АННОТАЦИЯ

“8D06102 – Компьютерлік ғылымдар” ББББ бойынша философия докторы (PhD) ғылыми дәрежесін алу үшін диссертацияға

БЕРЛІҚОЖА БАУЫРЖАН ӘСЕТҰЛЫ

Академиялық және мінез-құлық деректеріне негізделген жасанды интеллект әдістері арқылы IT мамандықтары студенттерінің мансаптық траекториясын болжау

Жұмыстың өзектілігі: Жаһандық еңбек нарығы цифрлық трансформацияның қарқынды үдерісінің нәтижесінде түбегейлі өзгерістерге ұшырады. Бұл жағдай икемді техникалық және талдамалық қабілеттерге ие жаңа буын мамандарын даярлауды қажет етті. Өнімділікті арттыру және инновацияны ынталандыру мақсатында әлем елдеріндегі кәсіпорындар автоматтандыру, деректерді талдау және жасанды интеллект (ЖИ) технологияларына барған сайын көбірек сүйеніп келеді. Соның салдарынан білім беру саласына технологиялық дамыған экономиканың өзгермелі талаптарына сәйкес оқыту мен кәсіби бағдарлау тәсілдерін бейімдеу бойынша қысым күшейіп отыр. Дүниежүзілік экономикалық форумның «Future of Jobs Report» (2023) есебіне сәйкес, автоматтандыру нәтижесінде шамамен 85 миллион жұмыс орны жойылуы мүмкін, алайда цифрлық экономикаға барынша бейімделген 97 миллион жаңа жұмыс орны құрылады. Мұндай парадигмалық өзгеріс оқушыларды әлі толық қалыптаса қоймаған мамандықтарға дайындау үшін деректерге негізделген, жекелендірілген тәсілді қажет етеді.

Осы тұрғыда ЖИ технологияларына негізделген кәсіби бағдарлау жүйелері университет пен еңбек нарығы арасындағы білім алшақтығын жоюдың тиімді құралы ретінде кеңінен танымал бола бастады. Мұндай жүйелер машиналық оқыту мен болжамды талдауды пайдалана отырып, студенттің жеке ерекшеліктері, мінез типі, әлеуетті мансаптық бағыттары мен оқу жетістіктері арасындағы өзара байланыстарды анықтайды. Соңғы зерттеулер көрсеткендей, жасанды интеллект дәстүрлі кәсіби кеңес беру әдістерімен салыстырғанда анағұрлым дәл және жекелендірілген ұсыныстар бере алады, бұл өз кезегінде жұмысқа орналасу мен кәсіби қанағаттану деңгейін арттырады. Бұдан бөлек, оқу талдауы (Learning Analytics - LA) мен білім беру деректерін өндіру (Educational Data Mining - EDM) бағыттарына деген қызығушылықтың артуы әлемдік ғылыми қауымдастықтың білім беру нәтижелерін жетілдірудегі ЖИ-дің әлеуетін мойындағанын көрсетеді. Оңтүстік Корея, Сингапур және Финляндия секілді мемлекеттер оқу бағдарламаларын бейімдеу мен студенттердің үлгерімін болжауға арналған ЖИ-негізделген шешімдерді білім беру жүйесіне енгізіп, кадрлық әлеуетті дамытудың тиімді тетіктерін

қалыптастырды.

Алайда, бұл технологияларды әртүрлі географиялық аймақтарда іске асыру деңгейінде елеулі айырмашылықтар бар. Дамушы елдерде деректер инфрақұрылымының жеткіліксіздігі, ЖИ саласындағы мамандардың тапшылығы және білім беру жүйелеріне деректерге негізделген шешімдерді интеграциялаудың әлсіздігі сияқты жүйелік кедергілер кездеседі. Ал жоғары табысты елдерде мықты деректер экожүйесі мен инновацияны қолдайтын институционалдық тетіктердің болуы айқын артықшылық береді.Нәтижесінде білім беру саласында «ЖИ қабылдау алшақтығы» (AI adoption gap) артып, бұл ұлттық бәсекеге қабілеттілік пен түлектердің еңбек нарығына бейімделу мүмкіндіктері арасындағы теңсіздікті күшейтуі ықтимал.

Осы зерттеу жұмысы дамушы экономикалардың контексіне бейімделген мансаптық траекторияны болжаудың ЖИ-негізделген моделін әзірлеу арқылы аталған алшақтықты қысқартуға тікелей үлес қосады.

Зерттеу жұмысының мақсаты: Бұл диссертациялық зерттеу жұмысының мақсаты – ақпараттық технологиялар (IT) мамандығында оқитын студенттердің академиялық және мінез-құлықтық деректеріне негізделген жасанды интеллект жүйесін әзірлеу арқылы олардың болашақ кәсіби траекториясын болжау болып табылады.

Зерттеу тапсырмалары: Бұл диссертациялық зерттеу бес негізгі мақсатқа негізделеді:

1) Академиялық, мінез-құлықтық және мотивациялық көрсеткіштерді қамтитын көпөлшемді студенттік деректерді жинау және алдын ала өңдеу.
2) Жасанды интеллектке негізделген мансаптық траекторияны болжау тәсілдерін зерттеу және талдау.
3) ЖИ технологиясына негізделген мансаптық траекторияны болжау қосымшасын жобалау.
4) Ұсынылған модельдің дәлдігін, есептеу тиімділігін және жадты пайдалану көлемін нақты университет деректері негізінде бағалау.
5) Мансаптық траекторияны болжауға арналған оңтайландырылған Колмогоров–Арнольд желісін (Kolmogorov–Arnold Network, KAN) зерттеу.

Бірінші тапсырма бойынша, Сүлеймен Демирел университетінің IT бағыты бойынша оқитын 692 студенттің деректері негізінде академиялық жетістіктер, мінез-құлық ерекшеліктері және кәсіби бейімділік арасындағы байланыстар талданды. Әрбір айнымалы GPA (орташа балл), пәндік үлгерім, жобаларға қатысу және мотивациялық деңгей студенттердің мамандану нәтижелерімен (мысалы, бағдарламалық қамтамасыз ету

әзірлеу, деректер ғылымы немесе жасанды интеллект бағыттары) арасындағы үрдістерді анықтау үшін зерттелді.

Екінші тапсырма бойынша, академиялық, мінез-құлықтық және мотивациялық деректер негізінде IT студенттерінің мансаптық бағыттарын болжауға арналған ең тиімді және дәл модельді анықтау үшін сегіз машиналық оқыту алгоритмі салыстырылды: K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Naïve Bayes (NB), Decision Tree (DT), Multi-Layer Perceptron (MLP) және TabTransformer. Әрбір модель 10 есе кросс-валидация әдісі арқылы оқытылып және тексерілді, бұл нәтижелердің статистикалық тұрақтылығы мен сенімділігін қамтамасыз етті. Талдау нәтижесінде Gradient Boosting ең жоғары өнімді эталондық модель ретінде танылды. Кейін зерттеудің екінші кезеңінде ол оңтайландырылған Колмогоров–Арнольд желісі (KAN) архитектурасымен салыстырылды. Осылайша, қарапайым машиналық оқыту әдістерінен күрделі нейрожелілік модельдерге біртіндеп көшу әдістемелік көпір қалыптастырды. Бұл дәлдік, түсіндіру мүмкіндігі және есептеу тиімділігі тұрғысынан айтарлықтай жетілдірулерге қол жеткізудің кезеңдік эволюциясын көрсетті.

