

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ
НАУК РЕСПУБЛИКИ КАЗАХСТАН
Казахский национальный
университет имени аль-Фараби

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
al-Farabi Kazakh National University

**SERIES
PHYSICO-MATHEMATICAL**

2 (342)

APRIL – JUNE 2022

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Қазақстан), **Н=5**

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

МАМЫРБАЕВ Өркен Жұмажанұлы, ақпараттық жүйелер мамандығы бойынша философия докторы (Ph.D), ҚР БҒМ Ғылым комитеті «Ақпараттық және есептеуші технологиялар институты» РМК жауапты хатшысы (Алматы, Қазақстан), **Н=5**

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат Нұрәділұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі (Алматы, Қазақстан), **Н=7**

БАЙГУНЧЕКОВ Жұмаділ Жанабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Сатпаев университетінің Қолданбалы механика және инженерлік графика кафедрасы, (Алматы, Қазақстан), **Н=3**

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физика), Люблин технологиялық университетінің профессоры (Люблин, Польша), **Н=23**

БОШКАЕВ Қуантай Авгазыұлы, Ph.D. Теориялық және ядролық физика кафедрасының доценті, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=10**

QUEVEDO Nemandó, профессор, Ядролық ғылымдар институты (Мехико, Мексика), **Н=28**

ЖҮСПОВ Марат Абжанұлы, физика-математика ғылымдарының докторы, теориялық және ядролық физика кафедрасының профессоры, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=7**

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина ҰҒА академигі, Қолданбалы математика және механика институты (Донецк, Украина), **Н=5**

РАМАЗАНОВ Тілекқабұл Сәбитұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университетінің ғылыми-инновациялық қызмет жөніндегі проректоры, (Алматы, Қазақстан), **Н=26**

ТАКИБАЕВ Нұрғали Жабағаұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=5**

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), **Н=42**

ХАРИН Станислав Николаевич, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Қазақстан-Британ техникалық университеті (Алматы, Қазақстан), **Н=10**

ДАВЛЕТОВ Асқар Ербуланович, физика-математика ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=12**

КАЛАНДРА Пьетро, Ph.D (физика), Наноқұрылымды материалдарды зерттеу институтының профессоры (Рим, Италия), **Н=26**

«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **№ 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика және ақпараттық коммуникациялық технологиялар сериясы.*

Қазіргі уақытта: *«ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Тиражы: *300 дана.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*
<http://www.physico-mathematical.kz/index.php/en/>

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2022

Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

ГЛАВНЫЙ РЕДАКТОР:

МУТАНОВ Галимжаир Мутанович, доктор технических наук, профессор, академик НАН РК, и.о. генерального директора «Института информационных и вычислительных технологий» КН МОН РК (Алматы, Казахстан), **Н=5**

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

МАМЫРБАЕВ Оркен Жумажанович, доктор философии (PhD) по специальности Информационные системы, ответственный секретарь РГП «Института информационных и вычислительных технологий» Комитета науки МОН РК (Алматы, Казахстан), **Н=5**

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

КАЛИМОЛДАЕВ Максат Нурадилович, доктор физико-математических наук, профессор, академик НАН РК (Алматы, Казахстан), **Н=7**

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Саптаева (Алматы, Казахстан), **Н=3**

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), **Н=23**

БОШКАЕВ Куантай Авгазыевич, доктор Ph.D, преподаватель, доцент кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=10**

QUEVEDO Hemando, профессор, Национальный автономный университет Мексики (UNAM), Институт ядерных наук (Мехико, Мексика), **Н=28**

ЖУСУПОВ Марат Абжанович, доктор физико-математических наук, профессор кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=7**

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), **Н=5**

РАМАЗАНОВ Тлеккабул Сабитович, доктор физико-математических наук, профессор, академик НАН РК, проректор по научно-инновационной деятельности, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=26**

ТАКИБАЕВ Нурғали Жабагаевич, доктор физико-математических наук, профессор, академик НАН РК, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=5**

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), **Н=42**

ХАРИН Станислав Николаевич, доктор физико-математических наук, профессор, академик НАН РК, Казахстанско-Британский технический университет (Алматы, Казахстан), **Н=10**

ДАВЛЕТОВ Аскар Ербуланович, доктор физико-математических наук, профессор, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=12**

КАЛАНДРА Пьетро, доктор философии (Ph.D, физика), профессор Института по изучению наноструктурированных материалов (Рим, Италия), **Н=26**

«Известия НАН РК. Серия физика-математическая».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **№ 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физика и информационные коммуникационные технологии.* В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

© Национальная академия наук Республики Казахстан, 2022
Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

EDITOR IN CHIEF:

MUTANOV Galimkair Mutanovich, doctor of technical Sciences, Professor, Academician of NAS RK, acting director of the Institute of Information and Computing Technologies of SC MES RK (Almaty, Kazakhstan), **H=5**

DEPUTY EDITOR-IN-CHIEF

MAMYRBAYEV Orken Zhumazhanovich, Ph.D. in the specialty information systems, executive secretary of the RSE “Institute of Information and Computational Technologies”, Committee of Science MES RK (Almaty, Kazakhstan) **H=5**

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, doctor in Physics and Mathematics, Professor, Academician of NAS RK (Almaty, Kazakhstan), **H=7**

BAYGUNCHEKOV Zhumadil Zhanabayevich, doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), **H=3**

WOICIK Waldemar, Doctor of Phys.-Math. Sciences, Professor, Lublin University of Technology (Lublin, Poland), **H=23**

BOSHKAYEV Kuantai Avgazievich, PhD, Lecturer, Associate Professor of the Department of Theoretical and Nuclear Physics, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=10**

QUEVEDO Hemando, Professor, National Autonomous University of Mexico (UNAM), Institute of Nuclear Sciences (Mexico City, Mexico), **H=28**

ZHUSSUPOV Marat Abzhanovich, Doctor in Physics and Mathematics, Professor of the Department of Theoretical and Nuclear Physics, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=7**

KOVALEV Alexander Mikhailovich, Doctor in Physics and Mathematics, Academician of NAS of Ukraine, Director of the State Institution «Institute of Applied Mathematics and Mechanics» DPR (Donetsk, Ukraine), **H=5**

RAMAZANOV Tlekkabul Sabitovich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Vice-Rector for Scientific and Innovative Activity, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=26**

TAKIBAYEV Nurgali Zhabagaevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=5**

TIGHINEANU Ion Mikhailovich, Doctor in Physics and Mathematics, Academician, Full Member of the Academy of Sciences of Moldova, President of the AS of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

KHARIN Stanislav Nikolayevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Kazakh-British Technical University (Almaty, Kazakhstan), **H=10**

DAVLETOV Askar Erbulanovich, Doctor in Physics and Mathematics, Professor, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=12**

CALANDRA Pietro, PhD in Physics, Professor at the Institute of Nanostructured Materials (Monterotondo Station Rome, Italy), **H=26**

News of the National Academy of Sciences of the Republic of Kazakhstan.

