

АННОТАЦИЯ

философия докторы (PhD) ғылыми дәрежесін алу үшін ұсынылған «Қазақстан Республикасының медиа кеңістігінде қоғамның әлеуметтік көңіл-күйін талдау мен бағалаудың интеллектуалды жүйесіне арналған модульді зерттеу және әзірлеу» тақырыбындағы диссертациялық жұмысы 6D070300 – Ақпараттық жүйелер мамандығы бойынша

Жұмыстың өзектілігі. Қазіргі уақытта интернет-технологиялардың дамуы әлемдегі түрлі оқиғаларды сипаттайтын жаңалықтар сайттары мен әлеуметтік желілер санының айтарлықтай артуына ықпал етті. Әлеуметтік желіде жергілікті және жаһандық болып жатқан оқиғалар туралы пікірлерді, ойларды және идеяларды жариялау әдеттегі тәжірибеге айналды. Көптеген әлеуметтік желілер, Twitter, Facebook, YouTube және басқалары танымал болып қала береді және көптеген пайдаланушыларды тартады. Ал жаңа платформалар, TikTok, Instagram, Pinterest және т.б. әлемде болып жатқан оқиғалардың үлкен санын егжей-тегжейлі көрсететін әлеуметтік медиа әлемінде танымал болуда.

Жаңалықтар тақырыптары мен пайдаланушылардың пікірлері керемет жылдам қарқынмен өсіп келе жатқандықтан, өмірдің әртүрлі салаларындағы (саясат, экономика, азаматтық қоғам, білім, денсаулық сақтау, экология, мәдениет) ең маңызды тақырыптарды қадағалаудың маңызды қажеттілігі бар. және спорт және т.б.). Әлеуметтік желілерде бөлісілген олар туралы фактілер мен пікірлердің көптігі аналитикалық платформалардың маңыздылығын арттыратын автоматтандырылған әдістерсіз мұндай бақылауды мүмкін емес етеді. Мұндай платформалардың ең маңызды бөлігі - көңіл-күйді талдау модулі. Жасанды интеллект әдістері адамның сезімін, эмоциясын, мәдениеті мен менталитетін толық түсіне алмаса да, аналитикалық құралдарды пайдалана отырып, белгілі бір оқиғаларға қоғамдық пікірдің тенденциясын анықтай алады. Қолмен талдау өте ұзақ және ресурсты қажет ететін процесс, сонымен қатар ол белгісіздік пен түсініксіздікті қалдырады. Алгоритмдерді пайдалану операциялық аналитиканы жылдам алуға және әртүрлі гибридік тәсілдерді енгізуге мүмкіндік береді: сөздік, машиналық оқыту үлгілерін және нейрондық желілерді пайдалану.

Қазіргі уақытта бірқатар шетелдік аналитикалық платформалар бар. Олардың ішінде Sproutsocial, Hubspot, Buzzsumo, Hootsuite, IQBuzz, Brandmention және Snaplytics ерекшеленеді. Бұл платформалардың бизнес секторына бағытталғанына қарамастан ұқсас ерекшеліктері бар, бұл өмірдің әлеуметтік-саяси және экономикалық аспектілерін талдауды жеткіліксіз көрсетеді. Сондай-ақ, бұл платформалар негізінен ағылшын, испан, итальян, француз және т.б. сияқты ресурстарға бай тілдермен жұмыс істейді. Орыс және қазақ тілдеріндегі мәтіндердің берілуі өте шектеулі. Осыған байланысты Қазақстанның әлеуметтік медиа кеңістігін қадағалап, талдайтын және елде болып жатқан түрлі өзекті мәселелерге үлкен көңіл бөлетін пікір мониторингі жүйесі (OMSsystem) ақпараттық жүйесі әзірленді. OMSsystem жетекші қазақстандық жаңалықтар порталдары мен Facebook, Vkontakte, Instagram, Twitter және YouTube сияқты әлеуметтік желілерге талдау жасайды. OMSsystem жүйесінің негізгі элементі әлеуметтік көңіл-күйді талдау модулі болып табылады, ол тондық сөздіктерді, машиналық оқыту үлгілерін, нейрондық желілерді және әлеуметтік көңіл-күй маркетингінің көрсеткіштерін пайдалана отырып, деректерді талдау әдісін пайдаланады.

