

**Тохтаунов Ильмурат Турдымагаметовичтің 8D06105 - «Деректер ғылымы» білім беру бағдарламасы бойынша философия докторы (PhD) ғылыми дәрежесін алу үшін ұсынылған «Мақсатты жарнамада "ұқсас аудиторияны" табуға арналған терең оқыту модельдері мен әдістері» тақырыбындағы диссертациялық жұмысына
АҢДАТПА**

Зерттеудің жалпы сипаттамасы

Бұл диссертациялық жұмыста жоғары өлшемді кестелік пайдаланушы деректерінің сызықтық емес өлшемін азайту және ұсыным үйрену әдістері зерттеліп, мақсатты жарнама жүйелерінде ұқсас аудиторияны анықтауға қолданылады. Зерттеуде классикалық сызықтық әдістер (PCA), сызықтық емес коллекторлық үйрену тәсілдері (t-SNE) және жинақталған автоэнкодерлерді пайдаланатын терең оқытуға негізделген ұсыным үйрену қарастырылады. Гетерогенді пайдаланушы белгі векторларын компактты 288 өлшемді латентті ендірулерге сығатын жинақталған автоэнкодер шеңбері ұсынылады; бұл міндетке тән модельді қайта оқытпай-ақ пайдаланушы ұқсастығын тиімді талдауға мүмкіндік береді. Шеңбер шамамен 900 000 жазылушыны және алты түрлі мінез-құлықтық субъект доменіне бөлінген 948 белгіні қамтитын ауқымды анонимдендірілген телекоммуникациялық деректер жиынтығында бағаланады.

Зерттеудің өзектілігі

Телекоммуникация, қаржы және цифрлық маркетинг салаларындағы өнеркәсіптік кестелік деректердің көлемі тез артып келеді. Мақсатты жарнама жүйелері жоғары өлшемділікке және гетерогенді белгі кеңістіктеріне ие ауқымды мінез-құлықтық деректер жиынтықтарына сүйенеді. Дәстүрлі машиналық оқыту желілері динамикалық пайдаланушы мінез-құлқымен және үздіксіз дамып отыратын деректер үлестірімдерімен масштабталуда қиындыққа тап болады.

Қолданыстағы ұқсас аудиторияны анықтау жүйелері, әдетте, қайталанатын оқыту мен қолмен жасалатын белгі инженериясын талап ететін науқанға тән классификаторларға негізделген. Жоғары өлшемді кестелік деректерде сенімді ұқсастықты модельдеуге жарамды бірыңғай ұсыным кеңістігі жоқ. Бастапқы белгілерге қолданылатын статикалық ұқсастық метрикалары деректердің гетерогенді геометриясына байланысты жиі сәтсіз болады. Ауысымды ұсынымдар мен бейімделгіш ұқсастық механизмдерін үйрене алатын терең оқыту модельдеріне қажеттілік бар.

Бұрынғы зерттеулер кестелік деректер үшін автоэнкодерге негізделген ұсыным үйренуге деген қызығушылықты көрсетті, алайда олардың гетерогенді телекоммуникациялық пайдаланушы деректеріне ұқсас аудиторияны модельдеу мақсатында жүйелі түрде қолданылуы, көп субъектілі ендіру стратегияларымен үйлесімде және классикалық әдістер мен нақты бизнес метрикаларына қарсы кешенді бағалаумен бірге, жеткілікті зерттелмеген. Бұл олқылық осы диссертацияның тақырыбы мен ғылыми мақсаттарын алдын ала анықтайды.

Зерттеудің мақсаты

Мақсатты жарнама жүйелерінде қолданылатын ауқымды гетерогенді кестелік деректерде “ұқсас аудиторияларды” анықтау үшін жалпылама ұсынымдар мен ұқсастық механизмдерін үйренуге арналған терең оқытудың модельдері мен әдістерін әзірлеу.

Зерттеудің міндеттері

Міндет 1. Семантикалық мазмұнды ұстауға арналған деректерді ұсыным техникаларын талдау: жоғары өлшемді гетерогенді кестелік деректердің түрлендірілген белгі кеңістігіндегі семантикалық құрылымды ұстауда қолданыстағы деректерді ұсыным техникаларының қолданылуын талдау.

Міндет 2. Сызықтық емес өлшемді азайтуға арналған терең оқыту моделі: гетерогенді кестелік деректерді сызықтық емес өлшемді азайту үшін жаңа терең оқыту моделін әзірлеу, бұл ретте авторлық шешімнің теориялық негіздемесін қамтамасыз ету.