Physical-mathematical series.

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-Ж**, issued 14.02.2018
Thematic scope: *series physics and information technology*.

Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*

Periodicity: *4 times a year.*

Circulation: *300 copies.*

Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

© National Academy of Sciences of the Republic of Kazakhstan, 2022

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

УДК – 004

HTAMP - 50.01.75

Б.Б. Тастемір, Ф.У. Маликова, Р.Б. Айтбаева

Алматы технологиялық университеті, Алматы, Қазақстан.

E-mail: tbeka97@gmail.com

ЭЛЕКТРОНДЫҚ ПОШТА СПАМДЫ СҮЗГІЛЕУГЕ АРНАЛҒАН RANDOM FORESTS МАШИНАЛЫҚ ОҚЫТУ ӘДІСІ

Аннотация. Электрондық пошта спамы-әлемдегі әрбір электрон-дық пошта пайдаланушысының күнделікті кездесетін негізгі мәсе-лелерінің бірі. Күн сайын электрондық поштаны пайдаланушылар робот-бағдарламалық агенттер автоматты түрде жасайтын анонимді мекенжайлардан жаңа мазмұндағы жүздеген спам-хаттарды алады. Қара және ақ тізімдерді (домендер, IP мекенжайлары, пошта мекенжайлары) пайдалану сияқты қалаусыз поштаны сүзудің дәстүрлі әдістері спам-хабарламалар қаупін төмендетуде өте тиімсіз болып шықты. Бұл жоғары сенімді электрондық пошта спам сүзгілерін ойлап табу қажеттілігіне әкелді. Жақында машиналарды оқыту тәсілі спам хаттарын анықтау және сүзу үшін сәтті қолданылды. Бұл мақалада спам электрондық пошталарын тиімді жіктеу үшін кездейсоқ орман машинасын оқыту алгоритмін қолдану ұсынылады. Негізгі мақсат - болжаудың дәлдігі мен мүмкіндіктері аз қалаусыз пошта сүзгісін жасау. Хам, спам және қарапайым электрондық пошталардың 5180 электрондық поштасынан тұратын Enron жалпыға қол жетімді мәліметтер жиынтығынан random forests алгоритмі арқылы алынған және қолданылған 99,92% жіктеу дәлдігі, жалған позитивтердің өте төмен жиілігі (0,01) және шынайы позитивтердің өте жоғары жиілігі 0,999. Барлық эксперименттер WEKA деректерді модельдеу және машиналық оқыту ортасында жүзеге асырылады. Қолданыстағы электрондық пошта спамдарын сүзгілеу әдістерінің көпшілігі спам жіберушілермен күнделікті жіберілетін

кейбір спамдарды тиімді өңдей алмайды. Себебі спам жіберушілер спам сүзгісі арқылы анықтаудан жалтарудың күрделірек әдістерін ойлап табуды жалғастырды. Спамерлердің жаңа техниканы үздіксіз қолдануымен электрондық поштаны спамды сүзу зерттеушілер үшін қызу зерттеу саласына айналды. Бұл зерттеуде біз электрондық пошта спамдарын тиімді және тиімді сүзгілеуге арналған Random Forests алгоритмін ұсындық. Алгоритмнің тиімділігі мен тиімділігін анықтау үшін дәлдік, TPR, FPR, дәлдік және F-өлшемі арқылы Enron спам деректер жиынындағы RFs алгоритмінің өнімділігін бағалады. Біз RF электрондық пошта пайдалану шыларының кіріс жәшігіндегі спам хабарламаларының көлемін одан әрі азайту үшін пошта серверінде немесе пошта клиенті жағында қабылдануы мүмкін перспективалы алгоритм екенін айта отырып қорытындылаймыз.

Түйін сөздер: машиналық оқыту, спамды сүзу, кездейсоқ ормандар, нейрондық желілер, тірек векторлық Машиналар, аңғал Байес.

Б.Б. Тастемир, Ф.У. Маликова, Р.Б. Айтбаева

Алматинский технологический университет, Алматы, Казахстан.

E-mail: tbeka97@gmail.com

МЕТОД МАШИННОГО ОБУЧЕНИЯ RANDOM FORESTS ДЛЯ ФИЛЬТРАЦИИ НЕЖЕЛАТЕЛЬНОЙ ПОЧТЫ

Аннотация. Спам по электронной почте является одной из основных проблем, с которыми ежедневно сталкивается каждый пользователь электронной почты в мире. Ежедневно пользователи электронной почты получают сотни спам-писем нового содержания с анонимных адресов, которые автоматически генерируются программными агентами-роботами. Традиционные методы фильтрации спама, такие как использование черных и белых списков (домены, IP-адреса, почтовые адреса), оказались совершенно неэффективными в снижении угрозы спам-сообщений. Это привело к необходимости изобретения высоконадежных фильтров электронной почты для спама. В последнее время подход машинного обучения успешно применяется для обнаружения и фильтрации спам-писем. В этой статье предлагается использовать алгоритм машинного обучения случайного леса для эффективной классификации спам-сообщений электронной почты. Основная цель — разработать спам-фильтр электронной почты с большей точностью предсказания и меньшим количеством функций. Из общедоступного

набора данных Enron, состоящего из 5180 электронных писем, как спама, так и обычных писем, был извлечен набор характерных признаков спама (из литературы), которые были применены с помощью алгоритма случайных лесов с результирующей точностью классификации 99,92%, очень низкий ложноположительный результат (0,01) и очень высокий истинно положительный уровень 0,999. Все эксперименты проводятся в среде моделирования интеллектуального анализа данных и машинного обучения WEKA. Большинство существующих методов фильтрации спама в электронной почте неэффективно справляются с частью ежедневно рассылаемого спамерами спама. Это связано с тем, что спамеры продолжают изобретать более изощренные способы избежать обнаружения с помощью спам-фильтров. С постоянным использованием спамерами новых методов фильтрация электронной почты от спама стала горячей областью исследований для исследователей. В этом исследовании мы предложили алгоритм Random Forests для эффективной и действенной фильтрации спама в электронной почте. Чтобы определить эффективность и действенность алгоритма, Enron оценила производительность алгоритма RFs в наборе данных о спа-ме, используя точность, TPR, FPR, точность и F-размер. В заключение мы скажем, что RF является многообещающим алгоритмом, который можно использовать на почтовом сервере или на стороне почтового клиента для дальнейшего уменьшения количества спам-сообщений в папке входящих сообщений пользователей электронной почты.