Машиналық оқыту үлгілерінде көңіл-күйді автоматты түрде анықтау үшін қазақстандық жаңалықтар порталдары мен әлеуметтік желілердегі 132 мың мәтіннің жинақталған дерекқоры пайдаланылды. Мәтіндер ең жақсы жіктеу нәтижелерін алу үшін *tf-idf* метрикасын және FastText сөздерді енгізу әдісін және классты теңестіруді қолдану арқылы алдын ала өңдеу, мүмкіндіктерді шығару кезеңдерін өтті. Классификация кезеңінде машиналық оқытудың бірқатар ең танымал алгоритмдері қолданылды (Support vector machine – SVM, Logistic regression – LR, Decision tree – DT, Random forest – RF, Naïve Bayes – NB, k-nearest neighbors – k-NN, және XGBoost) және нейрондық желілер (Deep Neural

networks – DNN, Convolutional neural networks – CNN және Recurrent neural networks – RNN) [36-40]. Жіктеу нәтижелері алгоритмдердің тиімділігін бағалауға арналған көрсеткіштердің жиынтық кестелері түрінде ұсынылған: дұрыстық, дәлдік, толықтық және F-өлшемі, қисық графиктер (Area under curve – Receiver operating characteristics – AUC–ROC) және қателік матрицалары. Қоғамның әлеуметтік көңіл-күйін талдау үшін әлеуметтік желілердегі маркетингтік көрсеткіштерді қолдану арқылы модельдер әзірленді: қоғамдағы тақырыпқа қызығушылық деңгейі, тақырыпты талқылау белсенділігі және әлеуметтік көңіл-күй.

Әзірленген модельдердің тиімділігі Covid-19-ға қарсы вакцинация тақырыбына эксперимент жүргізу арқылы бағаланды. Жиынтық талдау халықтың вакцинация науқанына, егу саясаты мен үкіметтің іс-әрекеттері мен пандемиямен күресу әдістеріне деген көзқарасын көрсетті. Келесі қадам OMSystem платформасы арқылы алынған деректерді талдайтын қосымша болып табылатын электрондық electronic Social Mood (eSM) модулін әзірлеу болды.

Диссертациялық жұмыстың мақсаты – машиналық оқыту үлгілерін, нейрондық желілерді және маркетингтік технологияларды пайдалана отырып, Қазақстан Республикасының медиа кеңістігінде қоғамның әлеуметтік көңіл-күйін бағалау әдістемесін әзірлеу.

Зерттеу мақсаттары:

1. OMSystem интеллектуалдық жүйесінің архитектурасы мен функционалдығын талдау.
2. OMSystem жүйесінің машиналық оқыту үлгілерін, нейрондық желілерді және маркетингтік технологияларды пайдалана отырып, Қазақстан Республикасының медиа кеңістігінде қоғамның әлеуметтік көңіл-күйін талдау және бағалау модулін әзірлеу.
3. Қарсы вакцинация тақырыбын талдау мысалында құрастырылған модульді бағалау Covid-19.
4. OMSystem жүйесін пайдалану арқылы алынған мәліметтерді талдайтын және қоғамның әлеуметтік көңіл-күйін бағалауды жүзеге асыратын electronic Social Mood (eSM) модулін әзірлеу.

Зерттеу нысаны: мәтіндік деректер, басылымдар, жаңалықтар ресурстары, Қазақстан Республикасының әлеуметтік медиа кеңістігі.

Зерттеу әдістері: Data mining, Web mining, Табиғи тіл өңдеу (NLP), Сентимент талдау (Sentiment Analysis), Машиналық оқыту, Нейрондық желілер, Әлеуметтік аналитикалық маркетингтік технологиялар.