Міндет 3. Латентті ендіру кеңістігіндегі үйренілген ұқсастық механизмдері: классификацияға негізделген ұқсас аудиторияны анықтаудың балама ретінде латентті ендіру кеңістігіндегі үйренілген ұқсастық механизмдерін зерттеу және олардың нақты жарнамалық науқандардағы тиімділігін бағалау.

Міндет 4. Масштабталатын өндірістік жүйе және доменаралық қолданылуы: ұсынылған шеңберді масштабталатын өндірістік жүйе ретінде іске асыру және оның бірнеше бизнес доменіндегі практикалық қолданылуын бағалау.

Зерттеу объектісі мен пәні

Зерттеу объектісі: мақсатты жарнама платформаларында қолданылатын пайдаланушы мінез-құлқын, құрылғы сипаттамаларын және контекстік белгілерді сипаттайтын ауқымды гетерогенді кестелік деректер жиынтықтары.

Зерттеу пәні: ұқсас аудиторияны анықтауға арналған көп субъектілі автоэнкодер ендірулері мен сиамдық нейрондық желілерді қоса алғандағы ұсыным үйрену мен ұқсастықты модельдеуге арналған терең оқытудың модельдері мен әдістері.

Зерттеу әдістері

Зерттеуде мынадай әдістер мен құралдар қолданылады:

- Өлшемді азайту әдістері: PCA (сызықтық базалық деңгей), t-SNE (сызықтық емес коллекторлық үйрену), жинақталған автоэнкодер (терең ұсыным үйрену);
- Ұсыным шеңбері: көп субъектілі ендіру стратегиясы - алты мінез-құлықтық домен бірыңғай 288 өлшемді латентті кеңістікке біріктіріледі;
- Ұқсастықты үйрену: нақты жарнамалық науқандардан алынған он/теріс пайдаланушы жұптарында оқытылған контрастты шығынға негізделген сиамдық желі;
- Төменгі ағымды бағалау: 17 тәуелсіз науқан деректер жиынтығы бойынша орташаланған kNN классификациясы;
- Бағалау метрикалары: Precision, Recall, F1-score, ROC AUC, Lift Top-1, Silhouette Score, Davies–Bouldin индексі, Calinski–Harabasz Score;
- Іске асыру стеги: Python, PyTorch, PyTorch Lightning, scikit-learn, Apache Spark, MLflow, DVC.

Ақпараттық база

Зерттеу телекоммуникациялық оператормен ынтымақтастық келісімі арқылы алынған ауқымды анонимдендірілген деректер жиынтығына негізделген. Деректер жиынтығы шамамен 900 000 жазылушының жазбаларын қамтиды, алдын ала өңдеуден кейін 2 814 бастапқы атрибуттан алынған 948 белгімен. Белгілер алты субъект доменіне бөлінген: Пайдаланушы субъектісі (~280 белгі: белсенділік, ARPU, демография, география), Веб субъектісі (~190 белгі: интернет белсенділігі мен қызығушылықтары), Қаржы субъектісі (~80 белгі: анонимдендірілген транзакция үлгілері), Құрылғы субъектісі (~120 белгі: аппараттық сипаттамалар мен баға деңгейлері), Базалық станция субъектісі (~180 белгі: желі сапасы мен геобелсенділік) және Тариф жоспары субъектісі (~98 белгі: қызмет жоспарының параметрлері мен бағасы). Тексеру оқыту кезінде болмаған сыртқы мақсатты айнымалылары бар 17 тәуелсіз жарнамалық науқан деректер жиынтығында жүргізілді, бұл бейтарап өнімділікті бағалауды қамтамасыз етіп, мақсатты ақпараттың ағып кетуін жойды.