Ключевые слова: машинное обучение, фильтрация спама, случайные леса, нейронные сети, опорные векторные машины, наивный Байес.

B.B. Tastemir, F.U. Malikova, R.B. Aitbayeva

Almaty Technological University, Almaty, Kazakhstan.

E-mail: tbeka97@gmail.com

RANDOM FORESTS MACHINE LEARNING TECHNIQUE FOR EMAIL SPAM FILTERING

Abstract. Email spam is one of the major problems that every email user in the world faces on a daily basis. Every day, email users receive hundreds of spam emails with new content from anonymous addresses that are automatically generated by robotic software agents. Traditional spam filtering methods such as blacklisting and whitelisting (domains, IP addresses, email addresses) have proven to be completely ineffective in

reducing the threat of spam messages. This necessitated the invention of highly reliable spam email filters. Recently, a machine learning approach has been successfully applied to detect and filter spam emails. This article proposes to use a random forest machine learning algorithm to effectively classify spam emails. The main goal is to develop an email spam filter with more predictive accuracy and fewer features. From Enron's public dataset of 5180 emails, both spam and regular, a spam feature set (from the literature) was extracted and applied using a random forest algorithm with a resulting classification accuracy of 99.92%, very poor. a false positive (0.01) and a very high true positive of 0.999. All experiments are carried out in the WEKA data mining and machine learning simulation environment. Most existing email spam filtering methods do not effectively handle some of the daily spam sent by spammers. This is because spammers have continued to invent more sophisticated ways to avoid detection through spam filters. With the constant use of new techniques by spammers, spam filtering of e-mail has become a hot field of research for researchers. In this study, we proposed a Random Forests algorithm for efficient and effective filtering of email spam. To determine the efficiency and effectiveness of the algorithm, Enron evaluated the performance of the RFs algorithm in the spam data set using accuracy, TPR, FPR, accuracy, and F-size. We conclude by saying that RF is a promising algorithm that can be adopted on the mail server or on the mail client side to further reduce the amount of spam messages in the inbox of e-mail users.

Key words: machine learning, spam filtering, random forests, neural networks, reference vector machines, Naive Bayes.

Кіріспе. Жақында спам деп аталатын сұраныссыз коммерциялық жаппай электрондық пошталар интернетте үлкен проблемаға айналды. Алаяқтық хаттарды жіберетін спамер әртүрлі веб-сайттарды, вирустарды және зиянды бағдарламаларды пайдаланып электрондық пошта мекенжайларын жинайды. Спам интернет пайдаланушыларына сақтау сыйымдылығын және желінің өткізу қабілеттілігін арттыруға кедергі келтіреді. Компьютерлік желілерде спам-хабарламалардың үлкен көлемінің болуы электрондық пошта серверінің жадын, өткізу қабілеттілігін, процессорды өңдеу жылдамдығын және пайдаланушы уақытын тиімді пайдалануға зиян келтіреді (Fonseca және т.б., 2016). Есептер спам-хабарламалар дүние жүзіндегі электрондық пошта трафигінің 77%-дан астамына жауап беретінін көрсетті (Касперский, 2017). Спам электрондық пошталар 419 интернет-хабарламаның және бейтаныс адамдарды пайдаланушы аты мен құпия

сөз, Банк растау нөмірі (BVN) және несие сияқты құпия ақпаратты шығаруға тарту мақсатында электрондық хаттарды жіберудің басқа алаяқтық тәжірибесінің құрбаны болған пайдаланушылар үшін өте тітіркендіргіш және жағымсыз. карта нөмірлері. Әдебиеттерде электрондық поштаны спам сүзгілеудің әртүрлі тәсілдерін ұсынған бірнеше жұмыс жарияланды. Электрондық хаттарды спам немесе спам емес деп жіктеу үшін сәтті қолданылды. Бұл әдістерге ықтималдық, шешім ағашы, жасанды иммундық жүйе (Бахгат және басқалар, 2016), тірек векторлық машинасы (SVM) (Bougila және Amaugі, 2009), жасанды нейрондық желілер (ANN) (Сао және басқалар, 2004) және жағдайға негізделген әдіс (Fdez-Riverola, 2007).

Электрондық хаттардағы белгілі бір мүмкіндіктерді (әдетте спам электрондық пошталарында жиі қолданылатын кілт сөздер) анықтау мүмкіндігі бар мазмұнға негізделген сүзу әдісін қолдану арқылы спам хаттарды сүзу үшін осы машинада оқыту әдістерін пайдалануға болатыны көрсетілді. Электрондық хаттарда бұл мүмкіндіктердің пайда болу жиілігі шекті мәнмен өлшенген кезде электрондық поштаның спам ретінде жіктелу ықтималдығын анықтайды. Шекті мәннен асатын электрондық пошта хабарлары спам ретінде жіктеледі (Mason, 2003).

