEARNING .....153

**T. Imanbekova, Zh. Ibrayeva, G. Jakanova, G. Askanbay**

DATA COMPRESSION ALGORITHM BASED ON WAVELET TRANSFORMER; ANALYSIS AND IMPLEMENTATION IN MATLAB .....167

**B.Z. Kenzhegulov, Zh.T. Bilyalova, K.N. Uteuliyeva, L. Nurgaliyeva, Sh.S. Nurzhanova**

A MATHEMATICAL AND ALGORITHMIC APPROACH TO THE DEVELOPMENT OF AN INTELLIGENT TEXT-TO-SQL SYSTEM BASED ON LARGE LANGUAGE MODELS .....185

**M. Kabdullin, A. Kairbekov, A. Kabdullin, L. Naizabayeva, A. Mukhit**

DESIGN AND IMPLEMENTATION OF A WEB-BASED PROTOTYPE FOR CARDIOLOGICAL IMAGE MANAGEMENT .....206

**L. Kurmangazyeva, Sh. Yelezhanova, S. Iskakova, Sh. Kodanova, H.İ. Bülbül**

APPROACHES TO BUILDING A HYBRID INTELLIGENT DECISION SUPPORT SYSTEM BASED ON FUZZY LOGIC, XAI, AND LARGE LANGUAGE MODELS .....219

**Z. Momynkulov, B. Omarov**

DDPG FOR TRAJECTORY GENERATION .....236

**F. Myrzabekov, N. Mustafin, M. Mukushev**

WHY IS THIS NEWS NEGATIVE?: AN ATTENTION-BASED VISUALIZATION TECHNIQUE FOR EXPLAINABLE SENTIMENT ANALYSIS. AN ANALYTICAL SURVEY AND CONCEPTUAL FRAMEWORK .....248

**N.Sh. Maxutova, J.A. Tussupov, A.A. Shekerbek, Zh.E. Kenzhebayeva, Q.O. Rakhimov**

MACHINE LEARNING FOR COMPREHENSIVE EVALUATION OF CARDIOVASCULAR DISEASE RISK AND BIOCHEMICAL ALTERATIONS: FOCUS ON ASPARTATE AMINOTRANSFERASE .....260

**L. Naizabayeva, G. Turken, R.A. Druba, A.Sh. Yawar**

LEVERAGING ENTERPRISE DATA WAREHOUSING TO IMPROVE WEB-BASED SALES

|                                                                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| OPERATIONS .....                                                                                                                                                                                                          | 275 |
| <b>S. Nurgaliyeva, N. Naiman, K. Makhmetova, M. Kali, N. Zhakiyev</b><br>COMPARATIVE ANALYSIS OF MACHINE LEARNING METHODS FOR RAPID DETECTION OF<br>VEGETABLE DISEASE USING COMPUTER VISION .....                         | 291 |
| <b>O.S. Salykova, V.A. Madin, B.R. Salykov, D.N. Komarov, N.V. Manuilov</b><br>INTEGRATION OF MEMS ACCELEROMETER SENSOR MODULES IN INDUSTRIAL MONITORING<br>SYSTEMS .....                                                 | 309 |
| <b>R. Taberkhan, M.A. Sambetbayeva, G. Kalman</b><br>KAZCAUSAL: THE FIRST CORPUS-BASED ANNOTATION OF CAUSAL RELATIONSHIPS IN THE<br>KAZAKH LANGUAGE .....                                                                 | 323 |
| <b>S.Tynymbayev, S.E. Mamanova, R. Berdybayev, Zh.E. Temirbekova, T. Chinibayeva</b><br>DIVIDING DEVICES WITH PRELIMINARY PREPARATION OF MULTIPLES OF THE DIVISOR .....                                                   | 335 |
| <b>K.N. Uteuliyeva, B.Z. Kenzhegulov, T.A. Karazhigitova, H.İ. Bülbül, Z.Zh. Zhanuzakova</b><br>MATHEMATICAL AND ALGORITHMIC APPROACHES TO THE DEVELOPMENT OF A<br>COLLABORATIVE FILTERING-BASED RECOMMENDER SYSTEM ..... | 351 |
| <b>S. Sharmukhanbet, G. Turmukhanova, O. Findik, V. Makhatova, L. Kurmangaziyeva</b><br>HIGH-PRECISION ROBOTIC ASSEMBLY UNDER VARIABLE ILLUMINATION: A ROBUST<br>MECHATRONIC ARCHITECTURE FOR VISUAL SERVOING .....       | 372 |

### INFORMATION SECURITY AND COMMUNICATIONTECHNOLOGIES

|                                                                                                                                                                                                                                                                                                                                 |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>A. Amirbay, Z. Amanbaikyzy, K. Maxutova, A. Mukhanova, M. Kassim</b><br>MACHINE LEARNING ALGORITHM FOR EARLY DETECTION OF AUTISM SPECTRUM DISORDERS<br>IN CHILDREN BASED ON MULTIMODAL ANALYSIS OF EYE MOVEMENTS AND FACIAL<br>EXPRESSIONS .....                                                                             | 390 |
| <b>K. Baisylbayeva, Sh. Mussiraliyeva, Zh. Yeltay</b><br>DETECTION OF EXTREMIST IDEOLOGY IN THE KAZAKH LANGUAGE: ANNOTATION<br>CHALLENGES AND DEEP LEARNING APPROACHES .....                                                                                                                                                    | 405 |
| <b>M.A. Bolatbek, A.M.Usmanova, K.B. Bagitova, G.B. Baispay</b><br>DEVELOPMENT AND RESEARCH OF A METHOD FOR ANALYZING NETWORK TRAFFIC TO<br>IDENTIFY A CYBER THREAT .....                                                                                                                                                       | 424 |
| <b>D.I. Prokopovych-Tkachenko, N.K. Zhumagaliyeva, D.N. Shchytov, N.F. Mormul, D.A. Cherkaskyi</b><br>FUZZY MODEL FOR EVALUATING INFORMATION SECURITY PARAMETERS OF INFORMATION<br>SYSTEMS UNDER INCOMPLETE AND QUALITATIVE DATA: CONSTRUCTION METHODOLOGY,<br>RULE BASE TUNING, AND DEMONSTRATION CASE FOR ORGANIZATIONS ..... | 442 |
| <b>E.A. Pustovoy, O.A. Pustovaya, A.N. Raushanova, I.S. Zaurbekov</b><br>EVALUATION OF THE EFFECTIVENESS OF SYNTHESIS OF STOCHASTIC MODELS WITH<br>CONTROLLED PROPERTIES .....                                                                                                                                                  | 468 |
| <b>Y. Serzhan, T. Umarov, A. Abilbayeva</b><br>FRAUD DETECTION IN CREDIT CARD TRANSACTIONS USING MACHINE LEARNING: A<br>COMPARATIVE ANALYSIS .....                                                                                                                                                                              | 484 |
| <b>T. Babenko, Ye. Bakhtiyarova, D. Yeskendirova, A. Sysoev, Sh. Makilenov</b><br>FLEXIBLE IDPS BASED ON MACHINE LEARNING ALGORITHMS FOR DIPTYCHS AND IOT EDGE<br>DEVICE PROTECTION .....                                                                                                                                       | 496 |

### МАЗМҰНЫ

#### ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

|                                                                                                                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Д.Е. Абжанов, А.А. Белоощицкий</b><br>ЭКОЛОГИЯЛЫҚ МОНИТОРИНГТІҢ ЗИЯТКЕРЛІК ЖҮЙЕСІНДЕГІ СТАЦИОНАРЛЫҚ ЛАСТАНУ КӨЗ-<br>ДЕРІНІҢ ШЫҒАРЫНДЫЛАРЫ ТУРАЛЫ ДЕРЕКТЕРДІ БАСҚАРУДЫҢ МОДЕЛІ МЕН ӘДІСІ .....              | 11 |
| <b>А.Е. Сланбекова, М.Б. Рахимжанова, А.И. Жанибекова, А.З. Алимагамбетова, М. Худойбергенов</b><br>КЕҢІСТІКТІК-УАҚЫТТЫҚ (SPATIOTEMPORAL) ТАЛДАУ НЕГІЗІНДЕ ГИДРОЛОГИЯЛЫҚ ҚАУІП-<br>ҚАТЕРДІ ЕРТЕ АНЫҚТАУ ..... | 27 |
| <b>А. Сахипов, Р. Сеитбек, Д. Едильхан</b><br>АВТОМАТТАНДЫРЫЛҒАН ДЕРЕКТЕРДІ ДАЙЫНДАУДЫ ПАЙДАЛАНА ОТЫРЫП, ЖОЛАУШЫЛАР                                                                                           |    |

|                                                                       |    |
|-----------------------------------------------------------------------|----|
| АҒЫНЫН НАҚТЫ УАҚЫТ РЕЖИМІНДЕ БОЛЖАУ .....                             | 38 |
| <b>К. Токаев, А. Молдагулова</b>                                      |    |
| ҚАЗАҚСТАНДАҒЫ ҚОР НАРЫҒЫН БОЛЖАУДЫҢ БОЛЖАМДЫҚ МОДЕЛЬДЕРІНЕ ШОЛУ ..... | 60 |

## АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

|                                                                                                                                                                                    |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Ф.Н. Абдраимова, А.А. Керейбаева, Д.С. Дюсенова, Д.А. Алиева, Т.Ж. Токтарова</b>                                                                                                |     |
| ТІЛ БІЛІМІНДЕ ЖАСАНДЫ ИНТЕЛЛЕКТ ТЕХНОЛОГИЯЛАРЫ: СТУДЕНТТЕР ҚОЛДАНУЫНЫҢ ПРАКТИКАЛЫҚ АСПЕКТІЛЕРІ МЕН МӘСЕЛЕЛЕРІ .....                                                                | 83  |
| <b>Г.Т. Азиева, М.Б. Есенова, А.К. Абжаппарова, Г.Б. Абдикеримова, Р. Schmidt</b>                                                                                                  |     |
| UAV ДЕРЕКТЕРІ НЕГІЗІНДЕ АУЫЛ ШАРУАШЫЛЫҒЫ DAҚЫЛДАРЫН ЖІКТЕУГЕ АРНАЛҒАН ГИБРИДТІ СТЕКИНГ МОДЕЛІ .....                                                                                | 97  |
| <b>Ә.Қ. Әйтiм</b>                                                                                                                                                                  |     |
| АГГЛЮТИНАТИВТІ ТІЛДЕРГЕ АРНАЛҒАН МОРФОЛОГИЯЛЫҚ ДИЗАМБИГУАЦИЯ МЕН POS-ТАҢБАЛАУДЫ БІРЛЕСІП МОДЕЛЬДЕУ .....                                                                           | 109 |
| <b>А. Әуезова, Ә. Әйтiм, Б. Синчев, В. Вуйчик</b>                                                                                                                                  |     |
| ҚАЗАҚСТАННЫҢ ҚҰҚЫҚТЫҚ АКТІЛЕРІНДЕГІ NP-ТОЛЫҚ АЛГОРИТМДЕР: SEMANTIC ІЗДЕУГЕ АРНАЛҒАН ГИБРИДТІ АРХИТЕКТУРА .....                                                                     | 125 |
| <b>Я. Гозак, С. Паллий, Т. Гончаренко</b>                                                                                                                                          |     |
| КҮРДЕЛІЛІГІ ТӨМЕН ДЕРЕКТЕР ЖИЙНТЫҚТАРЫНДА ТЕРЕҢ ЕМЕС ОРАМДЫ НЕЙРОНДЫҚ ЖЕЛІЛЕРДІҢ (CNN) ОҢТАЙЛЫ КОНФИГУРАЦИЯЛАРЫН БАҒАЛАУ .....                                                     | 138 |
| <b>С.А. Есниязова, С.Т. Каимов</b>                                                                                                                                                 |     |
| ТҮСІНДІРІЛЕТІН МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНА ОТЫРЫП АУЫР ЖҮК КӨЛІКТЕРІНЕ БОЛЖАМДЫ ТЕХНИКАЛЫҚ ҚЫЗМЕТ КӨРСЕТУ .....                                                                      | 153 |
| <b>Т.Д. Иманбеткова, Ж.Б. Ибраева, Г.Т. Джаканова, Г.Т. Асқанбай</b>                                                                                                               |     |
| МӨЛІМЕТТЕРДІ ВЕЙВЛЕТ-ТҮРЛЕНДІРГІШТІҢ НЕГІЗІНДЕ ҚЫСУ АЛГОРИТМІ; MATLAB ОРТАСЫНДА ТАЛДАУ ЖӘНЕ ІСКЕ АСЫРУ .....                                                                       | 167 |
| <b>Б.З. Кенжегулов, Ж.Т. Билялова, К.Н. Утеулиева, Л. Нургалиева, Ш.С. Нуржанова</b>                                                                                               |     |
| ҮЛКЕН ТІЛДІК МОДЕЛЬДЕР НЕГІЗІНДЕ ИНТЕЛЛЕКТУАЛДЫ TEXT-TO-SQL ЖҮЙЕСІН ӨЗІРЛЕУДІҢ МАТЕМАТИКАЛЫҚ-АЛГОРИТМДІК ТӘСІЛІ .....                                                              | 185 |
| <b>М. Кабдуллин, А. Каирбеков, А. Кабдуллин, Л. Найзабаева, А. Мухит</b>                                                                                                           |     |
| КАРДИОЛОГИЯЛЫҚ КЕСКІНДЕРДІ БАСҚАРУҒА АРНАЛҒАН ВЕБ-НЕГІЗДЕЛГЕН ПРОТОТИПТІ ӨЗІРЛЕУ ЖӘНЕ ІСКЕ АСЫРУ .....                                                                             | 206 |
| <b>Л. Құрманғазиева, Ш. Ележанова, С. Исакова, Ш. Коданова, Х. И. Бұлбұл</b>                                                                                                       |     |
| АЙҚЫН ЕМЕС ЛОГИКА, ХАІ ЖӘНЕ ҮЛКЕН ТІЛДІК МОДЕЛЬДЕР НЕГІЗІНДЕ ГИБРИДТІ ЗИЯТКЕРЛІК ШЕШІМ ҚАБЫЛДАУДЫ ҚОЛДАУ ЖҮЙЕСІН ҚҰРУ ТӘСІЛДЕРІ .....                                              | 219 |
| <b>З. Момынкулов, Б. Омаров</b>                                                                                                                                                    |     |
| ТРАЕКТОРИЯЛАРДЫ ГЕНЕРАЦИЯЛАУ ҮШІН DDPRG .....                                                                                                                                      | 236 |
| <b>Ф. Мырзабеков, Н. Мустафин, М. Мукушев</b>                                                                                                                                      |     |
| БҮЛ ЖАҒАЛЫҚ НЕЛІКТЕН ТЕРІС?: ТҮСІНДІРІЛЕТІН СЕНТИМЕНТТІК ТАЛДАУҒА АРНАЛҒАН НАЗАР АУДАРУ МЕХАНИЗМІНЕ НЕГІЗДЕЛГЕН ВИЗУАЛИЗАЦИЯ ӘДІСІ. ТАЛДАУЛЫҚ ШОЛУ ЖӘНЕ ТҰЖЫРЫМДАМАЛЫҚ НЕГІЗ ..... | 248 |
| <b>Н.Ш. Максүтова, Ж.А. Түсүпов, А.Ә. Шекербек, Ж.Е. Кенжебаева, К.О. Рахимов</b>                                                                                                  |     |
| ЖҮРЕК-ҚАН ТАМЫРЛАРЫ АУРУЛАРЫНЫҢ ҚАУІП-ҚАТЕРІН ЖӘНЕ БИОХИМИЯЛЫҚ ӨЗГЕРІСТЕРДІ КЕШЕНДІ БАҒАЛАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ: АСПАРТАМИНОТРАНСФЕРАЗАҒА ЕРЕКШЕ НАЗАР .....                      | 260 |
| <b>Л. Найзабаева, Г. Түркен, Р.А. Друба, А.Ш. Явар</b>                                                                                                                             |     |
| ВЕБ НЕГІЗІНДЕГІ САТУ ОПЕРАЦИЯЛАРЫН ЖЕТІЛДІРУ ҮШІН КӘСІПОРЫНДЫҚ ДЕРЕКТЕР ҚОЙМАСЫН ПАЙДАЛАНУ .....                                                                                   | 275 |
| <b>С. Нургалиева, Н. Найман, К. Махметова, М. Кали, Н. Жакиев</b>                                                                                                                  |     |
| КОМПЬЮТЕРЛІК КӨРУ ТЕХНОЛОГИЯСЫН ПАЙДАЛАНА ОТЫРЫП, КӨКӨНІС DAҚЫЛДАРЫНЫҢ АУРУЛАРЫН ЖЕДЕЛ АНЫҚТАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН САЛЫСТЫРМАЛЫ ТАЛДАУ .....                      | 291 |
| <b>О.С. Салыкова, В.А. Мадин, Б.Р. Салыков, Д.Н. Комаров, Н.В. Манұйлов</b>                                                                                                        |     |
| ӨНЕРКӘСІПТІК МОНИТОРИНГ ЖҮЙЕЛЕРІНДЕГІ MEMS-АКСЕЛЕРОМЕТРЛЕРДІҢ СЕНСОРЛЫҚ МОДУЛЬДЕРІН ИНТЕГРАЦИЯЛАУ .....                                                                            | 309 |
| <b>Р. Таберхан, М.А. Самбетбаева, Г. Қалман</b>                                                                                                                                    |     |

|                                                                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| KAZCAUSAL: ҚАЗАҚ ТІЛІНДЕГІ СЕБЕП-САЛДАРЛЫҚ ҚАТЫНАСТАРДЫҢ АЛҒАШҚЫ<br>КОРПУСТЫҚ АННОТАЦИЯСЫ .....                                                                                                                           | 323 |
| <b>С. Тынымбаев, С.Е. Маманова, Р. Бердібаев, Ж.Е. Темірбекова, Т. Чинибаева</b><br>БӨЛГІШТІҢ ЕСЕЛІ МӨНДЕРІН АЛДЫН АЛА ДАЙЫНДАУМЕН ЖҮЗЕГЕ АСЫРЫЛАТЫН БӨЛУ<br>ҚҰРЫЛҒЫЛАРЫ .....                                            | 335 |
| <b>К.Н. Утеулиева, Б.З. Кенжегулов, Т.А. Каражигитова, Х. Булбул, З.Ж. Жанузакова</b><br>КОЛЛАБОРАТИВТІК СҮЗГІЛЕУ НЕГІЗІНДЕГІ ҰСЫНЫМДЫҚ ЖҮЙЕНІ ӨЗІРЛЕУДІҢ<br>МАТЕМАТИКАЛЫҚ-АЛГОРИТМДІК ТӘСІЛДЕРІ .....                    | 351 |
| <b>С. Шармұханбет, Г. Тұрмұханова, О. Финдик, В. Махатова, Л. Курмангазиева</b><br>АЙНЫМАЛЫ ЖАРЫҚ ЖАҒДАЙЫНДАҒЫ ЖОҒАРЫ ДӘЛДІКТІ РОБОТТЫҚ ҚҰРАСТЫРУ: ВИЗУАЛ-<br>ДЫ СЕРВОТЕЖЕУДІҢ ТӨЗІМДІ МЕХАТРОНИКАЛЫҚ АРХИТЕКТУРАСЫ ..... | 372 |

## АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

|                                                                                                                                                                                                                                                                                                                                           |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>А. Амирбай, З. Аманбайқызы, К. МаксUTOBA, А. Муханова, М. Kassim</b><br>КӨЗ ҚОЗҒАЛЫСТАРЫ МЕН БЕТ МИМИКА БЕЛГІЛЕРІН МУЛЬТИМОДАЛЬДЫ ТАЛДАУҒА НЕГІЗ-<br>ДЕЛГЕН БАЛАЛАРДАҒЫ АУТИЗМ СПЕКТРІНІҢ БҰЗЫЛЫСТАРЫН ЕРТЕ АНЫҚТАУҒА АРНАЛҒАН<br>МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМІ .....                                                                      | 390 |
| <b>К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж. Елтай</b><br>ҚАЗАҚ ТІЛІНДЕГІ ЭКСТРЕМИСТІК ИДЕОЛОГИЯНЫ АНЫҚТАУ: АННОТАЦИЯЛАУ МӘСЕЛЕЛЕРІ<br>ЖӘНЕ ТЕРЕҢ ОҚЫТУ ТӘСІЛДЕРІ .....                                                                                                                                                                      | 405 |
| <b>М.А. Болатбек, А.М. Усманова, Қ.Б. Багитова, Г.Б. Байспай</b><br>КИБЕР ҚАУІПТІ АНЫҚТАУ ҮШІН ЖЕЛІЛІК ТРАФИКТІ ТАЛДАУ ӘДІСІН ӨЗІРЛЕУ ЖӘНЕ<br>ЗЕРТТЕУ .....                                                                                                                                                                               | 424 |
| <b>Д.И. Прокопович-Ткаченко, Н.К. Жумағалиева, Д.Н. Щитов, Н.Ф. Мормуль, Д.А. Черкасский</b><br>ТОЛЫҚ ЕМЕС ЖӘНЕ САПАЛЫҚ ДЕРЕКТЕР ЖАҒДАЙЫНДА АҚПАРАТТЫҚ ЖҮЙЕЛЕРДІҢ АҚПА-<br>РАТТЫҚ ҚАУІПСІЗДІК ПАРАМЕТРЛЕРІН БАҒАЛАУДЫҢ БҰЛЫҢҒЫР МОДЕЛІ: ҚҰРУ ӘДІСТЕМЕСІ,<br>ЕРЕЖЕЛЕР БАЗАСЫН БАПТАУ ЖӘНЕ ҰЙЫМДАРҒА АРНАЛҒАН ДЕМОНСТРАЦИЯЛЫҚ<br>КЕЙС ..... | 442 |
| <b>Е.А. Пустовой, О.А. Пустовая, А.Н. Раушанова, И.С. Заурбеков</b><br>БАСҚАРЫЛАТЫН ҚАСИЕТТЕРІ БАР СТОХАСТИКАЛЫҚ МОДЕЛЬДЕРДІ СИНТЕЗДЕУДІҢ<br>ТИІМДІЛІГІН БАҒАЛАУ .....                                                                                                                                                                    | 468 |
| <b>Е. Сержан, Т. Умаров, А. Әбілбаева</b><br>МАШИНАЛЫҚ ОҚУ ӘДІСІ АРҚЫЛЫ КРЕДИТ КАРТА ОПЕРАЦИЯЛАРЫНДАҒЫ АЛАЯҚТЫҚТЫ<br>АНЫҚТАУ: САЛЫСТЫРМАЛЫ ТАЛДАУ .....                                                                                                                                                                                   | 484 |
| <b>Т.В. Бабенко, Е.А. Бахтиярова, Д.М. Ескендинова, А. Сысоев, Ш.Н. Макиленов</b><br>ДИПТИХ ЖӘНЕ IOT ШЕТКІ ҚҰРЫЛҒЫЛАРЫН ҚОРҒАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ<br>АЛГОРИТМДЕРІНЕ НЕГІЗДЕЛГЕН ИКЕМДІ IDPS .....                                                                                                                                 | 496 |

## СОДЕРЖАНИЕ

### ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

|                                                                                                                                                                                                                         |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Д.Е. Абжанова, А.А. Белошицкий</b><br>МОДЕЛЬ И МЕТОД УПРАВЛЕНИЯ ДАННЫМИ О ВЫБРОСАХ СТАЦИОНАРНЫХ ИСТОЧНИКОВ<br>ЗАГРЯЗНЕНИЯ В ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЕ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА .....                                | 11 |
| <b>А.Е. Сланбекова, М.Б. Рахимжанова, А.И. Жанибекова, А.З. Алимагамбетова, М. Худойбергенов</b><br>РАННЕЕ ВЫЯВЛЕНИЕ ГИДРОЛОГИЧЕСКИХ ОПАСНОСТЕЙ НА ОСНОВЕ ПРОСТРАНСТВЕННО-<br>ВРЕМЕННОГО (SPATIOTEMPORAL) АНАЛИЗА ..... | 27 |
| <b>А. Сахипов, Р. Сеитбек, Д. Едильхан</b><br>ПРОГНОЗИРОВАНИЕ ПАССАЖИРОПОТОКА В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ АВ-<br>ТОМАТИЗИРОВАННОЙ ПОДГОТОВКИ ДАННЫХ .....                                                       | 38 |
| <b>К. Токаев, А. Молдагулова</b><br>ОБЗОР ПРОГНОСТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ФОНДОВОГО РЫНКА<br>В КАЗАХСТАНЕ .....                                                                                             | 60 |

## ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

|                                                                                                                                                                                                                                             |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Ф.Н. Абдраимова, А.А. Керейбаева, Д.С. Дюсенова, Д.А. Алиева, Т.Ж. Токтарова</b><br>ТЕХНОЛОГИИ ИИ В ЯЗЫКОВОМ ОБРАЗОВАНИИ: ПРАКТИЧЕСКИЕ АСПЕКТЫ И ПРОБЛЕМЫ<br>ПРИМЕНЕНИЯ СТУДЕНТАМИ .....                                                 | 83  |
| <b>Г.Т. Азиева, М.Б. Есенова, А.К. Абжаппарова, Г.Б. Абдикеримова, Р. Schmidt</b><br>ГИБРИДНАЯ МОДЕЛЬ СТЕКИНГА ДЛЯ КЛАССИФИКАЦИИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР<br>ПО ДАННЫМ UAV .....                                                        | 97  |
| <b>Ә.Қ. Әйтiм</b><br>СОВМЕСТНАЯ МОРФОЛОГИЧЕСКАЯ ДИЗАМБИГУАЦИЯ И POS-РАЗМЕТКА ДЛЯ<br>АГГЛЮТИНАТИВНЫХ ЯЗЫКОВ .....                                                                                                                            | 109 |
| <b>А. Ауезова, А. Айтим, Б. Синчев, В. Вуйчик</b><br>NP-ПОЛНЫЕ АЛГОРИТМЫ В LEGALTECH: ГИБРИДНАЯ АРХИТЕКТУРА ДЛЯ СЕМАНТИЧЕСКОГО<br>ПОИСКА В ПРАВОВЫХ АКТАХ КАЗАХСТАНА .....                                                                  | 125 |
| <b>Я. Гозак, С. Палий, Т. Гончаренко</b><br>ОЦЕНКА ОПТИМАЛЬНЫХ КОНФИГУРАЦИЙ НЕГЛУБОКИХ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ<br>НА ДАТАСЕТАХ НИЗКОЙ СЛОЖНОСТИ .....                                                                                    | 138 |
| <b>С.А. Есниязова, С.Т. Каимов</b><br>ПРЕДИКТИВНОЕ ТЕХНИЧЕСКОЕ ОБСЛУЖИВАНИЕ ТЯЖЕЛЫХ ГРУЗОВИКОВ С ИСПОЛЬЗОВАНИ-<br>ЕМ ОБЪЯСНИМОГО МАШИННОГО ОБУЧЕНИЯ .....                                                                                   | 153 |
| <b>Т.Д. Иманбекова, Ж.Б. Ибраева, Г.Т. Джаканова, Г.Т. Асканбай</b><br>АЛГОРИТМ СЖАТИЯ ДАННЫХ НА ОСНОВЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАТЕЛЯ: АНАЛИЗ И РЕАЛИЗА-<br>ЦИЯ В MATLAB .....                                                                    | 167 |
| <b>Б.З. Кенжегулов, Ж.Т. Билялова, К.Н. Утеулиева, Л. Нургалиева, Ш.С. Нуржанова</b><br>МАТЕМАТИКО-АЛГОРИТМИЧЕСКИЙ ПОДХОД К РАЗРАБОТКЕ ИНТЕЛЛЕКТУАЛЬНОЙ ТЕХТ-<br>TO-SQL СИСТЕМЫ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ .....                    | 185 |
| <b>М. Кабдуллин, А. Каирбеков, А. Кабдуллин, Л. Найзабаева, А. Мухит</b><br>РАЗРАБОТКА И РЕАЛИЗАЦИЯ ВЕБ-ПРОТОТИПА ДЛЯ УПРАВЛЕНИЯ КАРДИОЛОГИЧЕСКИМИ<br>ИЗОБРАЖЕНИЯМИ .....                                                                   | 206 |
| <b>Л. Курмангазиева, Ш. Ележанова, С. Искакова, Ш. Коданова, Х. И. Бюльбюль</b><br>ПОДХОДЫ К ПОСТРОЕНИЮ ГИБРИДНОЙ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ<br>ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ НЕЧЕТКОЙ ЛОГИКИ, ХАИ И БОЛЬШИХ ЯЗЫКОВЫХ<br>МОДЕЛЕЙ ..... | 219 |
| <b>З. Момынкулов, Б. Омаров</b><br>DDPG ДЛЯ ГЕНЕРАЦИИ ТРАЕКТОРИЙ .....                                                                                                                                                                      | 236 |
| <b>Ф. Мырзабеков, Н. Мустафин, М. Мукушев</b><br>ПОЧЕМУ ЭТА НОВОСТЬ НЕГАТИВНА?: МЕТОД ВИЗУАЛИЗАЦИИ НА ОСНОВЕ МЕХАНИЗМА<br>ВНИМАНИЯ ДЛЯ ОБЪЯСНИМОГО АНАЛИЗА ТОНАЛЬНОСТИ. АНАЛИТИЧЕСКИЙ ОБЗОР И<br>КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ .....                | 248 |
| <b>Н.Ш. МаксUTOва, Д.А. Тусупов, А.А. Шекербек, Ж.Е. Кенжебаева, К.О. Рахимов</b><br>МАШИННОЕ ОБУЧЕНИЕ ДЛЯ КОМПЛЕКСНОЙ ОЦЕНКИ РИСКА СЕРДЕЧНО-СОСУДИСТЫХ<br>ЗАБОЛЕВАНИЙ И БИОХИМИЧЕСКИХ ИЗМЕНЕНИЙ: АКЦЕНТ НА<br>АСПАРТАМИНОТРАНСФЕРАЗЕ ..... | 260 |
| <b>Л. Найзабаева, Г. Туркен, Р.А. Друба, А.Ш. Явар</b><br>ИСПОЛЬЗОВАНИЕ КОРПОРАТИВНОГО ХРАНИЛИЩА ДАННЫХ ДЛЯ УЛУЧШЕНИЯ ВЕБ-ОРИЕНТИ-<br>РОВАННЫХ ОПЕРАЦИЙ ПРОДАЖ .....                                                                        | 275 |
| <b>С. Нургалиева, Н. Найман, К. Махметова, М. Кали, Н. Жакиев</b><br>СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ БЫСТРОГО ОБНАРУЖЕ-<br>НИЯ ЗАБОЛЕВАНИЙ ОВОЩНЫХ КУЛЬТУР С ИСПОЛЬЗОВАНИЕМ КОМПЬЮТЕРНОГО<br>ЗРЕНИЯ .....               | 291 |
| <b>О.С. Салыкова, В.А. Мадин, Б.Р. Салыков, Д.Н. Комаров, Н.В. Мануйлов</b><br>ИНТЕГРАЦИЯ СЕНСОРНЫХ МОДУЛЕЙ MEMS-АКСЕЛЕРОМЕТРОВ В СИСТЕМАХ<br>ПРОМЫШЛЕННОГО МОНИТОРИНГА .....                                                               | 309 |
| <b>Р. Таберхан, М.А. Самбетбаева, Г. Калман</b><br>KAZCAUSAL: ПЕРВАЯ КОРПУСНАЯ АННОТАЦИЯ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ НА<br>КАЗАХСКОМ ЯЗЫКЕ .....                                                                                           | 323 |
| <b>С. Тынымбаев, С.Е. Маманова, Р. Бердибаев, Ж.Е. Темирбекова, Т. Чинибаева</b>                                                                                                                                                            |     |

|                                                                                                                                                                                                                        |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| УСТРОЙСТВА ДЕЛЕНИЯ ЧИСЕЛ С ПРЕДВАРИТЕЛЬНОЙ ПОДГОТОВКОЙ КРАТНЫХ ДЕЛИТЕЛЮ .....                                                                                                                                          | 335 |
| <b>К.Н. Утеулиева, Б.З. Кенжегулов, Т.А. Каражигитова, Х.Бюльбюль, З.Ж. Жанузакова</b><br>МАТЕМАТИКО-АЛГОРИТМИЧЕСКИЕ ПОДХОДЫ К РАЗРАБОТКЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ .....          | 351 |
| <b>С. Шармуханбет, Г. Турмуханова, О.Финдик, В.Махатова, Л. Курмангазиева</b><br>ВЫСОКОТОЧНАЯ РОБОТИЗИРОВАННАЯ СБОРКА ПРИ ПЕРЕМЕННОЙ ОСВЕЩЁННОСТИ: РОБАСТНАЯ МЕХАТРОННАЯ АРХИТЕКТУРА ВИЗУАЛЬНОГО СЕРВОУПРАВЛЕНИЯ ..... | 372 |

## ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

|                                                                                                                                                                                                                                                                                                                             |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>А. Амирбай, З. Аманбайкызы, К. МаксUTOва, А. Муханова, М. Kassim</b><br>АЛГОРИТМ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РАННЕГО ВЫЯВЛЕНИЯ РАССТРОЙСТВ АУТИСТИЧЕСКОГО СПЕКТРА У ДЕТЕЙ НА ОСНОВЕ МУЛЬТМОДАЛЬНОГО АНАЛИЗА ДАННЫХ ДВИЖЕНИЯ ГЛАЗ И МИМИЧЕСКИХ СИГНАЛОВ .....                                                                  | 390 |
| <b>К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж.Елтай</b><br>ОБНАРУЖЕНИЕ ЭКСТРЕМИСТСКОЙ ИДЕОЛОГИИ НА КАЗАХСКОМ ЯЗЫКЕ: ПРОБЛЕМЫ АННОТИРОВАНИЯ И МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ .....                                                                                                                                                     | 405 |
| <b>М.А. Болатбек, А.М. Усманова, К.Б. Багитова, Г.Б. Байспай</b><br>РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА АНАЛИЗА СЕТЕВОГО ТРАФИКА ДЛЯ ВЫЯВЛЕНИЯ КИБЕРУГРОЗЫ .....                                                                                                                                                               | 424 |
| <b>Д.И. Прокопович-Ткаченко, Н.К. Жумагалиева, Д.Н. Щитов, Н.Ф. Мормуль, Д.А. Черкасский</b><br>НЕЧЕТКАЯ МОДЕЛЬ ОЦЕНИВАНИЯ ПАРАМЕТРОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ СИСТЕМ В УСЛОВИЯХ НЕПОЛНЫХ И КАЧЕСТВЕННЫХ ДАННЫХ: МЕТОДИКА ПОСТРОЕНИЯ, НАСТРОЙКА БАЗЫ ПРАВИЛ И ДЕМОНСТРАЦИОННЫЙ КЕЙС ДЛЯ ОРГАНИЗАЦИЙ ..... | 442 |
| <b>Е.А. Пустовой, О.А. Пустовая, А.Н. Раушанова, И.С. Заурбеков</b><br>ОЦЕНКА ЭФФЕКТИВНОСТИ СИНТЕЗА СТОХАСТИЧЕСКИХ МОДЕЛЕЙ С УПРАВЛЯЕМЫМИ СВОЙСТВАМИ .....                                                                                                                                                                  | 468 |
| <b>Е. Сержан, Т. Умаров, А. Абильбаева</b><br>ВЫЯВЛЕНИЕ МОШЕННИЧЕСТВА С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ ПРИ ОПЕРАЦИЯХ С КРЕДИТНЫМИ КАРТАМИ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ .....                                                                                                                                                    | 484 |
| <b>Т. Бабенко, Е. Бахтиярова, Д. Ескендинова, А. Сысоев, Ш. Макиленов</b><br>ГИБКАЯ IDPS НА ОСНОВЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ МОНИТОРИНГА И ЗАЩИТЫ ПЕРИФЕРИЙНЫХ IoT-УСТРОЙСТВ .....                                                                                                                                  | 496 |

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 7. Is.2. Number 26 (2026). Pp. 38–59

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2026.26.2.003>

## REAL-TIME PASSENGER FLOW PREDICTION USING AUTOMATED DATA PREPARATION

*A. Sakhipov, R. Seiitbek\*, D. Yedilkhan*

Astana IT University, Astana, Kazakhstan.

E-mail: [242773@astanait.edu.kz](mailto:242773@astanait.edu.kz)

**Aivar Sakhipov** — Ph.D., Assistant Professor, School of Software Engineering, Astana IT University, Astana, Kazakhstan

<https://orcid.org/0000-0003-1045-4199>;

**Ramazan Seiitbek** — Master of Computer Science, Astana IT University, Astana, Kazakhstan

E-mail: [242773@astanait.edu.kz](mailto:242773@astanait.edu.kz), <https://orcid.org/0009-0007-3301-1872>;

**Didar Yedilkhan** — PhD, Professor at the School of Computer Engineering, Astana IT University, Astana, Kazakhstan

<https://orcid.org/0000-0002-6343-5277>.