Үшінші тапсырма бойынша, зерттеу шеңберінде ЖИ-негізделген мансаптық траекторияны болжау қосымшасы әзірленді. Қосымша студенттердің академиялық, мінез-құлықтық және мотивациялық деректерін автоматты түрде өңдеп, болжау қозғалтқышы арқылы нақты нәтижелерді шығарады және оларды интерактивті бақылау тақтасы арқылы студенттер мен кеңесшілерге ыңғайлы түрде ұсынады. Модульдік API-интерфейстер мен деректер қауіпсіздігі қағидаларына сүйенген бұл жүйе нақты уақытта жекелендірілген мансаптық ұсыныстарды ұсынуға және әртүрлі университеттерге бейімделуге мүмкіндік береді. Осылайша, теориялық зерттеу нәтижесі практикалық, масштабталатын шешімге айналды.

Төртінші тапсырма бойынша, Сүлеймен Демирел университетінің нақты студенттік деректері негізінде ұсынылған модельдің жадты пайдалану, есептеу тиімділігі және дәлдік тұрғысынан өнімділігі бағаланды. Оңтайландырылған қабатты Колмогоров - Арнольд желісі (KAN) көпөлшемді деректер жиынтығында кеңінен сынақтан өтті және дәстүрлі әдістермен MLP, Random Forest, Gradient Boosting салыстырылды. Бағалау ROC–AUC, precision, recall және орындау уақыты сияқты статистикалық және есептеу көрсеткіштері бойынша жүргізілді. Зерттеу нәтижелері ұсынылған модельдің жоғары дәлдікке аз ресурспен қол жеткізетінін және білім беру ортасында нақты қолдануға жарамды екенін көрсетті.

Бесінші мақсат бойынша, зерттеу барысында жоғары білім беру жүйесінде мансаптық траекторияны болжауға арналған Колмогоров – Арнольд желісінің (KAN) құрылымы жасалып, жетілдірілді және эксперименттік тұрғыдан тексерілді. Модельдің үйрену қабілетін,

жалпылау деңгейін және түсіндіру мүмкіндігін арттыру мақсатында ішкі және сыртқы қабаттары кеңейтілген оңтайлы қабатты KAN архитектурасы ұсынылды. KAN моделі Gradient Boosting, Random Forest және MLP сияқты дәстүрлі модельдермен салыстырылып, дәлдік, оқыту уақыты және жад тиімділігі тұрғысынан бағаланды. Нәтижелер көрсеткендей, жетілдірілген KAN моделі есептеу шығындарын едәуір төмендете отырып, болжау дәлдігін айтарлықтай арттырады. Бұл тәсіл жеке ерекшеліктерге бейімделген, деректерге негізделген кәсіби бағдарлаудың инновациялық және тиімді моделі ретінде өзін көрсетті.

Зерттеу әдістері: Бұл зерттеу жұмысы Сүлеймен Демирел университетінің (SDU) 692 IT студентінің анонимдендірілген деректерін пайдаланды. Деректер студенттердің академиялық үлгерімін, тұлғалық ерекшеліктерін, қосымша (extracurricular) белсенділіктерін және иемденген сертификаттарын қамтыды. Мәліметтер Beta Career Platform платформасынан жиналды — бұл SDU университетінің студенттерге тағылымдамалар мен бастапқы деңгейдегі жұмыс орындарын табуға көмектесетін ішкі цифрлық жүйесі. Платформа 692 студент бойынша сенімді, өкілді және практикалық тұрғыдан маңызды деректер ұсынды, бұл болжау модельдерін құруға берік эмпирикалық негіз қалыптастырды.

Деректер алдын ала өңделіп, жетіспейтін мәндер түзетілді, категориялық айнымалылар кодталды, сипаттамалар нормализацияланды, және маңызды белгілер (predictors) таңдалды. Бірнеше машиналық оқыту модельдері 10 есе кросс-валидация әдісімен сыналды және олардың өнімділігі accuracy, precision, және recall көрсеткіштері бойынша бағаланды. Қорытынды тексеру үшін дербес тест жиынтығы қолданылды. Барлық деректер этикалық нормаларға сай жиналды — анонимдік сақталды, институттық рұқсат алынды және қатысушылардың ақпараттандырылған келісімі қамтамасыз етілді.

Модельді оқытуға дейін деректерді кең көлемде алдын ала өңдеу жүргізілді. Сандық айнымалылардағы жетіспейтін мәндер орташа мәнмен (mean imputation) толтырылды, ал категориялық айнымалылар үшін мода бойынша толықтыру (mode imputation) тәсілі қолданылды. Тұлға типтері мен қосымша белсенділіктер сияқты категориялық белгілер one-hot encoding әдісі арқылы сандық форматқа ауыстырылды, бұл оларды машиналық оқыту алгоритмдерімен үйлесімді етті. Үздіксіз айнымалылар Min–Max масштабтау (scaling) арқылы нормализацияланып, барлық белгілердің салыстырмалылығын қамтамасыз етті. Бұған қоса, корреляциялық талдау және Recursive Feature Elimination (RFE) әдістері арқылы ең ақпараттық белгілер таңдалды, бұл модельдің тиімділігін арттырып, деректердегі артық шуды азайтты.

Болжамдық модельді құру үшін келесі машиналық оқыту алгоритмдері пайдаланылды: Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Gradient Boosting, Naïve Bayes, Decision

Tree, Multi-Layer Perceptron (MLP), Transformer-негізіндегі модель, және Колмогоров–Арнольд желісі (Kolmogorov–Arnold Network, KAN). Бұл алгоритмдер көпсанатты классификация (multi-class classification) міндеттерін шешудегі тиімділігіне байланысты таңдалды. Модельдерді оқыту стратификацияланған 10 есе кросс-валидация әдісі арқылы жүзеге асырылып, барлық жиынтықтарда сыныптардың тең бөлінуі қамтамасыз етілді. Бағалау accuracy, precision, recall, F1-score және ROC–AUC сияқты көрсеткіштер бойынша жүргізілді, бұл әр модельдің болжамдық қабілетін және тұрақтылығын кешенді бағалауға мүмкіндік берді.

Барлық деректерді өңдеу, модельді әзірлеу және бағалау Python бағдарламалау тілінде жүзеге асырылды. Scikit-learn, TensorFlow, және Pandas сияқты кеңінен қолданылатын кітапханалар пайдаланылды. Модельдің параметрлерін оңтайландыру үшін grid search және random search стратегиялары қолданылып, шамадан тыс үйрену (overfitting) қаупі азайтылды. Модельдің соңғы өнімділігі оқыту жиынтығынан тыс алынған тәуелсіз тест деректерінде тексерілді, бұл оның жалпылау қабілетін (generalization) дәлелдеді.