Ғылыми даму бағыттарына және мемлекеттік бағдарламаларға сәйкестігі

Зерттеу нәтижелері Қазақстан Республикасының мынадай стратегиялық құжаттарына сәйкес келеді:

1. Жасанды интеллектті дамыту тұжырымдамасы 2024–2029 жылдарға (Қазақстан Республикасы Үкіметінің 2024 жылғы №592 қаулысы) - диссертация Тұжырымдамамен айқындалған ғылыми зерттеулердің басым бағытына жауап беріп, ауқымды деректер үшін машиналық оқыту мен ұсыным үйрену саласындағы ұлттық жасанды интеллект құзыреттіліктерін дамытуға үлес қосады;
2. Қазақстан Республикасы Президентінің «Жасанды интеллект дәуіріндегі Қазақстан: цифрлық трансформация арқылы қазіргі заман сын-қатерлері мен олардың шешімдері» атты жолдауы (2025 жылғы 8 қыркүйек) және оны іске асыру жөніндегі Ұлттық іс-қимыл жоспары - диссертация телекоммуникация мен цифрлық жарнама секторларындағы жасанды интеллектке негізделген цифрлық трансформацияның стратегиялық мақсатын қолдайды;
3. Қазақстан Республикасының «Жасанды интеллект туралы» Заңы (№230-VIII, 2025 жылғы 17 қараша) - диссертация Заңмен белгіленген жасанды интеллектті басқарудың құқықтық шеңберіне сәйкес экономиканың басым секторларында жасанды интеллект жүйелерін әзірлеу мен енгізуді ілгерілетеді;
4. Қазақстан Республикасының 2025 жылға дейінгі Ұлттық даму жоспары және 2023–2029 жылдарға арналған Цифрлық трансформация, Ақпараттық-коммуникациялық технологиялар саласын дамыту мен Киберқауіпсіздік тұжырымдамасы - диссертация цифрлық экономиканы дамытудың ұлттық басымдықтарына сәйкес телекоммуникация секторының деректерге негізделген цифрландырылуына үлес қосады.

Ғылыми жаңалығы

1. PCA, t-SNE және автоэнкодерге негізделген ұсыным үйренудің ауқымды гетерогенді кестелік деректердегі салыстырмалы талдауы жүргізілді. Автоэнкодер бағаланған барлық метрикалар бойынша жоғары семантикалық бөлінуді көрсетеді (Silhouette Score 0.48 PCA үшін 0.21-ге қарсы; Davies–Bouldin индексі 0.79 пен 1.84-ке қарсы; kNN F1 0.71 пен 0.56-ға қарсы).

2. Batch Normalization және LeakyReLU ($\alpha=0.2$) бар жаңа авторлық жинақталған автоэнкодер (948→1000→1000→288→1000→1000→948) ұсынылды - жоғары өлшемді гетерогенді телекоммуникациялық кестелік деректерге арналған осы типтегі алғашқы архитектура. Аралас BatchNorm–LeakyReLU белсендірулері бар ұсынылған 7 қабатты архитектура үшін кері таралу туралы Лемма формальды түрде дәлелденді.

3. Көп субъектілі ендірулерді (6 мінез-құлықтық домен → 288 өлшемді латентті кеңістік) және бейімделгіш үйренілген ұқсастық функциялары ретінде косинустық ұқсастығы бар сиамдық нейрондық желілерді біріктіретін ұқсастыққа негізделген парадигма әзірленді - науқанға тән классификаторларды бірыңғай ауысымды модельмен ауыстырып, 17 нақты жарнамалық науқан бойынша Lift Top 1 = 12.9, ROC AUC = 0.79, Conversion Rate = 0.36 мәндеріне қол жеткізілді.

4. 900 000 жазылушыны өңдейтін модульдік өндірістік архитектура (HDFS + DVC + MLflow + MinIO) ұсынылып, орнатылды; ай сайынғы бумалық жаңарту ~45 минутта, сұраныс бойынша инференс ~3 минутта орындалады, бұл өнеркәсіптік орнатуға практикалық жарамдылығын растайды. Телекоммуникация, электрондық коммерция, денсаулық сақтау және қаржы салаларындағы доменаралық қолданылуы дәлелденді.