© A. Sakhipov, R. Seiitbek, D. Yedilkhan

**Abstract.** Public transportation systems are vital for sustainable urban development, addressing critical challenges such as traffic congestion, environmental impact, and urban mobility. However, the dynamic and unpredictable nature of passenger demand complicates resource allocation and operational efficiency. Traditional forecasting methods, including ARIMA, fail to account for the non-linear and spatio-temporal complexities inherent in modern transit systems, necessitating the adoption of more advanced methodologies. This study introduces an innovative real-time passenger flow prediction framework tailored for Astana’s public transport network. The framework leverages state-of-the-art machine learning models, including LightGBM, Random Forest, and conceptual Spatio-Temporal Graph Neural Networks (ST-GNNs) proposed for future implementation. While the ST-GNN component is outlined conceptually, the experimental evaluation in this study focuses on ensemble and boosting models such as LightGBM and Gradient Boosting Regressor. A comprehensive data preprocessing pipeline was developed to address common issues such as missing values, outliers, and temporal inconsistencies, ensuring high-quality input data for predictive modeling. The proposed framework achieved a Mean Absolute Error (MAE) of 1.06, a Root Mean Squared Error (RMSE) of 1.70, and a Mean Absolute Percentage Error (MAPE) of 34.8%, demonstrating superior predictive performance compared to traditional methods. This accuracy enables precise demand forecasting, optimized



resource allocation, and enhanced passenger experiences. Unlike traditional approaches, the framework integrates advanced spatio-temporal modeling techniques, bridging the gap between theoretical advancements and practical applications. This research offers actionable insights for the development of intelligent transportation systems, providing a scalable and adaptive solution to the challenges of urban mobility in rapidly growing cities.

**Keywords:** passenger flow, machine learning, transportation systems, optimization, prediction algorithms

**For citation:** Sakhipov A., Seiitbek R., Yedilkhan D. (2026). Real-time passenger flow prediction using automated data preparation // International journal of information and communication technologies. Vol. 7. No. 26. Pp. 38-59. <https://doi.org/10.54309/IJICT.2026.26.2.003>. (In Eng.).

**Conflict of interest:** The authors declare that there is no conflict of interest.

**Acknowledgements.** *This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No.BR24992852 “Intelligent models and methods of Smart City digital ecosystem for sustainable development and the citizens’ quality of life improvement”).*

## АВТОМАТТАНДЫРЫЛҒАН ДЕРЕКТЕРДІ ДАЙЫНДАУДЫ ПАЙДАЛАНА ОТЫРЫП, ЖОЛАУШЫЛАР АҒЫНЫН НАҚТЫ УАҚЫТ РЕЖИМІНДЕ БОЛЖАУ

*А. Сахипов, Р. Сеитбек\*, Д. Едильхан*

Astana IT University, Астана, Қазақстан.

E-mail: 242773@astanait.edu.kz

**Сахипов Айвар** — PhD, Компьютерлік инженерия мектебінің ассистент-профессоры, Astana IT University, Астана, Қазақстан  
<https://orcid.org/0000-0003-1045-4199>;

**Сеитбек Рамазан** — компьютерлік ғылымдар магистрі, Astana IT University, Астана, Қазақстан

E-mail: 242773@astanait.edu.kz, <https://orcid.org/0009-0007-3301-1872>;

**Едильхан Дидар** — PhD, Компьютерлік инженерия мектебінің профессоры, Astana IT University, Астана, Қазақстан  
<https://orcid.org/0000-0002-6343-5277>.

© А. Сахипов, Р. Сеитбек, Д. Едильхан

**Аннотация.** Қоғамдық көлік жүйелері тұрақты қалалық дамуды қамтамасыз етуде маңызды рөл атқарады және көлік кептелісі, қоршаған ортаға әсер ету және қалалық мобильділік сияқты өзекті мәселелерді шешуге мүмкіндік береді. Алайда жолаушылар сұранысының динамикалық және алдын ала болжануы қиын сипаты ресурстарды бөлуді және операциялық тиімділікті қамтамасыз етуді күрделендіреді. ARIMA сияқты дәстүрлі болжау әдістері заманауи көлік жүйелеріне тән сызықтық

емес және кеңістіктік-уақыттық күрделіліктерді ескере алмайды. Бұл анағұрлым жетілдірілген әдістемелерді қолдану қажеттілігін туындатады. Осы зерттеуде Астана қаласының қоғамдық көлік желісіне бейімделген жолаушылар ағынын нақты уақыт режимінде болжаудың инновациялық құрылымы ұсынылады. Ұсынылған құрылым заманауи машиналық оқыту модельдерін, соның ішінде LightGBM, Random Forest және болашақта іске асыруға ұсынылған кеңістіктік-уақыттық графтық нейрондық желілерді (ST-GNN) пайдаланады. ST-GNN компоненті тұжырымдамалық деңгейде сипатталғандықтан, осы зерттеудегі эксперименттік бағалау LightGBM және Gradient Boosting Regressor сияқты ансамбльдік және бустинг модельдеріне бағытталған. Болжамдық модельдерді құру үшін жоғары сапалы бастапқы деректерді қамтамасыз ету мақсатында бос мәндер, шектен тыс мәндер және уақыттық сәйкессіздіктер сияқты кең таралған мәселелерді шешуге арналған деректерді алдын ала өңдеудің кешенді конвейері әзірленді. Ұсынылған құрылым орташа абсолюттік қателік (MAE) — 1,06, түбірлік орташа квадраттық қателік (RMSE) — 1,70 және орташа абсолюттік пайыздық қателік (MAPE) — 34,8 % көрсеткіштеріне қол жеткізді. Бұл нәтижелер дәстүрлі әдістермен салыстырғанда болжау тиімділігінің жоғары екенін көрсетті. Алынған дәлдік сұранысты неғұрлым нақты болжауға, ресурстарды бөлуді оңтайландыруға және жолаушыларға көрсетілетін қызмет сапасын арттыруға мүмкіндік береді. Дәстүрлі тәсілдерден айырмашылығы, ұсынылған құрылым заманауи кеңістіктік-уақыттық модельдеу әдістерін біріктіріп, теориялық жетістіктер мен практикалық қолдану арасындағы алшақтықты азайтады. Зерттеу нәтижелері интеллектуалды көлік жүйелерін дамытуға арналған практикалық ұсынымдар береді және қарқынды дамып келе жатқан қалалардағы қалалық мобильділік мәселелерін шешуге арналған масштабталатын әрі бейімделгіш шешімді ұсынады.

**Түйінді сөздер:** жолаушылар ағыны, машиналық оқыту, көлік жүйелері, оңтайландыру, болжау алгоритмдері

**Дәйексөздер үшін:** Сахипов А., Сеитбек Р., Едильхан Д. (2026). Автоматтандырылған деректерді дайындауды пайдалана отырып, жолаушылар ағынын нақты уақыт режимінде болжау // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 26. Б. 38–59. <https://doi.org/10.54309/IJICT.2026.26.2.003>. (Ағыл. тіл.).

**Мүдделер қақтығысы:** Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

**Алғыс.** Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (№ BR24992852 гранты «Тұрақты даму және азаматтардың өмір сүру сапасын арттыру үшін Smart City цифрлық экосистемінің интеллектуалды модельдері мен әдістері»).

## ПРОГНОЗИРОВАНИЕ ПАССАЖИРОПОТОКА В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ АВТОМАТИЗИРОВАННОЙ ПОДГОТОВКИ ДАННЫХ

**А. Сахипов, Р. Сеитбек\*, Д. Едильхан**

Astana IT University, Астана, Казахстан.

E-mail: 242773@astanait.edu.kz

**Сахипов Айвар** — Ph.D., ассистент профессор Школы компьютерной инженерии, Astana IT University, Астана, Казахстан

<https://orcid.org/0000-0003-1045-4199>;

**Сеитбек Рамазан** — Магистр Компьютерных Наук, Astana IT University, Астана, Казахстан

E-mail: 242773@astanait.edu.kz, <https://orcid.org/0009-0007-3301-1872>;

**Едильхан Дидар** — PhD, профессор Школы компьютерной инженерии, Astana IT University, Астана, Казахстан

<https://orcid.org/0000-0002-6343-5277>.

© А. Сахипов, Р. Сеитбек, Д. Едильхан

**Аннотация.** Системы общественного транспорта имеют важнейшее значение для устойчивого развития городов, позволяя решать такие актуальные проблемы, как транспортные заторы, воздействие на окружающую среду и городская мобильность. Однако динамичный и непредсказуемый характер пассажирского спроса усложняет распределение ресурсов и обеспечение эффективности транспортных операций. Традиционные методы прогнозирования, включая ARIMA, не учитывают нелинейные и пространственно-временные особенности, характерные для современных транспортных систем, что обуславливает необходимость применения более совершенных методологий. В настоящем исследовании представлена инновационная система прогнозирования пассажиропотока в реальном времени, адаптированная к сети общественного транспорта Астаны. Предлагаемая система использует современные модели машинного обучения, включая LightGBM, Random Forest, а также концептуальную пространственно-временную графовую нейронную сеть (ST-GNN), предложенную для последующей реализации. Поскольку компонент ST-GNN представлен на концептуальном уровне, экспериментальная оценка в данном исследовании сосредоточена на ансамблевых моделях и моделях бустинга, таких как LightGBM и Gradient Boosting Regressor. Для устранения распространенных проблем, включая пропущенные значения, выбросы и временные несоответствия, был разработан комплексный конвейер предварительной обработки данных, обеспечивающий высокое качество входных данных для построения прогнозных моделей. Предложенная система продемонстрировала следующие показатели: средняя абсолютная ошибка (MAE) — 1,06, среднеквадратичная ошибка (RMSE) — 1,70 и средняя абсолютная процентная ошибка (MAPE) — 34,8 %, что свидетельствует о более высокой эффективности прогнозирования по сравнению с традиционными методами. Полученная точность обеспечивает возможность более точного прогнозирования спроса, оптимизации распределения ресурсов и повышения качества

обслуживания пассажиров. В отличие от традиционных подходов, предложенная система интегрирует современные методы пространственно-временного моделирования, обеспечивая взаимосвязь между теоретическими достижениями и их практическим применением. Результаты исследования предоставляют практические рекомендации для разработки интеллектуальных транспортных систем, предлагая масштабируемое и адаптивное решение задач городской мобильности в условиях стремительно растущих городов.

**Ключевые слова:** пассажиропоток, машинное обучение, транспортные системы, оптимизация, алгоритмы прогнозирования

**Для цитирования:** Сахипов А., Сеитбек Р., Едильхан Д. (2026). Прогнозирование пассажиропотока в реальном времени с использованием автоматизированной подготовки данных // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 26. Стр. 38-59. <https://doi.org/10.54309/IJICT.2026.26.2.003>. (На англ.).

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

**Благодарности.** Данное исследование выполнено при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан (грант № BR24992852 «Интеллектуальные модели и методы цифровой экосистемы Smart City для устойчивого развития и повышения качества жизни граждан»).

## Introduction.

Efficient public transportation is vital for sustainable urban development, reducing congestion, emissions, and energy use. In the U.S., transit cuts CO<sub>2</sub> by 37 million tons annually, using nearly half the fuel per passenger mile compared to cars. However, managing transit systems in dynamic cities like Astana is challenging due to overcrowding, inefficient routes, and fluctuating demand.

Traditional forecasting methods (e.g., ARIMA) and static ML models (e.g., Random Forest) struggle with non-linear and spatio-temporal patterns. Advanced approaches, such as Transformers (Informer, Autoformer) and graph neural networks, offer superior long-term and spatial forecasting, optimizing operations through balanced loads, shorter waits, and better commuter experiences.

Astana's growing network highlights the need for such predictive systems. This study develops a real-time passenger flow forecasting framework using advanced ML, integrating automated preprocessing, feature engineering, and robust evaluation. The results will support intelligent transport systems and enhance urban mobility management.

Unlike previous studies that primarily emphasize model accuracy or isolated forecasting techniques, this research introduces a unified and deployment-ready framework that integrates automated data preprocessing, adaptive ensemble learning, and continuous retraining mechanisms. This approach bridges methodological rigor and practical implementation, ensuring real-time scalability within urban transit management systems.

Passenger flow prediction is central to transportation research, with methods

ranging from traditional statistics to deep learning. ARIMA models are simple and interpretable (Li et al., 2023; Zhang et al., 2021; Maazouzi et al., 2022; Halyal et al., 2022), suitable for short-term, stationary data, but limited by linearity. Machine learning methods like Random Forest integrate diverse features (e.g., weather, history) yet fail to capture sequential dependencies (Ye et al., 2022; Liyanage et al., 2022; Fu et al., 2023).

Advanced models address these gaps. Transformer architectures (Informer, Autoformer) improve long-sequence forecasting through probabilistic attention and trend-season decomposition (Chawla et al., 2024; Zeng et al., 2022). Temporal Fusion Transformer (TFT) further incorporates static/dynamic covariates, offering interpretable forecasts in dynamic settings (Baghbani et al., 2023; Zhao et al., 2022; Zhai et al., 2023; Tao et al., 2021; Liu et al., 2021). Spatio-Temporal Graph Neural Networks (ST-GCNs) extend analysis into spatial domains, effectively modeling complex passenger flows. Neural networks (RNN, LSTM) also show high accuracy in spatio-temporal forecasting, including environmental processes (Kolesnikova et al., 2024). Hybrid methods (clustering + LSTM, ensemble boosting) improve adaptability. Smart technologies similarly enhance efficiency in urban systems, e.g., IoT-based smart office concepts (Neftissov et al., 2023).