Материалдар мен әдістер. Электрондық поштаны спамды сүзу әдістерінің көпшілігі мәтінді санаттау тәсілдерін пайдаланады. Демек, спам сүзгілері нашар жұмыс істейді және спам хаттардың пайдаланушылардың кіріс жәшігіне түсуіне тиімді жол бермейді. Бұл жұмыс электрондық пошталардан маңызды мүмкіндіктерді шығару және электрондық хаттарды ветчина, спам немесе қалыпты деп жіктеу үшін Кездейсоқ ормандар (RF) алгоритмін пайдалану ережелерін пайдаланады. Enron спам деректер жинағы эталондық деректер жинағы ретінде пайдаланылды. Random Forests машиналық оқыту алгоритмі WEKA көмегімен модельденді (Wang, 2005). WEKA-да деректерді алдын ала өңдеу, жіктеу, регрессия, кластерлеу және ассоциация ережелері үшін пайдалануға болатын машиналық оқыту алгоритмдерінің жиынтығы бар. WEKA жүйесінде енгізілген машиналық оқыту әдістері әртүрлі нақты әлемдегі мәселелерді шешуге көмектеседі. Құралдар жинағы зерттеушілер мен әзірлеушілерге машиналық оқытудың әртүрлі алгоритмдерімен тәжірибе жасау, олардың үлгілерін құру және бағалау үшін жақсы анықталған құрылымды ұсынады. Барлық эксперименттер AMD A 10-7300 Radeon R6, 10 есептеу ядросы 4C+6G, 1,90 ГГц, 8,00 ГБ жедел жады бар машинада жүргізілді.

Random Forest. Random Forests (RFs) ансамбльді оқытудың классикалық мысалы және деректерді жіктеу мәселелерін шешуге

жарамды регрессия әдісі. Breiman and Cutler (2007) RFs алгоритмін ұсынды. Алгоритм шешім ағаштарының көмегімен деректерді әртүрлі класстарға жіктейді. Жаттығу кезеңінде кейбір шешім ағаштары жасалады және кейінірек жіктеу тапсырмалары үшін пайдаланылады. Бұл жеке ағаштардың сайланған сыныбын ескере отырып жұмыс істейді және ең көп дауыс жинаған сынып соңғы нәтиже ретінде қарастырылады. RF алгоритмі жылдар бойы өте танымал болды және ол адам әрекетінің әртүрлі салаларындағы ұқсас мәселелерді шешу үшін қолданылады (Fette et al., 2007), (Koprinska, et al., 2007) және (Whittaker et al., 2010).). Кездейсоқ ормандардың бірнеше артықшылығы бар, мысалы: жіктеу қатесінің азаюы және кейбір басқа машиналық оқыту әдістерімен салыстырғанда жақсы f-баллдары. Оның үстіне, оның өнімділігі әдетте SVMs сияқты жақсы немесе одан да жоғары. Ол мәндері жоқ теңгерімсіз деректер жиындарын тиімді өңдей алады. Ол жетіспейтін деректердің болжалды мәнін есептеудің және деректердің елеулі бөлігі жетіспейтін жағдайларда деректердің дәлдігін сақтаудың тиімді алгоритмі ретінде қызмет етеді. RF үшін оқу уақыты әдетте SVM және нейрондық желілермен салыстырғанда қысқа (бірақ бұл жеке енгізуге байланысты). RF дәлдігі бойынша қолданыстағы машиналық оқыту алгоритмдерінің көпшілігінен жақсырақ. Оның үлкен дерекқорлардағы өнімділігі өте жақсы. Ол жүздеген мың кіріс айнымалыларын тиімді өңдей алады. RF орманды өңдеу кезінде ұжымдық қатенің ішкі бейтарап болжамын жасайды. Ол бейтарап деректер жиыны бар популяция класындағы қателерді жеңілдететін тәсілді ұсынады. RF таңбаланбаған деректерді тиімді өңдеу мүмкіндігі бар, бұл таңбаланбаған деректерді кластерлеудің өте қолайлы әдісі. Кездейсоқ ормандар күрделі емес және ол бақылаулар санымен салыстырғанда азырақ параметрлерді пайдаланады. RF пайдаланушыға мүмкіндігінше көп ағаштарды жоғары жылдамдықпен өсіруге мүмкіндік береді. RF ормандағы жеке ағаштардың жанында кіріс векторын қосу арқылы кіріс векторынан жаңа деректерді жіктейді. Әрбір ағаш өзінің классификациясын орындайды, ол әдетте сол сынып үшін ағаш «дауыстары» ретінде белгілі. Орман ормандағы жалпы ең көп дауыс жинаған сыныпты таңдайды. Ағаштарды өсірудің қадамдары төменде сипатталған:

1. N бар деректерден ауыстыруға болатын N даналарды кездейсоқ түрде көрсететін оқу даналарының саны делік. Мұндай мысалдар ағашты өсіруге арналған жаттығулар жиынтығы ретінде пайдаланылады.

2. P кіріс айнымалылары бар делік, $p \ll P$ саны сәйкес келетін әрбір

түйін үшін P айнымалылары P ішінен кездейсоқ таңдалатын және p нүктесіндегі ең жақсы бөлік түйінді бөлу үшін пайдаланылады, осылайша p енді болады. орман өсірудің барлық кезеңінде тұрақты құндылық.

3. Ағаштарды кесуге тыйым салынады, өйткені әрбір ағаш мүмкін болатын ең үлкен деңгейге дейін өсіріледі.

Ағаш қателік деңгейі аз болса, күшті классификатор деп аталады. Оның үстіне ормандағы әрбір ағаштың шоғырлануы артқан сайын орманның қателік деңгейі төмендейді. p мәнін азайту орманның қатынасын және күшін өзара төмендетеді, ал p мәнін арттыру әдетте өте кең болып табылатын p -тің ең жақсы шекарасының аймағында да артады. p мәнін қапшықтан тыс (OOB) қатесі арқылы есептеуге болады (сонымен қатар қаптан тыс бағалау ретінде белгілі) шек ішіндегі p мәнін дереу табуға болады. Бұл кездейсоқ ормандар оның дәл бапталуына аздап сезімтал болатын жалғыз сандық фактор.

Төмендегі алгоритм орман ағаштарын жасау үшін қажетті қадамдарды қысқаша сипаттады.

RF алгоритмін іске қосу

Кіріс: X : түйіндер саны

N : функциялар саны

Y : өсетін ағаштар саны

Шығыс: G : ең көп дауыс жинаған класс

While Тоқтату критерийлері дұрыс емес do:

Жаттығу деректерінен Y кездейсоқ түрде өздігінен басталатын S үлгісін таңдаңыз

Төмендегі қадамдарды қолдана отырып, таңдалған өздігінен басталатын S үлгісінен R_i ағашын жасаңыз:

(1) N ішінен кездейсоқ n мүмкіндікті таңдаңыз; мұндағы $n \ll N$

(2) n мүмкіндіктер арасындағы d түйіні үшін ең жақсы бөлу нүктесін есептеңіз

(3) Негізгі түйінді оңтайлы бөлу арқылы екі ұрпақ түйініне бөліңіз

(4) Түйіндердің максималды саны (x) жасалғанша 1-3-қадамдарды орындаңыз Y рет 1-4 қадамдарды қайталау арқылы орманыңызды жасаңыз

EndWhile

Әрбір жасалған ағаштар үшін өнім шығару $\{R_t\} 1Y$

Түбір түйінінен бастап әрбір жасалған ағаштар үшін жаңа үлгіні пайдаланыңыз Үлгіні жапырақ түйініне сәйкес келетін сыныпқа тағайындаңыз.