Этикалық қағидалар зерттеу процесінің барлық кезеңінде қатаң сақталды. Барлық студенттік деректер анонимдендірілді, жеке тұлғаларды сәйкестендіруге мүмкіндік беретін ақпарат алынбады. Деректерді жинау алдында тиісті институттық рұқсаттар алынып, қатысушылардың ақпараттандырылған келісімі (informed consent) қамтамасыз етілді.

Осылайша әзірленген әдістемелік негіз жасанды интеллектке негізделген мансаптық траекторияны болжау моделінің техникалық тұрғыдан сенімді, этикалық жағынан жауапты және білім беру жүйелерінде нақты қолдануға бейім екенін қамтамасыз етеді.

Ғылыми жаңалығы: Диссертацияның ғылыми жаңалығы – кәсіби бағдарды болжауға арнайы бейімделген көпқабатты Колмогоров–Арнольд желісі (KAN) архитектурасын әзірлеу мен оңтайландыруда. Ұсынылған модель дәстүрлі және базалық KAN үлгілеріне қарағанда жоғары дәлдікке, жылдам есептеу өнімділігіне және аз жад пайдалануына қол жеткізеді. Бұл зерттеу KAN желілерінің білім беру саласындағы кәсіби бағыттау жүйелерінде алғаш рет қолданылуын қамтамасыз етіп отыр.

Зерттеудің ғылыми, практикалық және теориялық маңыздылығы: Академиялық, мінез-құлықтық және мотивациялық деректерді біріктіретін гибридті болжамдық үлгіні (framework) құра отырып, бұл диссертация білім берудегі жасанды интеллект (AIED) және білім беру деректерін интеллектуалды талдау (EDM) сияқты кеңейіп келе жатқан пәнаралық салаларға үлес қосады. Бұл - жұмыстың теориялық маңыздылығының негізгі қыры.

Зерттеу Колмогоров–Арнольд желілерін (KAN) кәсіби бағдарды болжау процесіне қолдану арқылы, білім беру саласында жиі кездесетін құрылымдалған және шағын көлемді деректер жиынтығымен жұмыс істейтін зерттеушілер үшін әдістемелік құралдардың ауқымын кеңейтеді.

Біртұтас аналитикалық тәсіл арқылы бұл зерттеу жасанды интеллект модельдерінің студентке қатысты күрделі деректерді (мысалы, психологиялық ерекшеліктер, оқу жетістіктері, қосымша белсенділіктер) түсіну және сандық тұрғыдан сипаттау мүмкіндіктерін теориялық тұрғыдан тереңдетеді. Сонымен қатар, зерттеу түсіндірілетін жасанды интеллектті (Explainable AI, XAI) білім беру жүйелеріне енгізудің теориялық негізін нығайтады. Бұл академиялық кеңесшілерге жоғары техникалық білімді қажет етпей-ақ, модель нәтижелерін түсінікті түрде қолдануға мүмкіндік береді.

Ұсынылған әдіс машиналық оқыту технологияларының деректерге негізделген болжау мен адамға бағытталған шешім қабылдау арасындағы алшақтықты қысқартуға мүмкіндік беретінін көрсетеді. Ол жеке тұлғаға бейімделген және икемді кәсіби бағдар беру жүйелерін құруға арналған теориялық негіз ұсынады.

Зерттеудің практикалық маңыздылығы – әзірленген модельдің Beta Career Platform жүйесіне енгізілуімен дәлелденеді. Бұл цифрлық университеттік экожүйеде Сүлеймен Демирел университетінің студенттері тәжірибеден өту және мансаптық мүмкіндіктерге өтініш бере алады. Осы зерттеу аясында жасалған жасанды интеллект модулі студент деректерін автоматты түрде талдап, нақты уақытта жекелендірілген кәсіби ұсыныстар беруге мүмкіндік жасайды.

Университет ішіндегі сауалнама нәтижелеріне сәйкес, жүйені енгізу 2023-2024 және 2024-2025 оқу жылдары арасында студенттердің қанағаттану деңгейін 27%-ға арттырған. Сонымен қатар, әдістеме оқытушылар мен университет әкімшілігіне құзырет олқылықтарын анықтауға, оқу бағдарламаларын еңбек нарығының талаптарымен үйлестіруге, және мансаптық қызмет сапасын арттыруға арналған аналитикалық панельдер ұсыну арқылы көмек береді.

Қолданбалы тұрғыдан алғанда, әзірленген фреймворкті Қазақстанның басқа жоғары оқу орындарында да кеңейтіп қолдануға болады. Бұл академиялық кеңес беру жүйесін цифрландыруға және білім беру нәтижелерін ішкі және халықаралық еңбек нарығының сұраныстарымен сәйкестендіруге мүмкіндік береді.

Ғылыми гипотезалар, негізгі нәтижелер және зерттеу қорытындыларының маңыздылығы: Бұл диссертацияның басты алғышарты – академиялық және мінез-құлықтық деректерді талдау процесіне жасанды интеллектті енгізу арқылы IT мамандығы студенттеріне арналған кәсіби бағдар ұсыныстарының дәлдігі мен жекелендірілуін айтарлықтай арттыруға болады деген тұжырым. Зерттеудің аналитикалық және тәжірибелік негізін төрт негізгі гипотеза құрайды.

Бірінші гипотеза (H1) бойынша, студенттердің IT саласындағы идеалды мансап бағыты мен олардың академиялық жетістіктері (мысалы, бағдарламалау, алгоритмдер және операциялық жүйелер пәндері бойынша бағалары) арасында статистикалық тұрғыдан маңызды байланыс бар. Басқаша айтқанда, тұрақты оқу үлгерімі көбіне студенттің техникалық құзыреттілігін және мамандануына когнитивтік дайындық деңгейін көрсетеді. Сондықтан академиялық жетістік кәсіби бағдар сәйкестігінің күшті болжамды көрсеткіші болып қала береді.

Екінші гипотеза (H2) академиялық көрсеткіштермен қатар мінез-құлықтық сипаттамаларды да (мысалы, хакатондарға қатысу, топтық жобалар, көшбасшылық іс-шаралар, инновациялық байқаулар) қамтиды. Бұл гипотеза тек академиялық емес, мінез-құлықтық деректерді қоса пайдалану кәсіби ұсыныстардың болжамдық дәлдігін арттырады деп пайымдайды. Мұндай тұжырым білім беру деректерін интеллектуалды талдау (Educational Data Mining) саласындағы алдыңғы зерттеулерді қолдайды, өйткені ол когнитивтік және бейкогнитивтік факторларды біріктіріп, студенттің әлеуетін кешенді бағалаудың маңызын көрсетеді.

Үшінші гипотеза (H3) дәстүрлі, қолмен жүргізілетін кеңес беру жүйелеріне қарама-қарсы идеяны ұсынады: жасанды интеллект пен машиналық оқыту әдістерін, атап айтқанда Transformer үлгілерін, нейрондық желілерді және Gradient Boosting алгоритмін қолдану арқылы неғұрлым дәл әрі контекстке сай кәсіби бағыт ұсынуға болады. Машиналық оқыту алгоритмдері үлкен және көпөлшемді деректер жиындарындағы жасырын байланыстар мен бейсызық тәуелділіктерді анықтау арқылы кәсіби болжаудың неғұрлым интеллектуалды және жекелендірілген тәсілін қамтамасыз етеді.