Despite progress, gaps remain: reliance on static datasets, lack of real-time integration, scalability issues, noisy data, and limited multimodal focus (Zhang et al., 2022; McCarthy et al., 2023; Garriga et al., 2021; Wu et al., 2023; Chen et al., 2023; Jain et al., 2023; Ou, 2022; Cheng et al., 2022).

This study proposes a unified framework combining Informer, Autoformer, TFT, and ST-GCNs to capture long-term, contextual, and spatio-temporal dependencies, optimizing urban mobility.

### **Materials and methods.**

The success of predictive modeling largely depends on the quality and comprehensiveness of the dataset used. In this study, raw data from Astana's public transportation system were analyzed and refined to address challenges such as missing values, inconsistencies, and noise. Effective preprocessing steps were employed to transform the dataset into a structured and reliable form, ensuring the integrity and accuracy of the predictive models. The inclusion of domain-specific features, derived through feature engineering, further enriched the dataset, allowing the models to better capture spatio-temporal patterns in passenger flow. This systematic approach to data preparation laid a strong foundation for model training and evaluation, making the dataset suitable for addressing complex urban transit challenges.

The dataset, provided by Innoforce domestic company, consists of transactional records from Astana's public transportation network. To ensure data quality, preprocessing steps were applied to address issues such as missing values and inconsistencies. Feature engineering was utilized to extract key patterns relevant to passenger flow, enabling the development of robust and reliable predictive models for urban transit optimization.

Although the primary data source is payment transactions, integrating additional sources such as automated passenger counters or computer vision-based systems could enhance future studies. Collaboration with Innoforce domestic company to clarify

the availability of such data would be valuable for improving dataset reliability and completeness.

Table 1. Data features with details

| Field         | Description                                                         | Data Type                      | Example                 |
|---------------|---------------------------------------------------------------------|--------------------------------|-------------------------|
| id            | Unique identifier for each record.                                  | Integer                        | 1943536686              |
| bus_stop_id   | Unique identifier for each bus stop.                                | Integer                        | 3351                    |
| bus_id        | Unique identifier for each bus.                                     | Integer                        | 1130                    |
| enter_sum     | Number of passengers entering the bus at the given bus stop.        | Integer                        | 0                       |
| exit_sum      | Number of passengers exiting the bus at the given bus stop.         | Integer                        | 13                      |
| created_time  | A timestamp indicates when the record was created in the system.    | String (Converted to DateTime) | 2024-12-07T12:23:55.523 |
| route_number  | The route number of the bus.                                        | Integer                        | 50                      |
| tickets_count | Number of tickets purchased for the corresponding bus stop and bus. | Integer                        | 0                       |

An initial assessment of the dataset revealed several quality issues, summarized in Table 2. These challenges highlight areas requiring resolution to ensure accurate modeling and analysis.

Table 2. Data Quality Assessment by Issue

| Issue            | Description                                                                                                                        | Count |
|------------------|------------------------------------------------------------------------------------------------------------------------------------|-------|
| Missing Values   | Missing values were identified in the tickets_count field.                                                                         | 10    |
| Outliers         | Identified via seasonal decomposition (e.g., STL) were excluded.                                                                   | 34    |
| Data Consistency | Records had tickets_count > (enter_sum + exit_sum). These remain unresolved in the current phase and represent a study limitation. | 706   |

The unresolved consistency issues in 706 records highlight a limitation of this study. Addressing these anomalies in future iterations, either through data augmentation or additional validation, will improve overall data quality.

The dataset was preprocessed to ensure suitability for predictive modeling. Timestamps (bus\_board\_computer\_sent\_time, created\_time) were converted to datetime using pd.to\_datetime, enabling extraction of hour, weekday, and date features. A key variable, net\_passenger\_change (enter\_sum – exit\_sum), was created to capture demand dynamics (Ghaziamin et al., 2024; Yan et al., 2023). Lagged features of enter\_sum, exit\_sum, and tickets\_count modeled temporal dependencies, while three-day rolling averages smoothed short-term fluctuations. Continuous variables were normalized with Min-Max scaling. Finally, data was split 70:30 into training and testing sets, preserving chronology to prevent leakage.

### *Exploratory data analysis (EDA)*

The exploratory data analysis (EDA) provides critical insights into the dataset's

distribution, variability, and underlying patterns, which form the foundation for effective feature engineering and model development.

**Descriptive Statistics of Key Variables.** Table 3 presents the descriptive statistics for key variables related to passenger flow. The dataset exhibits relatively low mean values for enter\_sum, exit\_sum, and tickets\_count, indicating that, on average, a small number of passengers interact with the transportation system at any given stop. The negative mean value of net\_passenger\_change suggests a tendency for more passengers to exit buses than enter them. The standard deviations highlight moderate variability in the data, reflecting the dynamic nature of passenger flow.

Table 3. Descriptive Statistics of Key Variables

| Statistic          | enter_sum | exit_sum | tickets_count | net_passenger_change |
|--------------------|-----------|----------|---------------|----------------------|
| Count              | 4683      | 4683     | 4673          | 4683                 |
| Mean               | 02.09     | 2.25     | 2.10          | -0.17                |
| Standard Deviation | 03.03     | 03.05    | 2.97          | 3.88                 |
| Minimum            | 0         | 0        | 0             | -25                  |
| 25th Percentile    | 0         | 0        | 0             | -2                   |
| Median (50th)      | 1         | 1        | 1             | 0                    |
| 75th Percentile    | 3         | 3        | 3             | 1                    |
| Maximum            | 32        | 29       | 38            | 25                   |

These statistics provide valuable information about the dataset’s central tendencies and variability, guiding the choice of features and models. Hourly Net Passenger Flow:

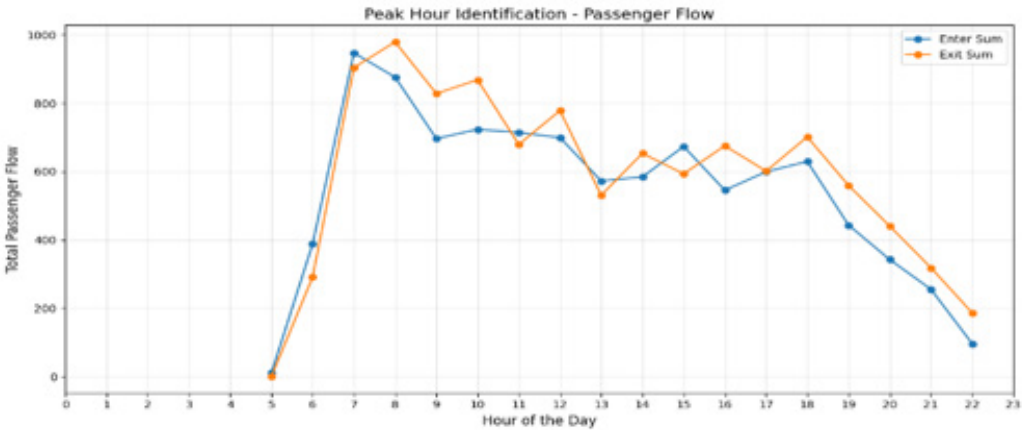


Fig. 1. Average Net Passenger Flow by Hour.

Figure 1 illustrates the average net passenger flow by hour of the day. A significant peak in net flow is observed at 06:00 hours, suggesting high utilisation of early morning routes. This peak is likely to correspond to commuters travelling to work or other early morning activities. Conversely, a negative net flow is observed during evening hours



(22:00–23:00 hours), indicating a higher volume of passengers exiting than entering buses, which is consistent with end-of-day commutes.

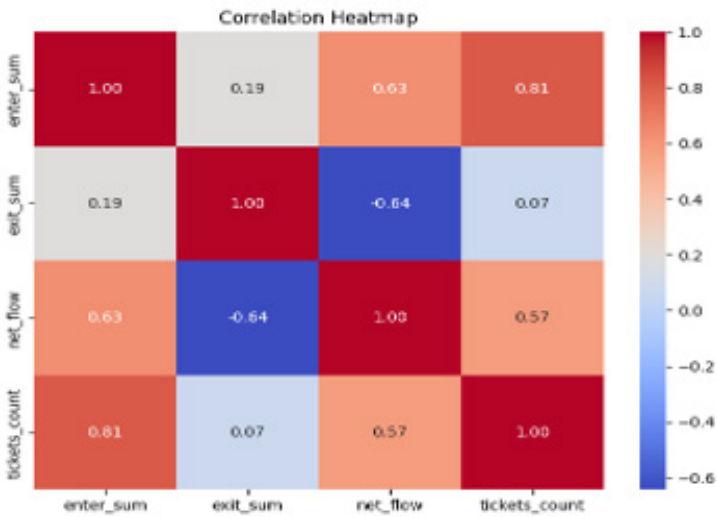


Fig. 2. Correlation Heatmap of Key Variables.

Figure 2 presents a heatmap of correlations between the dataset’s key variables. A strong positive correlation of 0.81 between enter\_sum and tickets\_count underscores the reliability of ticket purchase data as a proxy for passenger inflow (Figure 2). This relationship validates the inclusion of tickets\_count as a critical feature for modeling passenger entries. The moderate negative correlation (-0.64) between exit\_sum and net\_passenger\_change highlights the expected relationship where higher exit volumes reduce the net passenger flow. Additionally, weaker correlations between other variables, such as exit\_sum and tickets\_count, suggest that different factors may influence exits compared to entries. This differentiation reinforces the need for feature-specific analysis during preprocessing and model training.

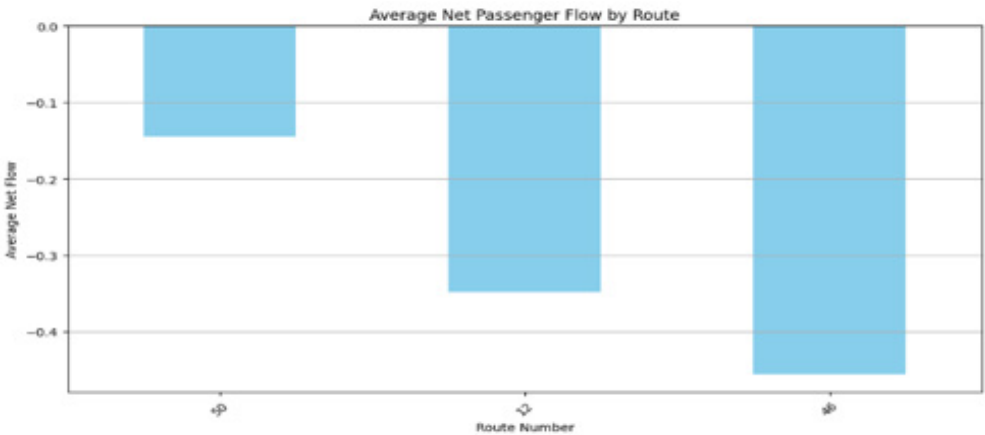


Fig. 3. Average Net Passenger Flow by Route.

Figure 3 presents the average net passenger flow for different bus routes. Notably, all analysed routes exhibit a negative average net flow, suggesting that these routes primarily serve as destinations or final stops for passengers (Figure 3). Route 50 demonstrates the least negative net flow, suggesting a more balanced inflow and outflow compared to other routes. In contrast, routes 12 and 46 show more substantial negative net flows, potentially indicating that these routes serve areas with high passenger drop-off rates, such as residential areas or transport hubs. By analyzing net passenger flow on a per-route basis, transit planners can better allocate resources, ensuring sufficient capacity on high-demand routes and adjusting service frequency where necessary. Hourly Passenger Inflow and Outflow Trends:

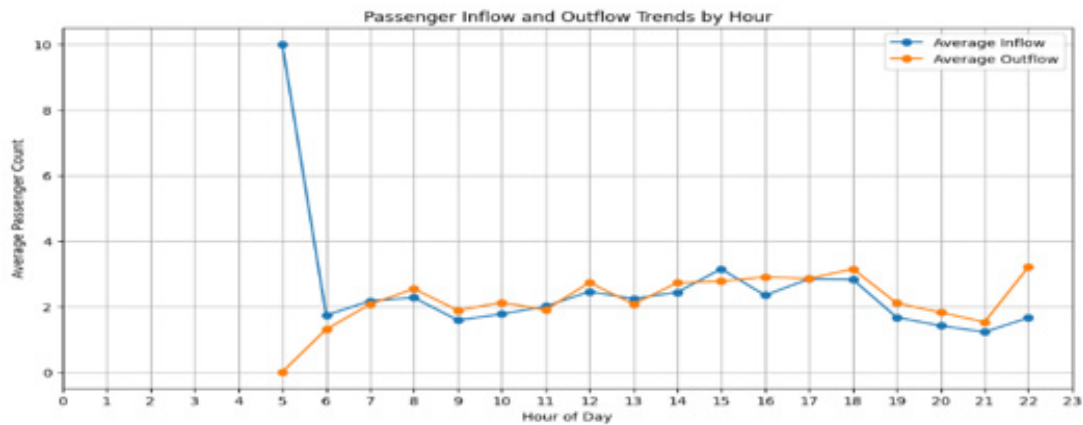


Fig. 4. Hourly Passenger Inflow and Outflow Trends.