Диссертациялық зерттеу барысында алынған ғылыми нәтижелер

- PCA, t-SNE және автоэнкодерге негізделген өлшемді азайтудың кешенді салыстырмалы талдауы ауқымды анонимдендірілген телекоммуникациялық деректер жиынтығында жүргізілді; автоэнкодер бағаланған барлық бес метрика бойынша екі әдісті де басып озды.
- 948→1000→1000→288→1000→1000→948 (~5.3M параметр) архитектурасы бар жинақталған автоэнкодер жасалып, оқытылып, тексерілді; тексеру MSE = 0.61 мәнімен 400 эпохадан кейін тұрақты жинақталуға қол жеткізілді (ерте тоқтату patience=20); 7 қабатты BatchNorm–LeakyReLU архитектурасы арқылы градиент таралуы туралы Лемма формальды түрде дәлелденді.
- Алты мінез-құлықтық доменді біріктіретін көп субъектілі ендіру стратегиясы әзірленді; Lift Top 1 = 11.7, ROC AUC = 0.76 және Conversion Rate = 0.31 мәндеріне қол жеткізілді - бұл ең жақсы дәстүрлі классификатормен (LightGBM, Lift = 6.6) салыстырғанда Lift бойынша 77.3% және бір субъектілі косинустық ұқсастықпен (Lift = 7.3) салыстырғанда 60% жақсарту.
- Алдын ала есептелген автоэнкодер ендірулерінде жұмыс істейтін сиамдық желілер 17 тәуелсіз нақты жарнамалық науқан бойынша Lift Top 1 = 12.9, F1 = 0.75, ROC AUC = 0.79 және Conversion Rate = 0.36 мәндеріне қол жеткізіп, барлық базалық деңгей әдістерін орта есеппен 41.6%-ға асып түсті; науқан ішіндегі конверсиялар ~17 100-ден (LightGBM) ~32 400-ге (Siamese) өсті.
- Модульдік өндірістік жүйе жасалып, эксперименттік тексерілді: ~900 000 жазылушыны қамтиды, 8 түйінді кластерде ай сайынғы бумалық жаңарту ~45 мин, сұраныс бойынша инференс сұрау басына ~3 мин.

Қорғауға шығарылатын негізгі қағидалар

1. Гетерогенді кестелік деректерді тиімді ұсынымды үйренуге арналған көп субъектілі терең оқыту шеңбері, барлық кластерлеу мен классификация сапа метрикалары бойынша сызықтық және параметрлік емес әдістерді басып озады: Silhouette Score 0.48 пен 0.21-ге қарсы (PCA); Davies–Bouldin индексі 0.79 пен 1.84-ке қарсы; kNN F1 0.71 пен 0.56-ға қарсы; Calinski–Harabasz Score 1020 пен 410-ға қарсы.
2. Автоэнкодерге негізделген латентті ендірулер (948→1000→1000→288, ~5.3M параметр) пайдаланушылардың мінез-құлықтық және контекстік сипаттамаларын ұстайтын құрылымдық ұсыным кеңістігін қалыптастырады; 400 эпохада тұрақты жинақталу расталды (MSE = 0.61) және BatchNorm–LeakyReLU 7 қабатты архитектурасы арқылы градиент таралуы туралы формальды дәлелденген Лемма арқылы теориялық негіздеме берілді.
3. Бейімделгіш үйренілген ұқсастық функциясы ретінде көп субъектілі автоэнкодер ендірулерінде жұмыс істейтін косинустық ұқсастығы бар сиамдық нейрондық желі, барлық метрикалар бойынша науқанға тән campaign-specific классификаторларды басып озады: Lift Top 1 = 12.9, F1 = 0.75, ROC AUC = 0.79, CR = 0.36 (17 нақты науқан бойынша SVM, Random Forest, LightGBM-нен орта есеппен +41.6%).

Зерттеудің теориялық маңыздылығы

Бұл диссертация жоғары өлшемді гетерогенді кестелік деректерге қолданылатын терең ұсыным үйренудің теориялық түсінігін дамытады. Batch Normalization және LeakyReLU белсендірулері бар терең жинақталған автоэнкодер арқылы кері таралу туралы Лемманың формальды дәлелі ұсынылған авторлық архитектураны қатаң негіздемемен қамтамасыз ететін түпнұсқа теориялық үлесті білдіреді. PCA, t-SNE және автоэнкодерге негізделген өлшемді азайту әдістерін бағалауға арналған жүйелі салыстырмалы шеңбер, гетерогенді деректердің бірыңғай ұсынымына арналған көп субъектілі ендіру стратегиясының формализациясы және науқанға тән классификаторларға балама ретіндегі ұқсастық парадигмасының теориялық негізі бірігіп кестелік деректер үшін ұсыным үйрену саласына жаңа әдіснамалық компоненттерді қосады.