Төртінші гипотеза (H4) алғашқы үшеуін біріктіре отырып, мінез-құлықтық және академиялық белгілерді бір уақытта қамтитын гибридті модель нақты оқу жағдайында жоғары болжамдық қуат пен

практикалық тиімділік көрсетеді деп болжайды. Бұл тәсіл студенттің қабілеттерін, қызығушылықтарын және белгілі бір IT мамандығына дайындығын толық сипаттау үшін құрылымдалған академиялық деректер мен құрылымдалмаған мінез-құлықтық деректерді біріктірудің маңыздылығын негіздейді.

Зерттеудің тұжырымдамалық негізін үш негізгі компонент құрайды: жүйелік интеграция, модельді оқыту және деректер жинау. Beta Career Platform платформасынан алынған деректер жиынтығына академиялық көрсеткіштер (мысалы, GPA, пән бағалары), мінез-құлықтық сипаттамалар (мысалы, клуб мүшелігі, көшбасшылық қабілеттер) және қосымша іс-әрекеттер (мысалы, сертификаттар мен хакатондар) енгізілді. Модельдің сенімділігін қамтамасыз ету мақсатында деректерді алдын ала өңдеу (нормализация, кодтау, белгілерді таңдау) аяқталғаннан кейін, бірнеше модель қабатталған 10 есе кросс-валидация әдісі арқылы оқытылып және бағаланды.

Салыстырмалы талдау нәтижесінде Gradient Boosting алгоритмі Random Forest, KNN, SVM, Naïve Bayes, Decision Tree, MLP және TabTransformer сияқты басқа сегіз модельден жоғары нәтижелер көрсетті. Алайда ұзақ мерзімді зерттеу кезеңінде Колмогоров–Арнольд желілері (KANs) бейсызық тәуелділіктер мен шағын деректер жиынтығымен жұмыс істегенде анағұрлым жоғары интерпретация мен жалпылау қабілетін көрсетті.

Зерттеу шеңберінің соңғы кезеңінде таңдалған модель Beta Career Platform жүйесіне интеграцияланып, зерттеу прототипінен нақты практикалық кеңес беру құралына айналды. Бұл интеграция студент деректерін нақты уақыт режимінде талдауға, жеке кәсіби ұсыныстар жасауға және модельдің тиімділігін бағалау үшін пайдаланушы пікірлерін жинауға мүмкіндік берді.

Осылайша, ұсынылған фреймворк болжам – қолдану – бағалау циклін толықтай қамтитын өзін-өзі жетілдіретін, бейімделгіш кәсіби кеңес беру жүйесін қалыптастырды. Бұл өз кезегінде Қазақстанның цифрлық экономикасы еңбек нарығының талаптарына сәйкес білім беру нәтижелерін үйлестіруге ықпал етеді.

Публикациялар және автордың үлесі: Бұл диссертацияның нәтижелері білім беру аналитикасы, мінез-құлықтық болжам және қолданбалы деректерді модельдеу салаларындағы машиналық оқыту әдістері бойынша

автордың зерттеуінің бірізді дамуын көрсететін ғылыми мақалалар сериясында көрініс тапқан. Әрбір жарияланған жұмыс диссертацияның әдіснамалық және тәжірибелік негізін қалыптастыруға өз үлесін қосты.

«Development of Method to Predict Career Choice of IT Students in Kazakhstan by Applying Machine Learning Methods» (Journal of Robotics and Control, 2025, Scopus Q2) мақаласында автордың IT мамандығы студенттерінің кәсіби бағдарын болжауға арналған кешенді зерттеуі ұсынылған. Диссертациялық жұмыс авторы 692 студенттің деректерін жинап, құрылымдап және анонимдендірді, сондай-ақ машиналық оқыту алгоритмдерін (Random Forest, KNN, SVM, Gradient Boosting, Decision Tree) оқыту мен салыстырудың толық әдістемесін жүзеге асырды. Ол статистикалық талдау жүргізіп, нәтижелерді визуализациялап, академиялық және мінез-құлықтық факторлар арасындағы байланыстарды түсіндірді. Бұл жұмыс диссертацияның 4-тарауында («Эксперименттік нәтижелер және талқылау») қамтылған және салыстырмалы метрикалар мен аналитикалық қорытындыларды қамтиды.

«Development of a Hybrid Machine Learning Model for Classification of Soil Types Based on Geophysical Parameters» (International Journal of Innovative Research and Scientific Studies, 2025, Scopus Q2) мақаласы гибридті модельді әзірлеуге арналған. Диссертациялық жұмыс авторы деректерді тазалау, алдын ала өңдеу және параметрлерді оңтайландыру жұмыстарын орындап, модельді оқыту үшін деректердің дұрыстығын қамтамасыз етті. Бұл зерттеу барысында алынған тәжірибе диссертацияның 3-тарауында («Зерттеу әдістері және деректерді алдын ала өңдеу») сипатталған жеке деректер өңдеу тізбегін (data pipeline) жасауға қолданылды.

«Intelligent Career Path Recommendations: Leveraging Blockchain and Machine Learning» (IEEE SIST, 2025) еңбегінде диссертациялық жұмыс авторы мансаптық ұсыныстардың айқындығы мен деректер сенімділігін арттыру үшін машиналық оқыту модулін блокчейн жүйесімен біріктіру архитектурасын әзірледі. Ол модельді тестілеу және таратылған платформаның прототипін жасауға қатысты. Бұл зерттеудің нәтижелері мен техникалық шешімдері диссертацияның 5-тарауында («Beta Career платформасына модельді интеграциялау және практикалық іске асыру») көрініс тапқан.

«Identification of Key Features for Career Prediction through Recursive Feature Elimination» (Абай атындағы ҚазҰПУ Хабаршысы,

2025) мақаласында диссертациялық жұмыс авторы Recursive Feature Elimination (RFE) әдісін қолданып, ең ақпаратты предикторларды анықтады, академиялық және мінез-құлықтық сипаттамалар арасындағы корреляциялық талдау жүргізді, аналитикалық кестелер мен графикалық материалдар дайындады. Бұл нәтижелер диссертацияның 3-тарауында («Зерттеу әдістері және деректерді өңдеу») қамтылған, мұнда белгілерді таңдау және модель параметрлерін оңтайландыру процесі егжей-тегжейлі сипатталған.

«Detecting Anxiety and Depression from Social Media Text by Applying Machine Learning Methods» (Вестник АГА, 2025) мақаласында Диссертациялық жұмыс авторы мәтіндік деректерді алдын ала өңдеу токенизация, тазалау және белгілер кеңістігін қалыптастыру жұмыстарын атқарды және машиналық оқыту модельдерін оқытуға қатысты. Неструктурланған мәтіндік деректерді талдауға арналған әзірленген әдістер диссертацияның 3-тарауында мінез-құлықтық деректерді өңдеу кезеңінде қолданылды.