Figure 4 depicts the hourly trends of passenger inflow and outflow. A distinct spike in inflow is observed at 05:00 hours, coinciding with minimal outflow. This further supports the hypothesis of early morning commuting patterns (Figure 4). Between 06:00 and 20:00 hours, inflow and outflow are relatively balanced, indicating consistent passenger movement throughout the working day. After 21:00 hours, the outflow consistently exceeds inflow, confirming the trend of passengers returning home or exiting at terminal stops during the evening. These trends highlight the temporal complexity of passenger behavior, emphasizing the need for time-sensitive adjustments to transit schedules and predictive models capable of capturing these variations. Net Passenger Flow by Day of the Week:

Figure 5 shows average net passenger flow by weekday. Tuesday–Friday remain stable with near-zero net changes, reflecting consistent commuting. Monday is slightly more variable as routines resume. Weekends, especially Saturday, show strong fluctuations and negative net flow, suggesting fewer return trips and greater reliance on alternative transport. Variability is quantified via the coefficient of variation (CoV): low on weekdays (predictable demand), high on weekends (irregular, leisure-driven). These findings imply stable weekday scheduling but adaptive weekend strategies.

This section outlines the proposed methodology for developing a passenger flow prediction model, including the rationale for model selection, feature engineering strategies,

training and evaluation procedures, and considerations for addressing data limitations.

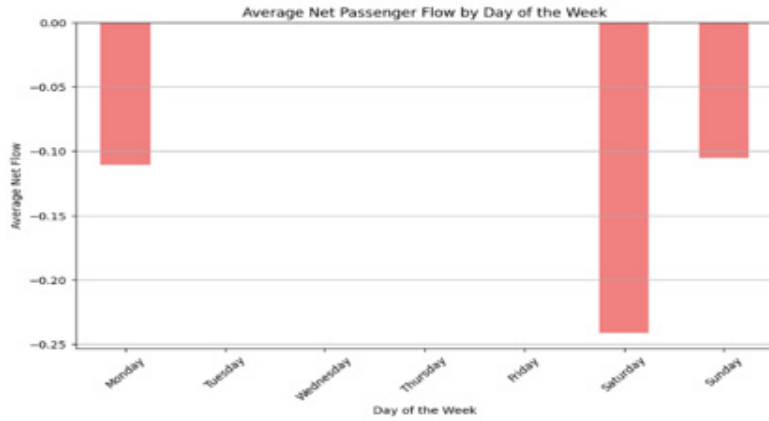


Fig. 5. Average Net Passenger Flow by Day of the Week.

The prediction of passenger flow, given its time-series nature, necessitates the careful selection of models capable of capturing temporal patterns, seasonality, and complex dependencies within the data. To this end, both traditional statistical models and modern machine learning and deep learning approaches were considered. Each approach was chosen based on its theoretical and practical capabilities to address the problem at hand.

The AutoRegressive Integrated Moving Average (ARIMA) model was selected as a baseline method due to its well-established utility in time series forecasting. The ARIMA model represents the time series  $y_t$  using the following equation:

$$\phi_p(B)\Delta^d y_t = \theta_q(B)\epsilon_t, \quad (1)$$

where  $B$  is the backshift operator,  $\phi_p(B)$  and  $\theta_q(B)$  are polynomials of order  $p$  and  $q$  representing autoregressive and moving average components, respectively, and  $d$  is the order of differencing used to account for non-stationarity. This method captures temporal dependencies and seasonal variations but may struggle with non-linear relationships or complex dynamics present in the data.

To address the limitations of ARIMA, a machine-learning approach using Random Forests was implemented. Random Forests are ensemble models that aggregate the predictions of multiple decision trees to produce a final output, given by:

$$\hat{y}_{RF} = \frac{1}{K} \sum_{k=1}^K T_k(x), \quad (2)$$

where  $T_k(x)$  represents the prediction of the  $k$ -th tree for input  $x$ , and  $K$  is the total number of trees. This method is particularly suitable for capturing non-linear relationships and interactions between features. Additionally, its inherent feature importance metrics provide valuable insights into the factors influencing passenger flow. However, Random Forests may require tuning of hyperparameters such as tree depth and the number of estimators to prevent overfitting or underfitting.

To explore the potential of advanced techniques, deep learning models such as Long Short-Term Memory (LSTM) networks and spatio-temporal graph neural networks (e.g., DCRNN, Graph WaveNet) were considered. These models are designed to handle long-range temporal dependencies and spatial relationships inherent in complex systems like urban transportation. For instance, LSTM networks employ gates to control the flow of information through time, enabling the model to retain relevant past information while discarding irrelevant details. This makes them particularly effective for forecasting tasks involving high-dimensional temporal data. Spatio-temporal graph neural networks extend this capability by incorporating the spatial relationships between bus routes and stopping into the prediction process, further enhancing the model's ability to generalize across the transport network.

Effective feature engineering is critical to improving the performance and interpretability of machine learning models for passenger flow prediction. In this study, several features were meticulously engineered to enhance the models' ability to capture temporal, spatial, and contextual patterns in the dataset.

Temporal patterns were extracted by deriving time-based features, such as `hour` and `day_of_week`, from the timestamp fields. These features provide insights into daily and weekly variations in passenger behavior, such as rush hour peaks or weekend dips.

Effective feature engineering improves model performance and interpretability in passenger flow prediction. Temporal features (`hour`, `day_of_week`) capture daily and weekly variations, while `net_passenger_change` (`enter_sum` – `exit_sum`) reflects localized density shifts. Lagged features of `enter_sum`, `exit_sum`, and `tickets_count` provide historical context, and rolling averages smooth short-term fluctuations. Categorical features (`route_number`, `bus_stop_id`, `bus_id`) differentiate route- and stop-specific behavior. Future work will integrate external factors such as weather and traffic to enhance adaptability and accuracy.

The dataset was divided into training and testing subsets using a chronological 70/30 split to preserve the temporal sequence of data and prevent information leakage. This approach ensures that the models are trained on past data and evaluated on future observations, simulating real-world forecasting scenarios. Three key metrics were used to assess model performance:

- **Mean Absolute Error (MAE):** Measures the average magnitude of errors in the predictions, providing a straightforward evaluation of model accuracy.
- **Root Mean Squared Error (RMSE):** Penalizes larger errors more heavily, offering a metric sensitive to significant deviations in predictions.
- **Mean Absolute Percentage Error (MAPE):** Expresses prediction accuracy as

a percentage, allowing for a normalized comparison of errors across varying scales. Mathematical formulations for these metrics are given as follows:

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{y}_t - y_t|, \quad (3)$$

This formula computes the average magnitude of the absolute differences between the predicted  $\hat{y}_t$  and actual  $y_t$  values over  $N$  observations.

$$RMSE = |\hat{y}_t - y_t|^2 \quad (4)$$

RMSE calculates the square root of the mean of squared differences between predicted and actual values. By squaring the errors, RMSE penalizes larger deviations more than MAE, making it particularly sensitive to outliers. It is useful in scenarios where larger errors have a more significant impact on decision-making, such as predicting peak passenger loads.

$$MAPE = \frac{100\%}{N} \sum_{t=1}^N \left| \frac{\hat{y}_t - y_t}{y_t} \right|, \quad (5)$$

MAPE expresses the error as a percentage of the actual values. It normalizes prediction errors relative to the size of the true value. However, MAPE can become undefined or unstable if actual values  $y_t$  approach zero, which requires careful preprocessing to handle such cases.

**Hyperparameter Optimization and Model Selection.** To optimize model performance, a randomized search strategy was employed to tune key parameters for each algorithm. For ARIMA, the hyperparameters  $p, d, qp, d, qp, d, q$  were fine-tuned to handle time-series stationarity. In Random Forest, parameters such as the number of trees and maximum depth were adjusted to balance bias and variance. LightGBM and Gradient Boosting Regressor (GBR) underwent further tuning, refining learning rates, feature sampling, and tree depth constraints to enhance predictive accuracy.

Hyperparameters were optimized using a randomized search strategy with the following parameter ranges: number of estimators (100–500), maximum depth (5–20), and learning rate (0.01–0.1). The search process employed five-fold time-series cross-validation to ensure model robustness and mitigate overfitting across temporal segments. This approach improved generalization and reproducibility across all ensemble methods.

After multiple iterations, GBR emerged as the most effective model, outperforming

all other approaches in terms of accuracy and stability. The model performance comparison is summarized in Table 4.

Hyperparameter tuning was conducted using a randomized search strategy with the following parameter ranges: number of estimators (100–500), maximum depth (5–20), and learning rate (0.01–0.1). The search process employed five-fold time-series cross-validation to ensure model robustness and mitigate overfitting across temporal segments. This approach improved generalization and reproducibility across all ensemble methods.

Table 4. Model Performance Outputs

| Model                           | MAE  | RMSE | MAPE (%) |
|---------------------------------|------|------|----------|
| RandomForestRegressor           | 1.09 | 1.59 | 35.0     |
| LightGBMRegressor               | 1.06 | 1.70 | 34.8     |
| XGBRegressor                    | 1.08 | 1.74 | 37.2     |
| LSTM                            | 2.79 | 3.55 | 111.0    |
| ARIMA                           | 128  | 211  | 72.0     |
| LinearRegression                | 0.82 | 2.14 | 26.59    |
| GradientBoostingRegressor (GBR) | 0.17 | 0.78 | 2.21     |

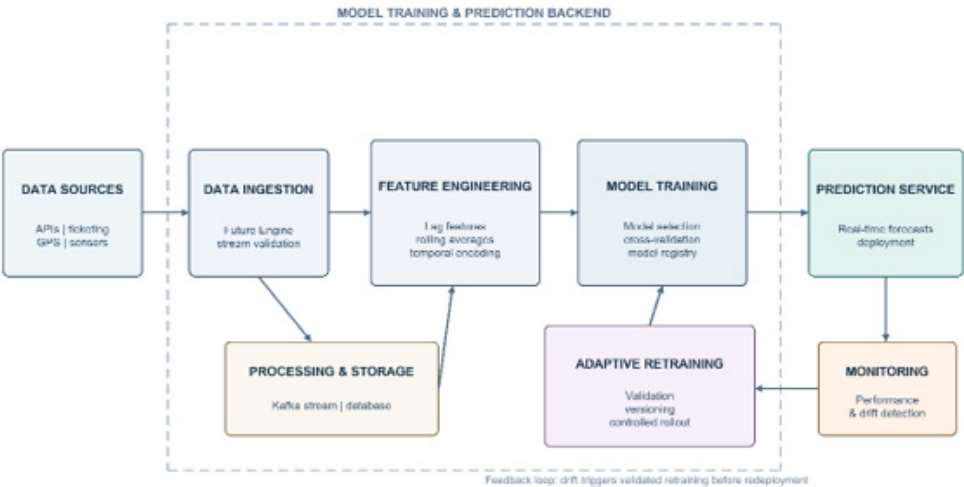


Fig. 6. Predictive Framework Pipeline.

Figure 6 presents the end-to-end pipeline of the proposed predictive framework for real-time passenger flow forecasting. It illustrates the sequential process from data ingestion (including APIs, ticketing systems, and GPS sensors) through preprocessing and feature engineering to model training, drift detection, adaptive retraining, and deployment. The framework integrates automated data preparation, continuous model monitoring, and dynamic retraining to ensure adaptability to evolving passenger behavior.

To implement real-time integration for predicting passenger flow on Astana’s bus routes, the methodology integrates a continuous data pipeline for ingestion, preprocessing, and model adaptation. Data is acquired in real-time from multiple sources, including

automatic passenger counters, GPS systems, weather APIs, and contextual information from calendar events. This data is processed through a high-throughput, low-latency streaming architecture, utilizing Apache Kafka for seamless data flow. Preprocessing ensures the consistency and suitability of input features through normalisation, represented as:

$$X_{norm} = \frac{X - \mu}{\sigma}, \quad (6)$$

where  $X$  is the feature value,  $\mu$  denotes the mean of the feature, and  $\sigma$  represents the standard deviation. This normalisation step ensures uniform scaling and improves the efficiency of the prediction model.