Әрбір ағаштың дауыстарын немесе нәтижелерін біріктіріңіз

Ең көп дауыс жинаған сыныпты шығарыңыз (G).

End RF алгоритмі

Эксперимент үшін қолданылатын мәліметтер жиынтығы. Enron спам деректер жинағы біздің тәжірибеміз үшін пайдаланылды (Копринска, және т.б., 2007). Бұл зерттеуде Enron корпорациясының Enron спам деректер жинағы пайдаланылады. Үш қалтада деректер жинағы ретінде 5180 электрондық пошта бар: қалыпты үшін норма, спам емес үшін ветчина және спам электрондық пошта үшін спам. Enron-да 5180 дана, 3672 ветчина, 8 норма және 1500 спам электрондық поштасы бар. Деректер жинағының мүмкіндіктері келесідей:

i. Электрондық хаттарда белгілі бір сөз немесе таңба қайталанып отырды.

ii. Орындау ұзақтығы атрибуттары (55-57) дәйекті бас әріптер тізбегінің ұзындығын өлшейді.

Деректерді Қалыпқа Келтіру Процесі. Біздің эксперименттерімізде пайдаланылған бастапқы деректер жинағы 5180 мәтіндік файлдан тұрады. Бұл файлдардағы деректер қалыпқа келтірілмеген. Бұл WEKA-ға кіріс ретінде қызмет ету үшін оларды қалыпқа келтіру керек дегенді білдіреді. Оқыту үшін WEKA-ға бермес бұрын, барлық деректерді бір .arff файлына түрлендіру қажет. Бұған қол жеткізу үшін WEKA пәрмен жолы интерфейсінде келесі пәрменді қолданамыз.

```
“java weka.core.converters.TextDirectoryLoader -dir D:/Enron > D:/Spam_mails.arff”
```

Қалыпқа келтіру процесінен кейін қалыпқа келтірілген файл алдын ала өңдеу үшін WEKA-ға берілді.

Белгілерді алу. Алдын ала өңдеу кезеңінен кейін мүмкіндікті шығару келеді. Мүмкіндіктерді шығару - оқу жинағында кездесетін терминдердің ішкі жиынын таңдау және мәтінді жіктеуде мүмкіндіктер ретінде тек осы жиынды пайдалану процесі. Бұған кейбір ережелер жиынтығын қолдану арқылы қол жеткізіледі. Мүмкіндіктерді шығару тиімді сөздіктің көлемін азайту арқылы жіктеуішті оқытуды және қолдануды тиімдірек етеді. Сондай-ақ әдетте шу мүмкіндіктерін жою арқылы жіктеу дәлдігін арттырады. Спамды сүзгілеу үшін пайдаланған кейбір маңызды электрондық пошта мүмкіндіктеріне мыналар жатады: Хабардың негізгі мазмұны және тақырыбы, Хабардың көлемі, Сөздердің пайда болу саны, Хабардағы мағыналық сәйкессіздік үлгілерінің саны, Алушының жасы, Жынысы мен елі, Алушының жауап бергені, Ересектерге арналған мазмұн, Хабарлама мазмұнындағы сөздер қапшығы, Домен атауы, IP мекенжайы, Негізгі бөлімде көбірек бос жолдар.

Нәтижелер және талқылау. Бұл бөлімде орындалған тәжірибелердің нәтижелері берілген. Деректер жиынын жіктеу және бағалау

үшін «Кездейсоқ орман» алгоритмі қолданылды, біз 10 еселік кросс-валидация сынағы қолданылды, бұл модельді оқыту үшін бастапқы жиынтықты оқу үлгісіне бөлетін болжамды модельдерді бағалауда қолданылатын тәсіл және оны бағалау. Біріншіден, деректер жиынын оқыту әрбір хабарламатақырыбын талдау, кілт сөздерді және ақтізім/қара тізімді тексеру арқылы алынған мүмкіндік векторларымен орындалды. Оқытылған үлгілердің өнімділігі оның жіктелу дәлдігі үшін 10 еселік айқаспалы валидация арқылы бағаланады. Жіктеу дәлдігі электрондық поштаның спам классификациясының өнімділік көрсеткіштерінің бірі болып табылады. Ол сынақ деректер жинағындағы дұрыс жіктелген даналардың саны мен сынақ жағдайларының жалпы санының қатынасы ретінде өлшенеді. Спамды сүзгілеуде жалған негативтер кейбір спам хаттардың спам емес деп қате жіктелгенін және пайдаланушының кіріс жәшігіне кіруіне рұқсат етілгенін білдіреді. Жалған позитив спам емес электрондық пошталардың қате спам ретінде жіктелгенін және спам қалтасына жылжытылғанын немесе жойылғанын білдіреді. Көптеген пайдаланушылар үшін жарамды электрондық пошталарды қате түрде спам ретінде жіктеу олардың кіріс жәшігіндегі спам хаттарды алудан өте қымбат болуы мүмкін. Жалған оң көрсеткіш электрондық пошта спам сүзгісінің тиімділігін бағалауда қолданылатын өнімділік көрсеткіштерінің бірі болып табылады. Төмендегі 1-суретте WEKA модельдеу ортасындағы нәтиженің скриншоты көрсетілген.

```

Classifier output
-----
@data

Classifier Model
RandomForest

Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S -2118349930 -do-not-check-capabilities

Time taken to build model: 45.2 seconds
=== Evaluation on training set ===

Time taken to test model on training data: 4.25 seconds

=== Summary ===

Correctly Classified Instances      5176          99.9228 %
Incorrectly Classified Instances     4             0.0772 %
Kappa statistic                    0.9981
Mean absolute error                 0.0296
Root Mean Squared Error            0.04
Relative absolute error             10.7404 %
Root relative squared error        16.1506 %
Total Number of Instances          5180

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
0.999   0.001   1.000   0.999   0.999   0.998   1.000   1.000   ham
1.000   0.000   1.000   1.000   1.000   1.000   1.000   1.000   norm
0.999   0.001   0.998   0.999   0.999   0.998   1.000   1.000   spam
Weighted Avg.   0.999   0.001   0.999   0.999   0.999   0.998   1.000   1.000

=== Confusion Matrix ===
 a  b  c  <-- classified as
3669  0  3  |  a = ham
  0  8  0  |  b = norm
  1  0 1499 |  c = spam
    
```