Зерттеудің теориялық маңыздылығы: OMSystem интеллектуалдық жүйесінің архитектурасын және функционалдығын дамытуды талдау, қоғамның әлеуметтік көңіл-күйін бағалау модулінің тиімділігін бағалау.

Зерттеудің практикалық маңыздылығы: OMSystem интеллектуалдық жүйесінің әзірленген деректерді өңдеу және талдау модулін пайдалана отырып, қоғамның әлеуметтік көңіл-күйін талдау.

Жүргізілген зерттеулер мен алынған нәтижелердің ғылыми жаңалығы :

1. Әлеуметтік көңіл-күйді талдау әдісі әзірленді, ол машиналық оқыту үлгілерін қолданумен және пайдаланушының тақырыпқа қызығушылығының маркетингтік көрсеткіштерімен, тақырыпты талқылаудағы белсенділікпен және әлеуметтік көңіл-күй деңгейімен ерекшеленеді.
2. Әлеуметтік көңіл-күйді талдауға арналған интеграцияланған оқыту моделі әзірленді, оның ішінде машиналық оқытудың жеті атрибут үлгісі, сондай-ақ терең оқытудың төрт моделі.
3. Қазақ тілінің тоналды сөздігі әзірленді, ол әлеуметтік көңіл-күйді талдаудың біріктірілген үлгісі үшін қолданылады.

Қорғаныс ережелері:

1. Машиналық оқыту үлгілері мен маркетингтік көрсеткіштерді пайдалана отырып, қоғамның әлеуметтік көңіл-күйін талдаудың әзірленген әдісі, ол әртүрлі әлеуметтік-саяси тақырыптарды және ағымдағы үкіметтік науқандарға пайдаланушылардың реакцияларын бағалауға мүмкіндік береді.

2. Машиналық оқытудың жеті атрибут үлгісін, сондай-ақ терең оқытудың төрт моделін қамтитын әлеуметтік көңіл-күйді талдау үшін әзірленген интеграцияланған оқыту моделі.

3. Әлеуметтік көңіл-күйді талдау әдісі арқылы қоғамның көзқарасы мен үкімет қызметін көрсететін Covid-19 вакцинасы тақырыбы бойынша эксперимент нәтижелері.

Жұмыстың көлемі мен құрылымы. Диссертациялық жұмыс 160 беттен тұрады және 69 сурет пен 30 кестеден тұрады. Мазмұны 6 бөлімнен тұрады.

Кіріспеде диссертациялық жұмыстың өзектілігі, мақсаты, міндеттері, зерттеу объектілері мен әдістері, теориялық және практикалық маңыздылығы, сонымен қатар жаңашылдығы сипатталған.

Бірінші бөлімде әлеуметтік желілерді мониторингілеуге арналған ақпараттық-талдамалық жүйелердің негізгі аспектілері сипатталған, әлеуметтік медиа кеңістігін бақылау мен талдауға арналған шетелдік және отандық платформалар егжей-тегжейлі қарастырылып, олардың артықшылықтары мен кемшіліктері көрсетілген. Мәтіндердің тоналдылығын және әлеуметтік көңіл-күйді анықтаудың негізгі әдістерінің сипаттамасы берілген: сөздік тәсілі, машиналық оқыту әдістері, нейрондық желілер және әлеуметтік аналитикалық маркетингтік технологиялар.