Соңында, **«Comprehensive Evaluation of Real-Time Object Detection Algorithm Based on Extended Criteria»** (КазАТК Хабаршысы, 2024) еңбегі компьютерлік көру алгоритмдерін дәлдік пен жылдамдық көрсеткіштері бойынша кеңейтілген критерийлер арқылы бағалауға арналды. Диссертациялық жұмыс авторы осы алгоритмдерді жобалау мен сынауға, сондай-ақ есептеу тиімділігін (оқыту уақыты, жадыны пайдалану) талдауға үлес қосты. Алынған тәжірибе диссертацияның 5-тарауында («Модельдердің өнімділігін салыстырмалы талдау») қолданылып, AI-модельдерді бағалау бойынша зерттеудің әдіснамалық негізін толықтырды.

Диссертацияның ұзақтығы мен құрылымы: Диссертациялық жұмыс логикалық тұрғыдан өзара байланысты алты негізгі бөлімнен тұрады, олардың соңынан қорытынды, пайдаланылған әдебиеттер тізімі және қосымшалар берілген. Мұндай құрылым зерттеудің теориялық негіздерін, әдістемелік үлгісін, эмпирикалық нәтижелерін және практикалық ұсыныстарын дәйекті түрде ұсынуға мүмкіндік береді.

Кіріспе бөлімінде зерттеу тақырыбының өзектілігі Қазақстанның жоғары білім беру жүйесінің контексінде ашылып, зерттеу мақсаты мен міндеттері, гипотезалары және ғылыми жаңалығы айқындалады. Бұл бөлім диссертацияның теориялық және практикалық маңызын айқындап, зерттеу логикасын қалыптастырады.

Әдебиеттерге шолу бөлімінде жасанды интеллект пен машиналық оқыту технологияларын қолдану арқылы мансаптық болжау жүйелерін дамыту бойынша алдыңғы зерттеулерге талдау жасалып, қазіргі ғылыми олқылықтар анықталады. Бұл талдау зерттеу нысанының ерекшелігін нақтылап, жұмыстың ғылыми жаңалығын негіздейді.

Әдістеме бөлімі зерттеу деректерін сипаттайды. Деректер жиынтығы Beta Career Platform жүйесінен алынған және 692 студенттің анонимдендірілген жазбаларын қамтиды. Мұнда деректерді алдын ала өңдеу, ерекшеліктерді (features) іріктеу және модельдеу стратегиялары егжей-тегжейлі түсіндіріледі. Сонымен қатар, сегіз машиналық оқыту алгоритмінің — Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, MLP және TabTransformer — салыстырмалы сынақ нәтижелері баяндалады. Осы тарауда модельдердің тиімділігін арттыру және интерпретациясын жақсарту мақсатында Колмогоров–Арнольд желісінің (Kolmogorov–Arnold Network, KAN) қолданылуы енгізіледі.

Нәтижелер мен талқылау бөлімінде модельдерді бағалау нәтижелері беріліп, алгоритмдердің өнімділігі F1-score, accuracy, precision және recall көрсеткіштері бойынша салыстырылады. Сондай-ақ модельдің Beta Career Platform жүйесіне интеграциялануы сипатталып, жасанды интеллектке негізделген кеңес беру жүйесінің студенттердің қанағаттанушылығына әсері талданады.

Ғылыми жаңалық және теориялық маңыздылық бөлімінде зерттеу жұмысының инновациялық нәтижелері жинақталып, KAN желісінің білім беру болжам жүйелерінде алғаш қолданылуы мен академиялық шешім қабылдау үдерісіне арналған жасанды интеллектке негізделген модельдің үлесі айқындалады.

Қорытынды мен ұсыныстар бөлімінде негізгі нәтижелер түйінделіп, зерттеу гипотезаларының расталғаны дәлелденеді және басқа жоғары оқу орындарына AI-негізделген мансаптық бағдар беру жүйелерін енгізуге арналған нақты ұсыныстар беріледі. Бұл ұсыныстар түлектердің еңбек нарығындағы бейімділігін арттыруға және академиялық қолдау қызметтерін жетілдіруге бағытталған.

Диссертацияның жалпы көлемі шамамен 123 бет, оның құрамында 41 сурет, 6 кесте және 228 библиографиялық дерек бар. Мұндай құрылым аналитикалық дәлдік пен нәтижелердің көрнекі бейнеленуін қамтамасыз етіп, зерттеу мақсаттары мен қорытындыларының дәйекті әрі түсінікті баяндалуына мүмкіндік береді.

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy (PhD) in EP "8D06102 – Computer Sciences"

BERLIKOZHA BAUYRZHAN ASSETULY

AI-Powered Approach to Career Path Prediction for IT Students Using Academic and Behavior Data

Relevance of the work. The global labor market has undergone a fundamental upheaval due to the rapid rate of digital transformation, necessitating the need for a new generation of workers with flexible technical and analytical abilities. To boost productivity and spur innovation, businesses all over the world are depending more and more on automation, data analytics, and artificial intelligence (AI) technologies. As a result, there is increasing pressure on the educational sector to adapt its methods of instruction and counseling to the changing demands of the technologically advanced economy. Around 85 million jobs could be lost to automation, while 97 million new positions that are better suited to a digital economy will be created, according to the World Economic Forum's Future of Jobs Report (2023). A data-driven, individualized strategy to prepare pupils for occupations that have not yet completely materialized is required due to this paradigm change.

AI-powered career counseling programs have become more well-known in this context as a means of bridging the knowledge gap between university and employment. These systems use machine learning and predictive analytics to find relationships between a student's personality, ideal job paths, and academic achievement. According to recent studies, artificial intelligence (AI) can offer more individualized and accurate career recommendations than conventional advising techniques, increasing job satisfaction and employability rates. Furthermore, the increasing interest in Learning Analytics (LA) and Educational Data Mining (EDM) shows that the academic community around the world recognizes AI's potential to improve educational outcomes. AI-based solutions for curriculum adaptation and student performance prediction have already been incorporated by nations like South Korea, Singapore, and Finland, leading to more effective talent development pipelines.

However, there is still disparity in how these technologies are actually implemented in various geographical areas. Developing countries encounter

systemic limitations like fragmented data infrastructure, a lack of AI expertise, and limited integration of data-driven solutions into education management systems, whereas high-income countries enjoy the advantages of robust data ecosystems and strong institutional support for innovation. As a result, there is a growing "AI adoption gap" in education that could exacerbate disparities in national competitiveness and graduate employability. By creating an AI-powered framework for predicting career paths that is especially suited to the setting of emerging economies, this research directly helps to close that gap.

Aim of the Research Work: This thesis aims to study and develop an AI-powered career prediction system based on academic and behavioral data to predict the future career paths of IT students.

Objectives of the Research Work: Five major objectives are established for the research done for this dissertation:

1. Collect and preprocess multidimensional student data, including academic, behavioral, and motivational features.
2. A Study and Analysis of AI-Powered Approaches for Career Path Prediction.
3. Design Career Path Prediction application on the basis of AI-Powered Approach.
4. Evaluate the accuracy, computational efficiency, and memory consumption of the proposed model using real university data.
5. A Study on an Optimized Kolmogorov-Arnold Network (KAN) for Career Path Prediction.