Зерттеудің практикалық маңыздылығы

Зерттеу науқанға тән модельді қайта оқыту қажеттілігін жоятын толық автоматтандырылған жалпыланған ұқсас аудитория қызметін ұсынып, орнатады. Телекоммуникациялық операторда жұмыс істейтін өндірістік жүйе 17 тексерілген нақты жарнамалық науқан бойынша B2B клиенттеріне қызмет көрсетеді. Conversion Rate 0.19-дан (LightGBM) 0.36-ға (сиамдық желі) жақсарды - 89% жетілдіру - ал науқан ішіндегі конверсиялар 90 000 жазылушылар пулынан ~17 100-ден ~32 400-ге өсті, қосымша шығынсыз. Аналитик еңбегі толығымен жойылды: қолмен жасалатын белгі инженериясы немесе науқан ішіндегі модельді оқыту талап етілмейді. Әдіснама домен-агностикалық болып табылады және электрондық коммерцияда (ұқсас сатып алушыларды мақсатты белгілеу), денсаулық сақтауда (ұқсас пациенттерді профильдеу) және қаржыда (транзакциялық ендірулерден тұтынушыларды сегменттеу) қолданылады. Қазақстан Республикасында №69310 авторлық куәлік тіркелген.

Таратылуы, жарияланымдар және докторанттың әрбір жарияланымға қосқан үлесі

Негізгі нәтижелер Халықаралық ақпараттық технологиялар университетінің Математикалық және компьютерлік модельдеу кафедрасында (2022–2026) және Нархоз Университетінің Цифрлық технологиялар мектебінде (2025–2026) ұсынылып, талқыланды. Диссертация тақырыбы бойынша жалпы 5 жарияланым шығарылды:

1. Tokhtakhunov I., Nurtas M., Neftissov A., Pirnaev S., Kazambayev I., Kirichenko L. Exploring Autoencoder-Based Representations for Tabular Data Classification. *Engineered Science*. 2025, vol. 37, 1703. DOI: 10.30919/es1703. [Scopus Q1, CiteScore: 10.2, Percentile: 88%] - Докторант зерттеу тұжырымдамасын дербес қалыптастырды, жинақталған автоэнкодер архитектурасын жасап іске асырды, телекоммуникациялық деректер жиынтығында барлық эксперименттерді жүргізді, статистикалық талдау жасады және қолжазбаны дайындады. Авторлас авторлар нәтижелерді талқылауға және редакциялауға үлес қосты.
2. Tokhtakhunov I., Altaibek A., Nurtas M. Optimizing Similar Audience Search in Targeted Advertising: Effectiveness of Siamese Networks for Autoencoder-Based User Embeddings. *Engineering, Technology & Applied Science Research*. 2025, vol. 15, no. 3, pp. 23367–23375. DOI: 10.48084/etasr.10527. [Scopus Q2, CiteScore: 2.8, Percentile: 56%] - Докторант сиамдық желі архитектурасын дербес жасады, оқыту процедурасы мен контрастты шығын шеңберін іске асырды, 17 нақты жарнамалық науқан бойынша бағалау жүргізді және қолжазбаны жазды. Авторлас авторлар жарнама метрикалары бойынша кеңес беріп, қолжазбаны шолды.
3. Tokhtakhunov I., Nurtas M., Altaibek A., Kozhamzharova D., Aitimov M. The Efficacy of Autoencoders in the Utilization of Tabular Data for Classification Tasks. *Procedia Computer Science*. 2024, vol. 238, pp. 492–502. DOI: 10.1016/j.procs.2024.06.052. [Scopus Q2, CiteScore: 3.6, Percentile: 62%] - Докторант эксперименттік жоспарды басқарды, барлық базалық деңгей мен автоэнкодер модельдерін іске асырды, kNN бағалауын жүргізді және қолжазбаны дайындады. Авторлас авторлар деректерді алдын ала өңдеуді талқылауға және қолжазбаны шолуға үлес қосты.
4. Tokhtakhunov I., Nurtas M. Nonlinear Dimensionality Reduction for Lookalike Audience Detection: From Manifold Learning to Autoencoder-Based Representations. *Journal of Problems in Computer Science and Information Technologies*. 2026, 4(1). DOI: 10.26577/jpcsit4120268. - Докторант PCA, t-SNE және автоэнкодер әдістерінің салыстырмалы талдауын дербес жүргізді, теориялық шеңберді формализациялады және қолжазбаны толығымен жазды. Ғылыми жетекші қолжазбаны шолып, мақұлдады.
5. Қазақстан Республикасының №69310 авторлық куәлігі. Мақсатты жарнама жүйелерінде ұқсас аудиторияны іздеуге арналған терең оқыту бағдарламалық моделі / Tokhtakhunov I., Nurtas M. Өтінім 29.03.2026; Жарияланды 31.03.2026. - Докторант автоэнкодерді оқыту желісін, көп субъектілі ендіру модулін, сиамдық желінің инференс модулін және модульдік өндірістік архитектураны (HDFS + DVC + MLflow + MinIO) қоса алғандағы толық бағдарламалық жүйені жасап, іске асырды. Ғылыми жетекші куәлік өтінімінің авторлас авторы болды.