Екінші бөлімде OMSystem әзірленген аналитикалық платформа егжей-тегжейлі берілген. Ол Қазақстан Республикасы медиа-кеңістігінің әлеуметтік желілері мен жаңалықтар порталдарын кеңейтілген талдау және мониторингі бойынша маманданған. OMSystem құрамына орыс және қазақ тіліндегі тон сөздіктері, машиналық оқыту және нейрондық желі үлгілері, сондай-ақ қоғамның әлеуметтік көңіл-күйі мен әл-ауқатын модельдеу және анықтау құралдары кіреді. Бұл бөлімде сонымен қатар диссертациялық жұмыстың ең маңызды бөлігі болып табылатын мәтіндердің екілік және көп класты классификациясының үлгілерін жасау қарастырылған. Нәтижелер графиктер, жиынтық кестелер және қорытындылар түрінде берілген. Әлеуметтік желілерде маркетингті басқару әдістеріне негізделген модельдерді әзірлеу ұсынылған, олар берілген тақырыптар бойынша қоғамның әлеуметтік көңіл-күйінің көрсеткіштерін анықтауға мүмкіндік береді.

Үшінші бөлімде Covid-19-ға қарсы вакцинацияға қатысты қоғамның әлеуметтік көңіл-күйін талдауға бағытталған эксперимент жүргізілді. Бұл тақырып әлемде пандемияның тез таралуына байланысты ерекше танымалдыққа ие болды. Ол жаңалықтар ресурстары мен әлеуметтік желілерде қызу талқыланды, осы тақырыпқа арналған жазбалардың астына мындаған пікірлер жазылды. Пайдаланушылардың пікірлерін бағалау тондық сөздіктерді, машиналық оқыту үлгілерін және нейрондық желілерді және маркетингтік технологияларды пайдалана отырып жасалды.

Төртінші бөлімде OMSystem платформасы арқылы алынған деректерді талдайтын қолданба болып табылатын Django Python құрылымында әзірленген электрондық electronic Social Mood (eSM) модулі ұсынылған. Бұл модуль келесі негізгі функцияларды орындайды: қоғамның әлеуметтік көңіл-күйін талдауға арналған тақырыптардың негізгі категорияларын құру, әрбір тақырып бойынша сандық деректерді алу, қоғамдағы тақырыпқа қызығушылық деңгейін есептеу, тақырыпты талқылау белсенділігі және әлеуметтік көңіл-күй, графиктер мен жиынтық кестелер түрінде алынған нәтижелерді көрнекі түрде көрсету.

Қорытында осы диссертациялық жұмыстың теориялық және практикалық нәтижелері жинақталып, оның машиналық оқыту үлгілері мен нейрондық желілерді

пайдалана отырып, қоғамның көңіл-күйін талдаудағы маңызды аспектілері және әлеуметтік көңіл-күй көрсеткіштері келтірілген.

Зерттеушінің жеке үлесі. Диссертацияда әлеуметтік медиа кеңістігін мониторингілеуге арналған қолданыстағы платформалар, OMSystem аналитикалық платформасының архитектурасы мен функционалдығы талданды; деректерді өңдеу және талдау модулі машиналық оқыту үлгілерін, нейрондық желілерді және әлеуметтік аналитикалық маркетингтік технологияларды пайдалана отырып әзірленді; Сонымен қатар, қоғамның әлеуметтік көңіл-күйін бағалайтын electronic Social Mood (eSM) модулі әзірленді.

Ғылыми нәтижелердің негізділігі мен сенімділік дәрежесі. Диссертацияның нәтижелері 12 ғылыми мақалада ұсынылды, оның ішінде кітаптағы 2 мақала және 1 тарау Scopus деректер базасында рецензияланған журналдар мен кітаптар сериясында, 4 мақала Білім беру сапасын қамтамасыз ету комитеті ұсынған журналдарда жарияланды. және Қазақстан Республикасы Білім және ғылым министрлігінің және 2 мақала – ғылыми конференцияларда, Scopus деректер базасында рецензиядан өткен және 3 мақала – халықаралық конференциялар материалдарында:

1. Karyukin, V., Mutanov, G., Mamykova, Z., Nassimova, G., Torekul, S., Sundetova, Z. & Negri, M. On the development of an information system for monitoring user opinion and its role for the public. *Journal of Big Data* 9, 110 (2022). <https://doi.org/10.1186/s40537-022-00660-w>.