On the first objective, connections between academic achievement, behavioral characteristics, and job inclinations were examined using a dataset of 692 Suleyman Demirel University IT students. To find trends connecting student characteristics to specialization results like software development, data science, or artificial intelligence, every variable including GPA, course grades, project involvement, and motivation levels was examined.

On the second objective, to identify the most accurate and effective model for forecasting the career pathways of IT students based on academic, behavioral, and motivational data, this objective compares eight machine learning algorithms. K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Naïve Bayes (NB), Decision Tree (DT), Multi-Layer Perceptron (MLP), TabTransformer, and Random Forest (RF) are among the algorithms that were put to the test. 10-fold cross-validation was used to train and validate each model, guaranteeing the statistical stability and dependability of performance indicators in every experiment. Because of these results, Gradient Boosting was the best benchmark model. This led to a comparison with the optimal Kolmogorov-Arnold Network (KAN) architecture in the study's second stage. This progression from simple machine learning

models to complex neural architectures establishes a crucial methodological bridge by demonstrating how incremental innovation and model optimization can lead to significant improvements in accuracy, interpretability, and computational efficiency for career path prediction in educational data mining.

The third objective centers on developing an AI-powered career path prediction application that implements the ML model in a practical and scalable system. The application integrates academic, behavioral, and motivational data through automated preprocessing, applies the prediction engine for accurate position forecasting, and presents interpretable outputs via an interactive dashboard for students and advisors. Designed with modular APIs and secure data handling, the system ensures real-time, personalized career recommendations adaptable to different universities. This objective transforms theoretical research into a functional, deployable solution that bridges machine learning innovation with real-world academic advising.

Using actual university data, the fourth objective seeks to assess the effectiveness and performance of the suggested AI-powered model with a particular emphasis on three crucial areas: memory usage, computational efficiency, and accuracy. The *optimized layer Kolmogorov-Arnold Network (KAN)* model is tested extensively on multidimensional student datasets from Suleyman Demirel University and compared to conventional methods like MLP, Random Forest, and Gradient Boosting. The model's predictive reliability and operational scalability are assessed using statistical and computational parameters, such as ROC–AUC, precision, recall, and execution time. This thorough examination confirms the model's supremacy in providing high accuracy with little resource requirements, guaranteeing its applicability for actual use in educational settings.

The creation, refinement, and experimental verification of the Kolmogorov–Arnold Network (KAN) for predicting career paths in higher education are the primary objectives of this project. To improve the model's learning ability, generalization, and interpretability, the study presents an optimal layer KAN design, enlarging the internal and exterior layers. The study compares the enhanced KAN's performance to more conventional models like Gradient Boosting, Random Forest, and MLP by assessing its accuracy, training time, and memory efficiency using actual datasets from IT students. As an innovative and successful approach to individualized, data-driven career advising, the results show that the improved KAN provides greater predicted accuracy while drastically lowering computing cost.

Research Methods: This study used anonymized data from 692 IT students at SDU University, including academic performance, personality traits, extracurricular activities, and certifications. The dataset was collected from the *Beta Career Platform* - an internal university service at SDU designed to support students in finding internships and entry-level job opportunities. This

platform provided reliable, representative, and practically significant data on 692 students, which served as a solid foundation for building predictive models. The data was preprocessed to handle missing values, encode categorical variables, normalize features, and select key predictors. Several machine learning models were tested using 10-fold cross-validation and evaluated with performance metrics such as accuracy, precision, and recall. An independent test set was used for final evaluation. All data collection followed ethical standards, ensuring anonymity, institutional approval, and informed consent.

Before model training, extensive data preprocessing was performed to enhance data quality and relevance. Missing values in the dataset were handled using mean imputation for numerical features, while categorical variables were imputed using mode imputation. Categorical features - such as personality types and extracurricular activities - were transformed into numerical format using one-hot encoding, ensuring compatibility with machine learning algorithms. Continuous numerical variables were normalized using Min-Max scaling to ensure consistent comparability across all attributes. Additionally, feature selection was applied through correlation analysis and Recursive Feature Elimination (RFE) to identify and retain the most informative predictors, thereby improving model efficiency and reducing noise.

To construct the predictive model, several machine learning algorithms were selected: Random Forest, K-Nearest Neighbors, Support Vector Machine, Gradient Boosting, Naive Bayes, Decision Tree, Multi-Layer Perceptron (MLP), Transformer-based Model and Kolmogorov-Arnold Networks. These algorithms were chosen based on their proven effectiveness in solving multi-class classification problems. Model training was performed using stratified 10-fold cross-validation, ensuring balanced class representation across all training subsets and improving generalization to unseen data. Model evaluation was carried out using multiple performance metrics - accuracy, precision, recall, F1-score, and ROC-AUC - providing a comprehensive understanding of each model's predictive performance and robustness.

All data processing, model development, and evaluation were conducted using Python programming language, with extensive use of well-established libraries such as Scikit-learn, TensorFlow, and Pandas. Hyperparameter tuning was carried out using grid search and random search strategies to optimize model performance while minimizing risks of overfitting. Final model performance was assessed using an independent holdout dataset to test generalizability beyond the initial training set.

Ethical considerations were strictly adhered to throughout the research process. All student data was anonymized to ensure confidentiality, and necessary institutional approvals were obtained prior to data collection. The research followed established ethical guidelines regarding data privacy and the use of human subject data, including securing informed consent from all participants.

This comprehensive methodological framework ensures that the resulting

AI-powered career path prediction model is not only technically sound but also ethically responsible and applicable for real-world use in academic advising systems.

Scientific novelty: The scientific novelty of the dissertation lies in the development and optimization of a layer Kolmogorov-Arnold Network (KAN) architecture specifically adapted for career path prediction, which achieves higher accuracy, faster computation, and lower memory consumption compared to traditional and baseline KAN models, marking the first application of KAN in the field of educational career recommendation systems.

Scientific, Practical, and Theoretical Significance of the Study: By creating a hybrid predictive framework that combines academic, behavioral, and motivational data, this dissertation adds to the expanding interdisciplinary fields of Artificial Intelligence in Education (AIED) and Educational Data Mining (EDM). This is where its theoretical significance lies. The study broadens the range of methodological tools available to researchers working with structured and small-sample datasets, which are common in educational environments, by applying Kolmogorov–Arnold Networks (KANs) to career path prediction.

Through a unified analytical approach, the research increases the theoretical knowledge of how AI models may comprehend and quantify complex student-related data, including psychological qualities, performance indicators, and extracurricular engagement. Additionally, it strengthens the theoretical foundation for implementing explainable AI (XAI) in educational systems by providing interpretable model outputs that academic advisers can use without the need for highly skilled technical knowledge.

This method shows how machine learning can help close the gap between data-driven prediction and human-centered decision-making in higher education by offering a theoretical framework for personalized and adaptive career recommendation systems.