Тараулар бойынша шолу

1-тарау өлшемділік қарғысын, классикалық және сызықтық емес өлшемді азайту әдістерін (PCA, t-SNE, LLE, Isomap, UMAP), ұсыным үйренуге арналған автоэнкодер архитектураларын, ұқсас аудиторияны модельдеу тәсілдерін және сиамдық желі әдістерін қамтитын кешенді әдебиеттерге шолу береді. Тарауда зерттеу олқылығы анықталып, ұсынылған тәсілдің теориялық негізі қалыптастырылады.

2-тарау деректер жиынтығын (алты субъект доменіндегі 2 814 бастапқы атрибуттан алынған 948 белгімен шамамен 900 000 жазылушы) және толық алдын ала өңдеу желісін сипаттайды: бір жылу кодтауы, жоқ мәндерді толтыру (MICE), Isolation Forest арқылы аномалияларды анықтау, көпколлинеарлықты азайту (шек 0.77) және min-max белгі нормализациясы.

3-тарау ұсынылған жинақталған автоэнкодер архитектурасын (948→1000→1000→288→1000→1000→948), формальды дәлелденген градиент таралуы туралы Лемманы, көп субъектілі ендіру стратегиясын және латентті кеңістіктегі косинустық ұқсастықты есептеуді қоса алғандағы барлық өлшемді азайту әдістерінің математикалық тұжырымын ұсынады.

4-тарау автоэнкодер ендірулерінде бейімделгіш ұқсастықты үйренуге арналған сиамдық желі архитектурасын ([256, 512, 224, 160, 128] салмақты бөлісетін тармақтар) сипаттайды; 17 науқанның бастапқы аудиторияларын пайдалана отырып жұп құру процедурасы мен есептеу ауыстырылымы талдауы қамтылған.

5-тарау модульдік өндірістік жүйе архитектурасын ұсынады: HDFS таратылған сақтауы, DVC нұсқаны басқару, MLflow эксперименттерді қадағалау, MinIO артефакт сақтауы және екі инференс режимі (таратылған Spark-негізді ай сайынғы бумалық жаңарту ~45 мин; жергілікті Pandas инференсі ~3 мин).

6-тарау барлық эксперименттік нәтижелерді ұсынады: өлшемді азайту әдістерінің көрнекі және сандық салыстырмасы, кластерлеу сапасын бағалау (3-кесте), 17 науқан бойынша ұқсас аудиторияны анықтау өнімділігі (4–5-кестелер), сиамдық желінің шатасу матрицасын талдау (6-кесте). Талқылау бөлімінде нәтижелер контекстке орналастырылып, шектеулер талданады және болашақ зерттеу бағыттары белгіленеді. Қорытынды қорғауға шығарылған барлық төрт қағиданы растайды.

Авторлық жеке үлесі

Диссертацияда сипатталған барлық негізгі нәтижелер автормен дербес орындалды. Автор мынадай жұмыстарға жауапты болды: зерттеу тұжырымдамасын әзірлеу; деректер жиынтығын алу және алдын ала өңдеу желісін жобалау; жинақталған автоэнкодер архитектурасын жасау мен іске асыру және градиент таралуы туралы Лемманы формальды дәлелдеу; көп субъектілі ендіру стратегиясын әзірлеу; сиамдық желіні жобалау және оқыту; толық өндірістік жүйені (HDFS + DVC + MLflow + MinIO) бағдарламалық іске асыру; 17 нақты жарнамалық науқан бойынша эксперименттік бағалау; статистикалық талдау; нәтижелерді түсіндіру; және барлық қолжазбаларды дайындау. Диссертацияға қатысты барлық жарияланымдарда автор жетекші рөл атқарды.

Диссертацияның құрылымы мен көлемі

Диссертация кіріспені, алты негізгі тарауды, талқылауды, қорытындыны, пайдаланылған әдебиеттер тізімін және қосымшаны қамтиды. Негізгі мәтіннің жалпы көлемі қосымшаларды есептемегенде 98 бетті құрайды. Диссертацияда 17 сурет, 5 кесте және 97 библиографиялық сілтеме бар.