2. G. Mutanov, V. Karyukin and Z. Mamykova, "Multi-class sentiment analysis of social media data with machine learning algorithms," *Computers, Materials & Continua*, vol. 69, no.1, pp. 913–930, 2021. <https://doi.org/10.32604/cmc.2021.017827>.

3. Mutanov, G., Mamykova, Z., Karyukin, V., Yessenzhanova, S. The Approach to Building a Context-Dependent Sentiment Dictionary. In: Mutanov, G., Serikbekuly, A. (eds) *Digital Transformation in Sustainable Value Chains and Innovative Infrastructures. Studies in Systems, Decision and Control*, vol 443, 2022. Springer, Cham. https://doi.org/10.1007/978-3-031-07067-9_1.

4. Мутанов Г.М., Мамыкова Ж.Д., Карюкин В.И., Жақсыкелді А.Ж. Разработка машинно-обучаемого алгоритма определения тональности пользовательского восприятия контента. Вестник КазНУ Серия Технические Науки, Казахстан, 135 (5), 2019.

5. Alimzhanova L.M. Karyukin V.I. A classification model based on decision-making processes. Вестник КазНУ Серия Технические Науки, Казахстан, 138 (2), 2020.

6. Рахимова Д.Р., Тұрарбек А.Т., Карюкин В.И., Карибаева А.С., Тұрғанбаева А.О. Қазақ тіліне арналған заманауи машиналық аударма технологияларына шолу. Вестник КазНУ Серия Технические Науки, Казахстан, 141 (5), 2020.

7. Karibayeva A., Karyukin V.I., Turganbayeva A., Turarbek A. The translation quality problems of machine translation systems for the Kazakh language. *Journal of Mathematics, Mechanics and Computer Science*, Kazakhstan, vol. 111, n. 3, 2021.

8. Vladislav Karyukin, Aidana Zhumabekova, and Sandugash Yessenzhanova. 2020. Machine Learning And Neural Network Methodologies of Analyzing Social Media. In *Proceedings of the 6th International Conference on Engineering & MIS 2020 (ICEMIS'20)*. Association for Computing Machinery, New York, NY, USA, Article 9, 1–7. <https://doi.org/10.1145/3410352.3410739>.

9. Diana Rakhimova, Vladislav Karyukin, Aidana Karibayeva, Assem Turarbek, and Aliya Turganbayeva. 2021. The Development of the Light Post-editing Module for English-Kazakh Translation. In *The 7th International Conference on Engineering & MIS 2021 (ICEMIS'21)*. Association for Computing Machinery, New York, NY, USA, Article 69, 1–5. <https://doi.org/10.1145/3492547.3492651>.

10. Карюкин В., Есенжанова С. Построение контекстно-зависимого тонального словаря. Международная научная конференция студентов и молодых ученых «ФАРАБИ ӘЛЕМІ», Алматы, Казахстан, 2020.

11. Карюкин В. Подход к построению приложения eSM. Международная научная конференция студентов и молодых ученых «ФАРАБИ ӘЛЕМІ», Алматы, Қазақстан, 2020.

12. Карюкин В. Многоклассовая классификация с применением алгоритмов машинного обучения. Международная научная конференция студентов и молодых ученых «ФАРАБИ ӘЛЕМІ», Алматы, Қазақстан, 2021.

Диссертацияның зерттеу жұмыстарымен байланысы. Бұл зерттеу ғылыми және (немесе) ғылыми-техникалық қызмет нәтижелерін коммерцияландыру жобасы аясында «OMSsystem пікір мониторингінің ақпараттық жүйесі (Пікірлерді бақылау жүйесі)» 0101-18-ГК. (Диссертанттың негізгі рөлі әлеуметтік көңіл-күйді, машиналық оқыту модельдерін және нейрондық желілерді талдау және бағалау модулін әзірлеу, Covid-19-ға қарсы вакцинация тақырыбы бойынша қоғамдық көңіл-күйді талдау бойынша эксперимент жүргізу және electronic Social Mood модуль әзірлеу болды).