Сурет-1 - Enron спам электрондық поштасының деректер жиыны үшін Random Forests жіктеу шығысының скриншоты

Тиімділік. Бұл бөлімде біз барлық машиналық оқыту классификаторларының тиімділігін үлгіні жасауға кететін уақыт, дұрыс жіктелген даналар, қате жіктелген даналар және жіктеу дәлдігі тұрғысынан бағалаймыз. Нәтижелер 1-кестеде көрсетілген.

1-кесте - RFS алгоритмінің жұмысын бағалау

Бағалау критерийлері	
Модельді (модельдерді) жасауға жұмсалған уақыт	17,75
Дұрыс жіктелген даналар	5176
Қате жіктелген даналар	4
Дәлдік (%)	99,92

Біз қарастырып жатқан машиналық оқыту алгоритмдерінің өнімділігін әділ және жақсырақ бағалау үшін бұл жұмыста модельдеу қатесі де ескеріледі. Бұл алгоритмдердің тиімділігі келесі терминдер арқылы бағаланады: Карра статистикасы (KS), Орташа абсолютті қате (MAE), Орташа түбірлік қате (RMSE), Салыстырмалы абсолютті қате (RAE), Түбір салыстырмалы квадраттық қате (RRSE). KS, MAE және RMSE сандық мәндерде. RAE және RRSE пайызбен берілген. Нәтижелер 2-кестеде көрсетілген.

Кесте 2 - RFS алгоритмдерін оқыту және модельдеу қатесі

Бағалау критерийлері	RFS
Каппа статистикасы (KS)	0,9981
Орташа абсолютті қате (MAE)	0,0296
Орташа квадраттық қате (RMSE)	0,06
Салыстырмалы абсолюттік қателік (RAE) %	10.7404
Квадраттың орташа квадраттық абсолютті қатесі (RRAE) %	16.1506

Тиімділік. Болжамдық модельді жасағаннан кейін RFs алгоритмінің тиімділігі төмендегі 3-кестеде көрсетілгендей бағаланды.

Кесте 3 - TPR, FTR, дәлдік және F-Score негізіндегі RFS алгоритмдерінің жұмысын бағалау

Техника	TPR	FPR	Дәлдігі	F-есебі	Классы
RFS	0.999	0.001	1.000	0.998	Хам
	1.000	0.000	1.000	1.000	Қарапайым
	0.999	0.001	0.998	0.998	Спам

1-кестеден оның моделін құру үшін RF шамамен 17,75 секунд қажет болды. RFs 99,92% жіктеу дәлдігіне ие. Сондай-ақ

нәтижелерден RF өте жоғары дұрыс жіктелген даналар және қате жіктелген даналардың өте аз саны бойынша тамаша орындағаны анық. 2-кестеде бейнеленген оқыту және модельдеу қатесі RF тамаша классификация нәтижесін (0,9992%) және өте төмен қателік көрсеткішін (0,0296) бергенін көрсетеді. Модель жасалғаннан кейін келесі қадам қарастырылатын алгоритмдердің тиімділігін анықтау үшін алынған нәтижелерді талдау болып табылады. 3-кестеде RF-ның ветчина, норма және спам кластары үшін TPR, FTR, Precision және F-Score терминдерінде өте жақсы нәтиже бар екені көрсетілген. Төменде 4-кестеде классификаторлардың өнімділігін бағалаудың практикалық әдісін қамтамасыз ететін RF алгоритмінің шатастыру матрицалары берілген, кестенің әрбір жолы сыныптың нақты жылдамдығын білдіреді, ал әрбір баған болжамды көрсетеді.

Кесте 4 - RF алгоритмі үшін шатастыру матрицасы

	Хам	Қарапайым	Спам	Class
RFs	3669	0	3	Хам
	0	8	0	Қарапайым
	1	0	1499	Спам

Жоғарыдағы 4-кестеден RF 3672 дананың 3669 данасын (шынында ветчина болып табылатын 3669 ветчина данасы және шын мәнінде спам болып табылатын 1 спам данасы) және қате болжанған 3 дананы (спам ретінде болжанған ветчина класының 3 данасы) дәл болжайды. Біздің эксперименттерімізден оның жіктелу дәлдігін, TPR, FPR, дәлдік және F-бағалылығын ескере отырып, RF тиімділік пен тиімділік тұрғысынан тамаша орындағаны анық. Ол сондай-ақ 1500 дананың 1499 данасын (шынымен спам болып табылатын 1499 спам данасы және шын мәнінде спам болып табылатын 1 ветчина данасы) және қате болжанған 3 данасын (ветчина ретінде болжанған спам класының 3 данасы) дұрыс болжайды.

Information about the author:

B.B. Tastemir – Almaty Technological University, Almaty, Kazakhstan,
E-mail: tbeka97@gmail.com, OCRID ID - 0000-0002-3869-1973.

F.U. Malikova – Almaty Technological University, Almaty, Kazakhstan, E-mail: feruza-malikova@mail.ru, OCRID ID - 0000-0003-0527-322X .

R.B. Aitbayeva – Almaty Technological University, Almaty, Kazakhstan,
E-mail: rakhatay@mail.ru, OCRID ID/0000-0002-4266-3887.

ӘДЕБИЕТТЕР

Breiman L., Cutler A. (2007). Random forests-classification description, Department of Statistics Homepage, 2007, <http://www.stat.berkeley.edu/~breiman/RandomForests/cchome.htm>.