The proposed model's practical incorporation into the Beta Career Platform, a university-level digital ecosystem where Suleyman Demirel University students apply for internships and early-career possibilities, demonstrates the dissertation's practical value. The AI module created for this study makes it possible to automatically analyze student data and provides real-time, personalized career recommendations.

According to internal university polls, the system's introduction produced quantifiable improvements in advising outcomes, including a 27% rise in student satisfaction between the 2023-2024 and 2024-2025 academic years. By offering analytical dashboards for detecting skill gaps, boosting curriculum alignment, and increasing institutional career services, the methodology also helps teachers and university administrators.

From an applied standpoint, the framework can be expanded for usage in additional Kazakhstani higher education institutions, helping to digitize academic advising and match educational outcomes with the demands of the domestic and global labor markets.

Scientific Hypotheses, Main Findings, and Significance of Results: The premise of this dissertation is that the accuracy and personalization of career path recommendations for IT students can be significantly improved by incorporating artificial intelligence into academic and behavioral data analysis. Four main theories serve as the study's compass and together they establish its analytical and experimental framework.

According to the first hypothesis (H1), there are statistically significant correlations between students' ideal career paths in IT sectors and their academic achievement, including their grades in programming, algorithms, and operating systems. In other words, since consistent academic performance frequently reflects both technical competence and cognitive readiness for specialization, academic achievement continues to be a strong predictor of career alignment.

The second hypothesis (H2) incorporates behavioral traits like involvement in hackathons, group projects, leadership exercises, and innovation competitions in addition to academic markers. It makes the case that using behavioral data instead of only academic indicators improves the prediction accuracy of job suggestions. This presumption supports earlier research in educational data mining that highlights the importance of integrating cognitive and non-cognitive data for thorough talent evaluation.

In contrast to conventional, manual advising systems, the third hypothesis (H3) contends that the application of artificial intelligence and machine learning techniques, particularly Transformer-based models, Neural Networks, and Gradient Boosting, can produce more precise and contextually relevant career guidance. A more sophisticated and customized method of career prediction is made possible by machine learning algorithms, which may uncover hidden correlations and nonlinear interactions between factors in big, multidimensional datasets.

By suggesting that a hybrid model that concurrently incorporates behavioral and academic characteristics will exhibit higher predictive power and practical application in a real-world academic situation, the fourth hypothesis (H4) synthesizes the first three. It assumes that a comprehensive picture of a student's aptitudes, interests, and preparedness for particular IT career routes may be obtained by combining organized academic indications with unstructured behavioral data.

Three main elements form the study's conceptual framework: system integration, model training, and data collection. Academic markers (e.g., GPA, course grades), behavioral traits (e.g., club membership, leadership skills), and

extracurricular information (e.g., certifications and hackathons) are all included in the dataset, which was sourced from the Beta Career Platform. To guarantee robustness and dependability, several models were trained and evaluated using stratified 10-fold cross-validation following data preprocessing, which included normalization, encoding, and feature selection.

Gradient Boosting outperformed the other eight algorithms, according to a comparative analysis of Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, MLP, and Tab Transformer. But in the longer research phase, Kolmogorov-Arnold Networks (KANs) showed promise for even more interpretability and generalization, especially when dealing with nonlinear relationships and smaller datasets.

In the framework's last stage, the selected model is integrated into the Beta Career Platform, turning it from a research prototype into a useful advising tool. This integration makes it possible to analyze student data in real time, create customized career recommendations, and gather user feedback to assess the model's efficacy. The framework creates a self-improving, adaptive system for career advising by completing the loop between prediction, application, and evaluation. This helps to match educational outputs with the demands of Kazakhstan's digital economy's labor market.

Publications and Author's Contribution: The results of this dissertation were reflected in a series of scientific papers that demonstrate the progressive development of the author's research on machine learning methods for educational analytics, behavioral prediction, and applied data modeling. Each publication contributed to the formation of the methodological framework and experimental foundation of the dissertation.

The paper **"Development of Method to Predict Career Choice of IT Students in Kazakhstan by Applying Machine Learning Methods"** (Journal of Robotics and Control, 2025, Scopus Q2) presented the author's comprehensive research on predicting career trajectories of IT students. Author of the dissertation collected, structured, and anonymized data from 692 IT students and implemented the full methodology for training and comparing machine learning algorithms (Random Forest, KNN, SVM, Gradient Boosting, Decision Tree). He performed statistical analysis, visualization of results, and interpretation of relationships between academic and behavioral factors. This work is reflected in Chapter 4 ("Experimental Results and Discussion") of the dissertation, which includes comparative metrics and analytical conclusions.

The article **"Development of a Hybrid Machine Learning Model for Classification of Soil Types Based on Geophysical Parameters"** (International Journal of Innovative Research and Scientific Studies, 2025, Scopus Q2) focused on hybrid model design. Author of the dissertation conducted data cleaning, preprocessing, and optimization of parameters,

ensuring data integrity for model training. The experience gained from this research was applied to develop the custom data pipeline described in Chapter 3 ("Research Methods and Data Preprocessing").

In **"Intelligent Career Path Recommendations: Leveraging Blockchain and Machine Learning"** (IEEE SIST, 2025), Author of the dissertation designed the architecture for integrating a machine learning module with a blockchain system to enhance transparency and data reliability in career recommendations. He participated in model testing and prototype development of a distributed platform. The outcomes and technical solutions from this study are reflected in Chapter 5 ("Integration and Practical Implementation of the Model in the Beta Career Platform").

The paper **"Identification of Key Features for Career Prediction through Recursive Feature Elimination"** (Bulletin of Abai Kazakh National Pedagogical University, 2025) presented a feature selection study. Author of the dissertation applied the Recursive Feature Elimination (RFE) method to identify the most informative predictors, performed correlation analysis of academic and behavioral features, and prepared analytical tables and graphical visualizations. The findings are included in Chapter 3 ("Research Methods and Data Processing"), which details the feature selection and model parameter optimization process.

In **"Detecting Anxiety and Depression from Social Media Text by Applying Machine Learning Methods"** (Bulletin of AGA, 2025), author of the dissertation was responsible for text data preprocessing, including tokenization, cleaning, and feature space formation, and participated in training machine learning classifiers. The developed techniques for analyzing unstructured textual data were integrated into Chapter 3 of the dissertation during the design of the behavioral feature processing pipeline.

Finally, the work **"Comprehensive Evaluation of Real-Time Object Detection Algorithm Based on Extended Criteria"** (Bulletin of KazATK, 2024) evaluated computer vision algorithms by extending accuracy and speed assessment criteria. Author of the dissertation contributed to designing and testing these algorithms, as well as analyzing computational efficiency (training time, memory usage). The experience was adapted in Chapter 5 ("Comparative Analysis of Model Performance"), enriching the dissertation's methodological foundation for evaluating AI models.

Volume and Structure of the Dissertation: The dissertation is organized logically into six major chapters, which are followed by appendices, references, and conclusions. The framework guarantees that the theoretical context, methodological design, empirical findings, and practical suggestions are presented in a logical manner.