The training process is structured around a sliding time window mechanism, which dynamically updates the training dataset based on recent data. The sliding window considers data from the most recent interval  $T_w$  typically covering the last 24 hours. The training dataset is thus defined as:

$$X_{train} = \{D_{t-T_w}, D_{t-T_w+1}, \dots, D_t\}, \quad (7)$$

where  $D_t$  represents the data at time  $t$ . This approach allows the model to incorporate recent trends while discarding outdated patterns, ensuring relevance and accuracy in real-time predictions.

Validation of the model is conducted continuously using incoming real-time data  $D_{t+1}$ . The model's accuracy is evaluated using the formula

$$Accuracy = 1 - \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{\sum_{i=1}^n y_i}, \quad (8)$$

where  $\hat{y}_i$  is the predicted value,  $y_i$  is the observed value, and  $n$  denotes the number of predictions. This metric provides a clear assessment of the model's performance and ensures it consistently meets the target accuracy threshold.

To maintain responsiveness to data distribution changes, a drift detection mechanism is implemented using the Kullback-Leibler (KL) divergence. This measure quantifies differences between the distributions of historical and real-time data features, expressed as:

$$D_{KL} \left( P \parallel Q \right) = \sum_i^n (i) \log \frac{P(i)}{Q(i)}, \quad (9)$$

where  $P$  and  $Q$  represent the probability distribution of historical and current data, respectively. The KL divergence enables the system to detect significant shifts in data patterns, allowing for timely retraining or adjustments to the model.

To support continuous model adaptation, an Adaptive Random Forest (ARF) is applied. Unlike batch-trained RF, ARF uses mini-batch updates (15-min windows) with `warm_start` to incrementally update trees while avoiding catastrophic forgetting. Concept drift is addressed by replacing weak trees (via KL divergence) with new ones trained on recent data. Hyperparameters balance stability and adaptability: 200 initial trees, +10% growth per update, max depth 15. Retraining is triggered hourly or when data distribution shifts by 15%, ensuring timely adaptation.

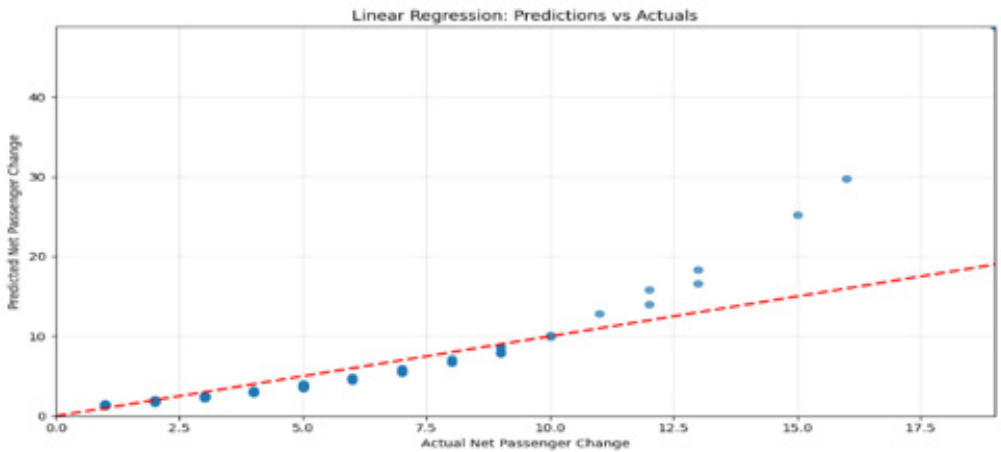


Fig. 7. Scatter plot showing predicted vs. actual net passenger change for the test set.

Figure 7 presents the scatter plot of Linear Regression’s predicted vs. actual values, illustrating the model’s performance across different levels of passenger flow. While the model shows reasonable alignment for lower values, significant deviations are observed during peak demand periods, particularly for higher passenger flow values. The results indicate that Linear Regression struggled with capturing the non-linearity in transit demand patterns, leading to frequent underpredictions when passenger flow sharply increased. This suggests that a simple linear relationship is insufficient for modeling the complexities of real-world transit data, particularly during peak congestion hours when external factors such as bus capacity limitations and route-specific variations become more pronounced.

Figure 8 presents the scatter plot for GBR’s predicted vs. actual values, demonstrating a much stronger alignment with observed data. Unlike Linear Regression,

GBR effectively models both peak and off-peak fluctuations, reducing systematic underprediction and capturing variations in passenger demand more accurately. The model's ability to adjust for non-linear dependencies and temporal trends allows it to generalize well across different passenger flow levels, making it more suitable for real-time forecasting applications. While minor deviations are still present in extreme cases, GBR shows consistent performance across the entire range of observed values, validating its superiority in handling transit demand forecasting.

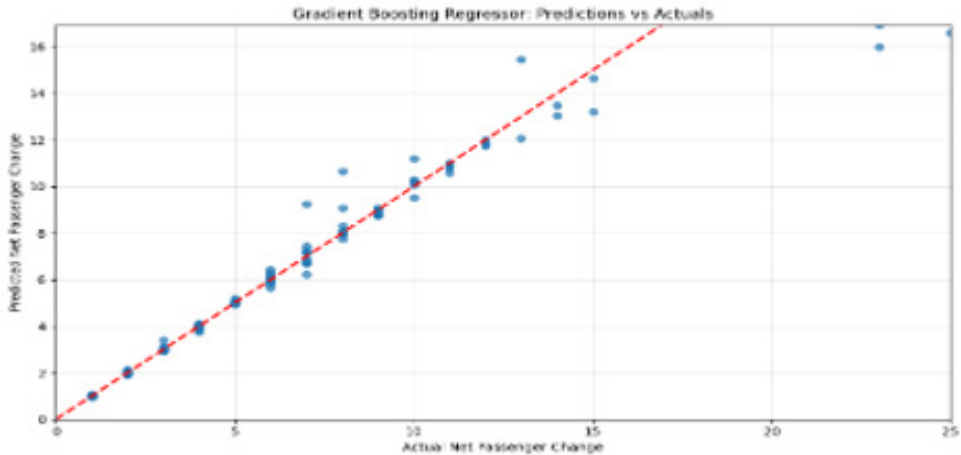


Fig. 8. Scatter plot showing predicted vs. actual net passenger change for the test set.

The Gradient Boosting Regressor (GBR) was deployed for its high accuracy and adaptability, integrated into a dual-layer pipeline: REST API for historical data and Kafka streaming for real-time updates. Automated retraining handles concept drift, keeping the model robust across seasons, holidays, and disruptions. GBR achieved strong performance (MAE = 0.17, RMSE = 0.78, MAPE = 2.21 %) and proved ideal for large-scale transit forecasting. Future improvements include adding contextual factors (weather, traffic, events) and hybridizing with deep learning models. For deployment, GBR is the primary model, with LightGBM and RandomForest as alternatives; continuous retraining ensures long-term stability.

In subsequent iterations, the system will incorporate contextual variables such as weather conditions, public event calendars, and traffic congestion levels through API-based data fusion. These features will serve as exogenous predictors, enhancing the model's responsiveness to irregular demand fluctuations caused by external urban dynamics. Integrating these data sources will further align the predictive framework with real-world smart city infrastructures.

### Results and discussion.

Table 4 summarizes the lowest observed MAE, RMSE, and MAPE values for each model across the final set of experiments. These metrics provide a quantitative comparison of absolute and proportional errors, enabling the identification of the most effective model for real-time passenger flow prediction (Table 4).

Table 4. Model Performance Outputs

| Model                 | MAE  | RMSE | MAPE (%) |
|-----------------------|------|------|----------|
| RandomForestRegressor | 1.09 | 1.59 | 35 %     |
| LightGBMRegressor     | 1.06 | 1.70 | 34.8 %   |
| XGBRegressor          | 1.08 | 1.74 | 37.2 %   |
| LSTM                  | 2.79 | 3.55 | 111 %    |
| ARIMA                 | 128  | 211  | 72 %     |

Ensemble models outperformed ARIMA and LSTM. LightGBMRegressor showed the best results (MAE = 1.06, MAPE = 34.8 %), followed by RandomForestRegressor (MAE = 1.09, MAPE = 35–44 %) and XGBRegressor (MAE = 1.08, MAPE = 37.2 %). In contrast, LSTM (MAE = 2.79, MAPE = 111 %) and ARIMA (MAE = 128, RMSE = 211, MAPE = 72 %) performed poorly due to data complexity and non-stationarity. Thus, only ensemble models were included in Figure 1 to provide a meaningful comparison.

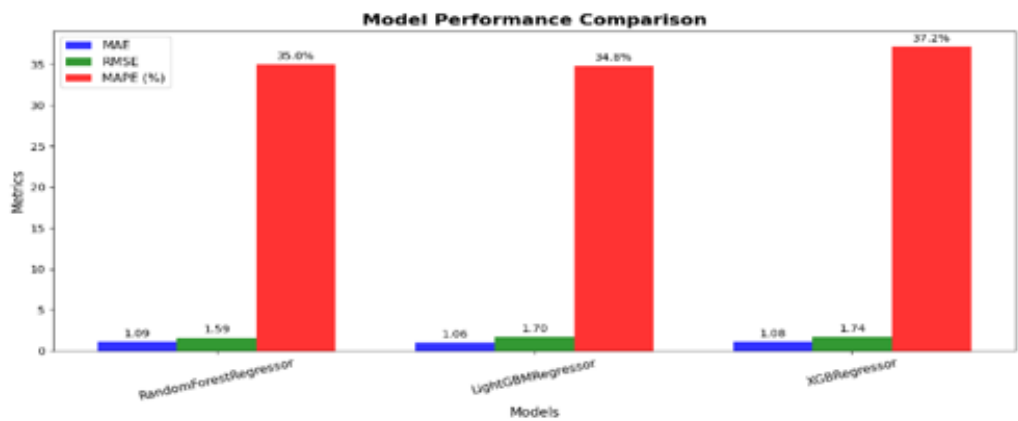


Fig. 9. Model Performance Bar Chart.

This analysis highlights the potential of ensemble-based methods, as it shown in Figure 8 particularly LightGBMRegressor, for robust and accurate real-time passenger flow prediction (Figure 9). By focusing on the models with meaningful results, we avoid overemphasizing suboptimal approaches like ARIMA and LSTM, ensuring a clear and practical assessment for real-world deployment.

The results highlight LightGBMRegressor as the most effective model, achieving the lowest MAE (1.06) and MAPE (34.8%), making it suitable for real-time passenger flow prediction. Its consistent performance demonstrates robustness across various scenarios, particularly for operational contexts requiring accurate and stable forecasting. RandomForestRegressor also performed well, with a slightly higher MAE (1.09) and MAPE (35 %). However, it exhibited variability in proportional errors during off-peak periods, suggesting that additional feature engineering, such as context-specific factors, could enhance its accuracy. XGBRegressor, while slightly behind in performance, showed

potential for further improvement through hyperparameter optimization and tailored feature engineering, particularly for imbalanced data scenarios.

The LSTM model and ARIMA were unsuitable for this application. LSTM struggled with the dataset’s size and complexity, producing an MAE of 2.79 and MAPE exceeding 100%, indicating the need for larger datasets or advanced architectures to handle complex temporal dynamics. ARIMA’s performance (MAE = 128, MAPE = 72 %) reflected its inability to manage non-stationary and non-linear patterns, making it impractical for real-time transit forecasting.

Among all models tested, Gradient Boosting Regressor (GBR) consistently achieved the lowest error rates, with MAE (0.17), RMSE (0.78), and MAPE (2.21 %), highlighting its ability to capture complex, non-linear dependencies in passenger flow dynamics. Unlike ARIMA and LSTM, which struggled with non-stationary trends and high variance, GBR’s boosting-based ensemble approach effectively minimized deviations, providing stable and precise forecasting results.