- Cao Y., Liao X., Li Y. (2004). An e-mail filtering approach using neural network, In International Symposium on Neural Networks, Springer Berlin Heidelberg, 688-694.
- Dhanaraj KR, Palaniswami V (2014). Firefly and Bayes Classifier for Email Spam Classification in a Distributed Environment. Australian Journal of Basic and Applied Sciences, 8(17):118-130.
- Fdez-Riverola F., Iglesias E.L., Diaz F., Méndez J.R., Corchado J.M. (2007). SpamHunting: An instance-based reasoning system for spam labelling and filtering, Decision Support Systems, 43(3):722-736.
- Fonseca D.M., Fazzion O.H., Cunha E., Las-Casas I., Guedes P.D., Meira W., Chaves M. (2016). Measuring Characterizing, and Avoiding Spam Traffic Costs. IEEE Internet Computing, 99.
- Karthika R., Visalakshi P. (2015). A Hybrid ACO Based Feature Selection Method for Email Spam Classification. WSEAS Transaction on Computers, 14, pp. 171-177.
- Kaspersky lab Spam Report (2017). Visited on May 15, 2018 https://www.securelist.com/en/analysis/204792230/Spam_Report_April_2012, 2012.
- Koprinska I., Poon J., Clark J., Chan J. (2007). Learning to classify e-mail, Information Sciences, 177(10): 2167–2187.
- Sharma A. and Suryawansi A. (2016). A Novel Method for Detecting Spam Email using KNN Classification with Spearman Correlation as Distance Measure. International Journal of Computer Applications, 136 (6):28-34.

REFERENCES

- Breiman L., Cutler A. (2007). Random forests-classification description, home page of the Department of Statistics, 2007, <http://www.stat.berkeley.edu/~Breiman/randomforests/cchome.htm>.
- Cao Yi, Liao X., Li Yi (2004). An approach to email filtering using a neural network, at the International Symposium on neural networks, Springer Berlin Heidelberg, 688-694.
- Dhanaray kr, Palaniswami in (2014). Firefly and Bayesian email classifier. Classification in distributed environments. Australian Journal of fundamental and Applied Sciences, 8 (17): 118-130.
- Fdes-Riverola F., Iglesias El, Diaz F., Mendes Jr., Corchado Jr.Moscow (2007). Spam hunting: a reasoned thinking system for tagging and filtering spam, a decision-making support system, 43 (3): 722-736.
- Fonseca D.M., Facion O., Kunya E., Las Casas I., Guedes P.D., Meiram J., Chavez M. (2016). Measuring characteristics and preventing spam traffic losses. IEEE Internet Computing, 99.
- Kartika R., Visalakshi P. (2015). A hybrid method for selecting ACO-based functions to classify email spam. Wseas transactions on computers, pp. 14, 171-177.
- Kaspersky Lab spam report (2017). Visited on May 15, 2018 https://www.securelist.com/en/analysis/204792230/Spam_Report_April_2012, 2012.

МАЗМҰНЫ

Т.И. Ганиева, Н.С. Семенов, С.Р. Семенов ЖАҒАНДЫҚ ҚОҒАМНЫҢ АҚПАРАТТЫҚ ИНФРАҚҰРЫЛЫМЫ САЛАСЫНДАҒЫ АҚПАРАТТЫҚ ҚАТЫНАСТАРДЫҢ КИБЕРҚАУПСІЗДІГІ.....	5
Е.С. Голенко, А.А. Исмаилова, А.С. Жумаханова «GENE ONTOLOGY» БАЗАСЫН ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ ҮЛГІЛЕРІН ПАЙДАЛАНА ОТЫРЫП АҚУЫЗ ФУНКЦИЯЛАРЫН БОЛЖАУ.....	19
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали СУ ЭКОЖҮЙЕЛЕРІН ЗЕРТТЕУДІҢ АҚПАРАТТЫҚ ТАЛДАУ ЖҮЙЕСІН ӨЗІРЛЕУ.....	39
А.А. Мырзатай, Л.Г. Рзаева, Г. Абитова, М.А. Жакенов ОҚИҒАЛАРДЫ БОЛЖАУ ЖҮЙЕЛЕРІНІҢ КІРІСТЕРІН ЖҮЙЕЛЕУ ҮШІН LAN МОНИТОРИНГ ЖҮЙЕСІН ЕНГІЗУ ЖӘНЕ ПАЙДАЛАНУ.....	54
Ж.С. Иксебаева, К. Жетписов, Ж.М. Муратова ГАНТ ДИАГРАММАСЫН ҚҰРУДЫҢ АҚПАРАТТЫҚ ЖҮЙЕСІ.....	64
Қ.Т. Қырғызбай, Е.Х. Какимжанов, Ж.М. Сагинтаев ГАЖ-ТЕХНОЛОГИЯЛАРЫ НЕГІЗІНДЕ АЛМАТЫ ОБЛЫСЫН АГРОКЛИМАТТЫҚ АУДАНДАСТЫРУ.....	76
А.А. Мухитова, А.С. Еримбетова, В.Б. Барахнин, Э.Н. Дайырбаева, А. Адалбек РЕЛЯЦИЯЛЫҚ ЖӘНЕ УАҚЫТҚА ТӘУЕЛДІ XML-ДЕРЕКТЕР ҚОРЫНДАҒЫ XML-ДЕРЕКТЕРДІ ӨНДЕУДІҢ ЗАМАНАУИ ӘДІСТЕРІ....	92
Б.Б. Оразбаев, Ж.Ж. Молдашева, В.И. Гончаров, К.Н. Оразбаева МАГИСТРАЛДЫ ҚҰБЫРЛАРМЕН МҰНАЙ ТАСМАЛДАУДЫ ДИАГНОСТИКАЛАУ ЖӘНЕ БАСҚАРУ ЖҮЙЕЛЕРІ.....	112
Б.Б. Тастемір, Ф.У. Маликова, Р.Б. Айтбаева ЭЛЕКТРОНДЫҚ ПОШТА СПАМДЫ СҮЗГІЛЕУГЕ АРНАЛҒАН RANDOM FORESTS МАШИНАЛЫҚ ОҚЫТУ ӘДІСІ.....	130
А. Урынбасарова, Д. Урынбасарова, Э. Ал-Хуссам ҚАЗАҚ ТІЛІНІҢ ЛАТЫН ГРАФИКАСЫНА АРНАЛҒАН ВЕБ-САЙТ.....	142
Э.Э. Эльдарова, В.В. Старовойтов, К.Т. Искаков БҰРМАЛҒАН КОНТРАСТТЫ ЦИФРЛЫҚ БЕЙНЕНІҢ ВИЗУАЛДЫ САПАСЫН ЖАҚСARTУ.....	153

СОДЕРЖАНИЕ

Т.И. Ганиева, Н.С. Семенов, С.Р. Семенов КИБЕРБЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ОТНОШЕНИЙ В СФЕРЕ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ ГЛОБАЛЬНОГО ОБЩЕСТВА.....	5
Е.С. Голенко, А.А. Исмаилова, А.С. Жумаханова ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКОВ ПРИ ПОМОЩИ БАЗЫ ДАННЫХ «GENE ONTOLOGY» И МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ.....	19
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали К РАЗРАБОТКЕ ИНФОРМАЦИОННОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ ИССЛЕДОВАНИЯ ВОДНЫХ ЭКОСИСТЕМ.....	39
А.А. Мырзатай, Л.Г. Рзаева, Г. Абитова, М.А. Жакенов ВНЕДРЕНИЕ И ИСПОЛЬЗОВАНИЕ СИСТЕМ МОНИТОРИНГА ЛВС ДЛЯ СИСТЕМАТИЗИРОВАНИЯ ВХОДНЫХ ДАННЫХ СИСТЕМ ПРОГНОЗИРОВАНИЯ ИНЦИДЕНТОВ.....	54
Ж.С. Иксебаева, К. Жетписов, Ж.М. Муратова ИНФОРМАЦИОННАЯ СИСТЕМА ПОСТРОЕНИЯ ДИАГРАММЫ ГАНТА.....	64
Қ.Т. Қырғызбай, Е.Х. Какимжанов, Ж.М. Сагинтаев АГРОКЛИМАТИЧЕСКОЕ РАЙОНИРОВАНИЕ АЛМАТИНСКОЙ ОБЛАСТИ С ПРИМЕНЕНИЕМ ГИС-ТЕХНОЛОГИЙ.....	76
А.А. Мухитова, А.С. Еримбетова, В.Б. Баракнин, Э.Н. Дайырбаева, А. Адалбек СОВРЕМЕННЫЕ МЕТОДЫ ОБРАБОТКИ XML-ДАННЫХ В РЕЛЯЦИОННЫХ И ВРЕМЕННЫХ XML-БАЗАХ ДАННЫХ.....	92
Б.Б. Оразбаев, Ж.Ж. Молдашева, В.И. Гончаров, К.Н. Оразбаева ДИАГНОСТИРОВАНИЕ СИСТЕМЫ УПРАВЛЕНИЯ ТРАНСПОРТИРОВКИ НЕФТИ ПО МАГИСТРАЛЬНЫМ ТРУБОПРОВОДАМ.....	112
Б.Б. Тастемир, Ф.У. Маликова, Р.Б. Айтбаева МЕТОД МАШИННОГО ОБУЧЕНИЯ RANDOM FORESTS ДЛЯ ФИЛЬТРАЦИИ НЕЖЕЛАТЕЛЬНОЙ ПОЧТЫ.....	130
А. Урынбасарова, Д. Урынбасарова, Э. Ал-Хуссам ВЕБ-САЙТ ЛАТИНСКОЙ ГРАФИКИ КАЗАХСКОГО ЯЗЫКА.....	142
Э.Э. Эльдарова, В.В. Старовойтов, К.Т. Искаков УЛУЧШЕНИЕ ВИЗУАЛЬНОГО КАЧЕСТВА КОНТРАСТНО ИСКАЖЕННЫХ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ.....	153

CONTENTS

T.I. Ganieva, N.S. Semenov, S.R. Semenov CYBERSECURITY OF INFORMATION RELATIONS IN THE FIELD OF INFORMATION INFRASTRUCTURE OF A GLOBAL SOCIETY.....	5
Y.S. Golenko, A.A. Ismailova, A.S. Zhumakhanova PREDICTING PROTEIN FUNCTIONS USING THE «GENE ONTOLOGY» DATABASE AND MACHINE LEARNING MODELS.....	19
R.M. Moldasheva, A.A. Ismailova, A.K. Zhamangara, A.M. Zadagali ABOUT DEVELOPMENT OF AN INFORMATION ANALYTICAL SYSTEM FOR THE STUDY OF AQUATIC ECOSYSTEMS.....	39
A.A. Myrzatay, L.G. Rzayeva, G. Abitova, M.A. Zhakenov THE IMPLEMENTATION AND THE USE OF THE LAN MONITORING SYSTEMS FOR SYSTEMATISATION OF THE INPUT DATA OF THE INCIDENT FORECASTING SYSTEMS.....	54
Zh.S. Ixebayeva, K. Jetpisov, Zh.M. Muratova INFORMATION SYSTEM FOR CONSTRUCTING GANTT CHARTS.....	64
K.T. Kyrgyzbay, E.Kh. Kakimzhanov, Jay Sagin AGRO-CLIMATIC ZONING OF ALMATY REGION USING GIS TECHNOLOGIES.....	76
A.A. Mukhitova, A.S. Yerimbetova, V.B. Barakhnin, E. Daiyrbayeva, A. Adalbek MODERN METHODS OF PROCESSING XML DATA IN RELATIONAL AND TEMPORARY XML DATABASES.....	92
B.B. Orazbayev, Zh.Zh. Moldasheva, B.I. Goncharov, K.N. Orazbayeva DIAGNOSTICS AND SYSTEMS OF OIL TRANSPORTATION THROUGH MAIN PIPELINES.....	112
B.B. Tastemir, F.U. Malikova, R.B. Aitbayeva RANDOM FORESTS MACHINE LEARNING TECHNIQUE FOR EMAIL SPAM FILTERING.....	130
A. Urynbassarova, D. Urynbassarova, E. Al-Hussam WEBSITE FOR THE LATIN SCRIPT OF THE KAZAKH LANGUAGE.....	142
E.E. Eldarova, V.V. Starovoytov, K.T. Iskakov IMPROVED VISUAL QUALITY OF CONTRAST DISTORTED DIGITAL IMAGES.....	153

**Publication Ethics and Publication Malpractice
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

www.nauka-nanrk.kz

<http://physics-mathematics.kz/index.php/en/archive>

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Заместитель директор отдела издания научных журналов НАН РК *Р. Жәліқызы*

Редакторы: *М.С. Ахметова, Д.С. Аленов*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 29.06.2022.

Формат 60x881/8. Бумага офсетная. Печать – ризограф.

9,0 п.л. Тираж 300. Заказ 1.