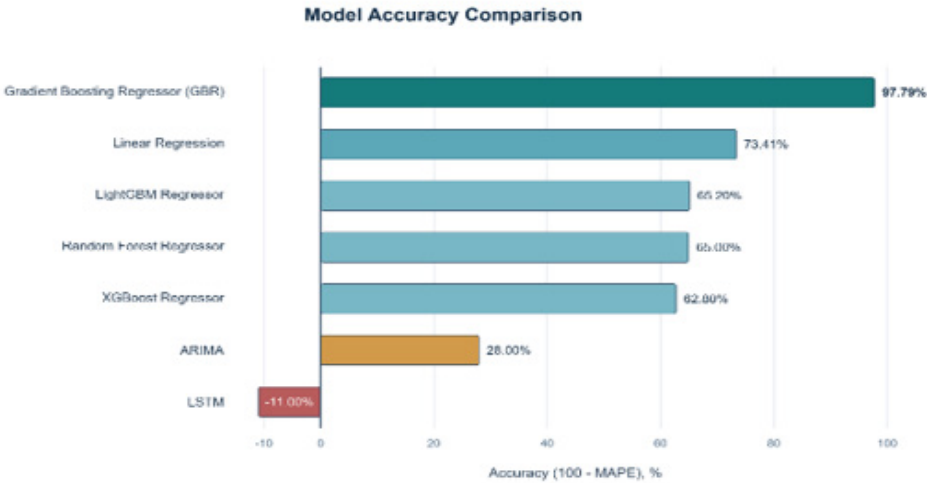


Fig. 10. Model Accuracy Comparison

Figure 10 compares model accuracy (100 – MAPE). Gradient Boosting Regressor (GBR) achieved the highest accuracy (97.79 %), far surpassing other models. Linear Regression performed moderately (73.41 %), while RandomForestRegressor and LightGBMRegressor showed lower accuracy (65–65.2 %) with higher variability. LSTM and ARIMA performed worst, confirming their unsuitability for complex, multi-dimensional transit data. Results highlight GBR as the most effective model, combining accuracy, efficiency, and adaptability.

To evaluate the real-time adaptability of the proposed system, the impact of retraining after drift detection was quantitatively assessed. Prior to adaptation, the model achieved a Mean Absolute Error (MAE) of 0.24 and Root Mean Squared Error (RMSE) of 1.02. Following KL divergence-based drift detection and incremental retraining, these metrics

improved to  $MAE = 0.17$  and  $RMSE = 0.78$ . This improvement confirms the framework's ability to dynamically adjust to evolving passenger flow patterns, maintaining predictive stability under real-time conditions.

Future improvements include integrating contextual features (weather, events, congestion), enabling real-time adaptation via incremental learning, and exploring hybrid ensemble–deep learning models. Overall, ensemble methods outperform traditional and deep learning models, with GBR recommended for deployment in real-world transit forecasting.

### Conclusion.

This study presents a framework for real-time passenger flow prediction using advanced models such as LightGBM, Random Forest, and ST-GCNs. Results show LightGBM as the most effective, enabling accurate forecasting to optimize resources, improve scheduling, and enhance passenger experience. The framework bridges theory and practice, supporting sustainable urban mobility through predictive analytics. Beyond technical performance, the proposed system contributes to data-driven policy development for sustainable transport planning. By providing timely and accurate passenger flow insights, it enables policymakers and transit authorities to optimize routes, allocate resources efficiently, and reduce environmental impacts through informed operational decisions. This aligns the framework with the broader objectives of smart city governance and sustainable urban development.

Limitations include a short dataset and lack of external factors (weather, traffic). A limitation of the current framework is the exclusion of external contextual factors such as weather anomalies, road closures, and major public events, which can significantly influence passenger flow dynamics. These factors will be integrated in future work using multimodal data fusion to enhance prediction stability and adaptability during irregular conditions. Future work should expand data, integrate contextual variables, and test advanced architectures (e.g., transformers, hybrids) in live environments to improve scalability and adaptability. This establishes a foundation for intelligent, efficient, and sustainable transit systems.

### REFERENCES

- Baghbani A., Bouguila N., Patterson Z. (2023). Short-term passenger flow prediction using a bus network graph convolutional long short-term memory neural network model // *Transportation Research Record*. — Vol. 2677. — No. 2. Pp. 1331–1340. DOI:10.1177/03611981221112673. [in Eng.].
- Bureau of National Statistics, Transport in the Republic of Kazakhstan (January–March 2024), Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. 2024. URL: <https://stat.gov.kz/en/industries/business-statistics/stat-transport/publications/157652>. [in Eng.].
- Cheng C. H., Tsai M. C., Cheng Y. C. (2022). An intelligent time-series model for forecasting bus passengers based on smartcard data, *Applied Sciences*. — Vol. 12. — No. 9. Pp. 4763. DOI:10.3390/app12094763. [in Eng.].
- Chawla P., Hasurkar R., Bogadi C. R., Korlapati N. S., Rajendran R., Ravichandran S., Gao J. Z. (2024). Real-time traffic congestion prediction using big data and machine learning techniques // *World Journal of Engineering*. — Vol. 21. — No. 1. Pp. 140–155. DOI:10.1108/WJE-07-2021-0428. [in Eng.].
- Chen Y., Li T., Sun Y., Wu J., Guo X., Liu D. (2023). Dynamic data-driven computation method for the number of waiting passengers and waiting time in the urban rail transit network, *IET Intell. Transp. Syst.* — Vol. 17. — No. 1. Pp. 165–179. DOI:10.1049/itr2.12245. [in Eng.].
- El Maazouzi Q., Retbi A., Bennani S. (2022). Automatisation hyperparameters tuning process for times series forecasting: application to passenger's flow prediction on a railway network, *Int. Arch. Photogramm. Remote Sens.*



Spatial Inf. Sci. — Vol. 48. — Pp. 53–60. DOI:10.5194/isprs-archives-XLVIII-4-W3-2022-53-2022. [in Eng.].

Fu X., Wu M., Ponnarasu S., Zhang L. A hybrid deep learning approach for real-time estimation of passenger traffic flow in urban railway systems, *Buildings*. — Vol. 13. — No. 6. Pp. 1514. DOI:10.3390/buildings13061514. [in Eng.].

Garriga M., Aarns K., Tsigkanos C., Tamburri D. A., Heuvel W. V. D. (2021). DataOps for cyber-physical systems governance: the airport passenger flow case // *ACM Trans. Internet Technol.* — Vol. 21. — No. 2. Pp. 1–25. DOI:10.1145/3432247. [in Eng.].

Ghaziamin P., Bajaj K., Bouguila N., Patterson Z. (2024). A privacy-preserving edge computing solution for real-time passenger counting at bus stops using overhead fisheye camera, in: *Proc. 2024 IEEE 18th Int. Conf. on Semantic Computing (ICSC)*, February. Pp. 25–32. DOI:10.1109/ICSC59802.2024.00011. [in Eng.].

Halyal S., Mulangi R. H., Harsha M. M. (2022). Forecasting public transit passenger demand: with neural networks using APC data, *Case Studies on Transport Policy*. — Vol. 10. — No. 2. Pp. 965–975. DOI: 10.1016/j.cstp.2022.03.011. [in Eng.].

Jain R., Dhingra S., Joshi K., Rana A., Goyal N. (2023). Enhance traffic flow prediction with real-time vehicle data integration // *Journal of Artificial Intelligence*. — Vol. 6. — No. 2. DOI:10.32629/jai.v6i2.574. [in Eng.].

Kolesnikova K., Naizabayeva L., Myrzabayeva A., Lisnevskiy R. (2024). Use the neural networks in prediction of environmental processes, in: *Proc. 2024 IEEE 4th Int. Conf. on Smart Information Systems and Technologies (SIST)*. — Astana. Pp. 625–630. DOI: 10.1109/SIST61555.2024.10629330. [in Eng.].

Kramar V., Alchakov V. (2023). Time-series forecasting of seasonal data using machine learning methods, *Algorithms*. — Vol. 16. — No. 5. Pp. 248. DOI:10.3390/a16050248. [in Eng.].

Liyanage S., Abduljabbar R., Dia H., Tsai P. W. (2022). AI-based neural network models for bus passenger demand forecasting using smart card data // *Journal of Urban Management*. — Vol. 11. — No. 3. Pp. 365–380. DOI: 10.1016/j.jum.2022.05.002. [in Eng.].

Liu S., Li X., He C. (2021). Study on dynamic influence of passenger flow on intelligent bus travel service model, *Transport*. — Vol. 36. — No. 1. Pp. 25–37. DOI:10.3846/transport.2021.14343. [in Eng.].

Li B., Guo T., Li R., Wang Y., Gandomi A. H., Chen F. (2023). Self-adaptive predictive passenger flow modeling for large-scale railway systems // *IEEE Internet of Things Journal*. — Vol. 10. — No. 16. Pp. 14182–14196. DOI:10.1109/IJOT.2023.3270427. [in Eng.].

McCarthy C., Ghaderi H., Jayaraman P. P., Dia H. (2023). A deep learning approach to real-time video analytics for people and passenger counting // in: *Handbook on Artificial Intelligence and Transport*, Edward Elgar Publishing. Pp. 348–379. DOI:10.4337/9781803929545.00022. [in Eng.].

Marczyk M., Kempski A., Socha M., Cogiel M., Foszner P., Staniszewski M. (2024). Passenger location estimation in public transport: evaluating methods and camera placement impact, *IEEE Trans. Intell. Transp. Syst.* DOI:10.1109/TITS.2024.3434746. [in Eng.].

Neftissov A., Sarinova A., Kazambaev I., Kirichenko L., Lisnevskiy R., Rzayeva L. (2023). Development of the Smart Office Concept, in: *Proc. 2023 IEEE Int. Conf. on Smart Information Systems and Technologies (SIST)*. — Astana. Pp. 382–386. DOI:10.1109/SIST58284.2023.10223512. [in Eng.].

Ou Y. (2022). AI for real-time bus travel time prediction in traffic congestion management, in: *Humanity Driven AI: Productivity // Well-being, Sustainability and Partnership*, Springer. Pp. 63–84. DOI:10.1007/978-3-030-72188-6\_4. [in Eng.].

Tao Z., Tang J., Hou K. (2021). Online estimation model for passenger flow state in urban rail transit using multi-source data, *Comput.-Aided Civil Infrastruct. Eng.* — Vol. 36. — No. 6. Pp. 762–780. DOI:10.1111/mice.12671. [in Eng.].

Wu J., Du B., Gong Z., Wu Q., Shen J., Zhou L., Cai C. (2023). A GTFS data acquisition and processing framework and its application to train delay prediction, *Int. J. Transp. Sci. Technol.* — Vol. 12. — No. 1. Pp. 201–216. DOI: 10.1016/j.ijtst.2022.01.005. [in Eng.].

Ye J., Xu Z., Gou X. (2022). An adaptive Grey–Markov model based on parameters self-optimization with application to passenger flow volume prediction, *Expert Systems with Applications*. — Vol. 202. — Article 117302. DOI: 10.1016/j.eswa.2022.117302. [in Eng.].

Yan M., Zhou K. (2023). Bus passenger flow prediction using BO-optimized XGBoost, in: *Proc. 2023 2nd Int. Conf. on Big Data, Information and Computer Network (BDICN)*, January. Pp. 134–141. DOI:10.1109/BDICN58493.2023.00035. [in Eng.].

Zeng L., Li Z., Yang J., Xu X. (2022). CEEMDAN-IPSO-LSTM: a novel model for short-term passenger flow prediction in urban rail transit systems, *Int. J. Environ. Res. Public Health*. — Vol. 19. — No. 24. Pp. 16433. DOI:10.3390/ijerph192416433. [in Eng.].

Zhang X., Lauber L., Liu H., Shi J., Xie M., Pan Y. (2022). Travel time prediction of urban public transportation

based on detection of single routes, PLoS One. — Vol. 17. — No. 1. e0262535. DOI: 10.1371/journal.pone.0262535. [in Eng.].

Zhai X., Shen Y. (2023). Short-term bus passenger flow prediction based on graph diffusion convolutional recurrent neural network // Applied Sciences. — Vol. 13. — No. 8. Pp. 4910. DOI:10.3390/app13084910. [in Eng.].

Zhang X., Yan M., Xie B., Yang H., Ma H. (2021). An automatic real-time bus schedule redesign method based on bus arrival time prediction // Advanced Engineering Informatics. — Vol. 48. Article 101295. DOI: 10.1016/j.aei.2021.101295. [in Eng.].

Zhao Y., Ma Z. (2022). Naïve Bayes-based transition model for short-term metro passenger flow prediction under planned events // Transportation Research Record. — Vol. 2676. — No. 9. Pp. 309–324. DOI:10.1177/03611981221086645. [in Eng.].



**INTERNATIONAL JOURNAL OF INFORMATION AND  
COMMUNICATION TECHNOLOGIES**

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ  
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И  
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**Собственник:**

АО «Международный университет информационных  
технологий» (Казахстан, Алматы)

**Главный редактор:**

Колесникова Катерина Викторовна

**Ответственный редактор:**

Мрзабаева Раушан Жалиевна

**Компьютерная верстка:**

Калабай Замзагуль Ертугановна

Сайт журнала: <https://journal.iitu.edu.kz>

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Подписано в печать 30.06.2026.

050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).