The introduction sets the research environment within Kazakhstan's higher education system, identifies the research's relevance, formulates the

research topic, and discusses the study's objectives, hypotheses, and innovation.

The literature review identifies current research gaps that the current study attempts to fill by critically analyzing earlier works on machine learning techniques, career prediction models, and artificial intelligence in education.

Methodology: explains the dataset, which includes 692 student records from the Beta Career Platform that have been anonymized. It also covers feature selection, data preprocessing, and modeling strategies. The comparative testing of eight algorithms-Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, MLP, and TabTransformer-is also explained in this chapter. Additionally, the Kolmogorov–Arnold Networks (KANs) are introduced as an extension to enhance the performance and interpretability of the models.

Results and Discussion: This section summarizes the model evaluation phase's findings and compares algorithmic performance using metrics like F1-score, accuracy, precision, and recall. The model's incorporation into the Beta Career Platform is also covered, and the effect of AI-based advising on student happiness is examined.

The research's main innovative features are summed up in Scientific Novelty and Theoretical Implications, which highlights the application of KANs in educational prediction and the contribution to AI-based frameworks for academic decision support.

In order to improve employability and academic support services, the conclusion and recommendations summarize the key findings, validate the theories, and offer doable suggestions for other universities looking to implement AI-powered advising systems.

The dissertation's roughly 123 pages of text are backed up by 41 figures, 6 tables, and 228 bibliographic references, which demonstrate a balance between analytical rigor and results visualization. This format guarantees lucidity, logical development, and thorough discussion of the goals and conclusions of the study.

АННОТАЦИЯ

диссертации на соискание ученой степени доктора философии (PhD) по ОП “8D06102 - Компьютерные науки”

БЕРЛІҚОЖА БАУЫРЖАН ӘСЕТҰЛЫ

Прогнозирование карьерной траектории студентов IT-направлений с использованием академических и поведенческих данных на основе методов искусственного интеллекта

Актуальность работы: Глобальная цифровизация и внедрение технологий искусственного интеллекта формируют новые требования к подготовке специалистов. Это требует от системы образования перехода к персонализированным и аналитически обоснованным методам карьерного консультирования.

Цель исследовательской работы: Настоящая диссертационная работа направлена на изучение и разработку системы прогнозирования карьерных траекторий студентов IT-специальностей на основе академических и поведенческих данных с применением технологий искусственного интеллекта.

Задачи исследовательской работы: В рамках диссертационного исследования были поставлены пять основных задач:

1. Сбор и предварительная обработка многомерных данных студентов, включающих академические, поведенческие и мотивационные характеристики.
2. Изучение и анализ подходов, основанных на технологиях искусственного интеллекта, для прогнозирования карьерных траекторий.
3. Разработка приложения для прогнозирования карьерных путей на основе AI-подхода.
4. Оценка точности, вычислительной эффективности и объема памяти предлагаемой модели с использованием реальных данных университета.
5. Исследование оптимизированной сети Колмогорова–Арнольда (Kolmogorov–Arnold Network, KAN) для прогнозирования карьерных

траекторий.

По первой задаче были исследованы взаимосвязи между академической успеваемостью, поведенческими характеристиками и профессиональными предпочтениями студентов на основе данных 692 студентов IT-направления Сулейман Демирель Университета. Для выявления закономерностей, связывающих особенности студентов с результатами специализации (например, в области разработки программного обеспечения, науки о данных или искусственного интеллекта), были проанализированы такие показатели, как средний балл (GPA), оценки по дисциплинам, участие в проектах и уровень мотивации. Во второй задаче, с целью определения наиболее точной и эффективной модели прогнозирования карьерных траекторий IT-студентов на основе академических, поведенческих и мотивационных данных, было проведено сравнение восьми алгоритмов машинного обучения: K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Naïve Bayes (NB), Decision Tree (DT), Multi-Layer Perceptron (MLP) и TabTransformer. Каждая модель была обучена и протестирована методом 10-кратной кросс-валидации, что обеспечило статистическую устойчивость и достоверность полученных результатов. По итогам экспериментов модель Gradient Boosting продемонстрировала наилучшие показатели и была выбрана в качестве эталонной. На втором этапе исследования она была сопоставлена с оптимизированной архитектурой Kolmogorov–Arnold Network (KAN). Такой переход от базовых алгоритмов машинного обучения к более сложным нейросетевым архитектурам сформировал важный методологический мост, показав, как поэтапная оптимизация модели способствует повышению точности, интерпретируемости и вычислительной эффективности при прогнозировании карьерных траекторий в образовательной аналитике. Третья задача была направлена на разработку практического приложения для прогнозирования карьерных траекторий, реализующего модель машинного обучения в масштабируемой системе. Приложение объединяет академические, поведенческие и мотивационные данные посредством автоматизированной предварительной обработки, применяет механизм предсказания для точного прогноза карьерных направлений и отображает интерпретируемые результаты на интерактивной панели для студентов и консультантов. Система, разработанная с использованием модульных API и безопасных механизмов обработки данных, обеспечивает персонализированные карьерные рекомендации в режиме реального времени и может быть адаптирована под различные университетские контексты. Таким образом, теоретические результаты исследования преобразованы в прикладное решение, соединяющее инновации машинного обучения с практикой академического консультирования. В рамках четвертой задачи была проведена оценка

эффективности и производительности предложенной AI-модели с акцентом на три ключевых параметра: использование памяти, вычислительная эффективность и точность. Оптимизированная слоистая архитектура Kolmogorov–Arnold Network (KAN) была протестирована на многомерных студенческих данных Сулейман Демирель Университета и сопоставлена с традиционными моделями — MLP, Random Forest и Gradient Boosting. Надёжность прогнозирования и масштабируемость работы модели были оценены с использованием статистических и вычислительных метрик: ROC–AUC, precision, recall и время выполнения. Проведённый анализ подтвердил преимущество модели KAN, обеспечивающей высокую точность при минимальных затратах вычислительных ресурсов, что делает её применимой в реальных образовательных условиях.

Методы исследования: В данном исследовании использовались анонимизированные данные 692 студентов IT-направления Сулейман Демирель Университета (SDU), включающие показатели академической успеваемости, личностные характеристики, участие во внеучебной деятельности и наличие профессиональных сертификатов. Данные были собраны с платформы Beta Career Platform — внутреннего университетского сервиса, предназначенного для поддержки студентов в поиске стажировок и вакансий начального уровня. Эта платформа предоставила надёжный, репрезентативный и практически значимый набор данных по 692 студентам, который послужил прочной эмпирической основой для построения прогнозных моделей. Перед обучением модели была проведена комплексная предварительная обработка данных, направленная на повышение их качества и аналитической значимости. Отсутствующие значения в числовых переменных заполнялись с использованием метода среднего значения (mean imputation), а в категориальных признаках применялась мода (mode imputation). Категориальные признаки - такие как тип личности или участие во внеучебных мероприятиях - были преобразованы в числовой формат методом one-hot encoding, что обеспечило их совместимость с алгоритмами машинного обучения. Непрерывные числовые переменные нормализовались при помощи Min–Max масштабирования, что позволило сохранить сопоставимость атрибутов в едином диапазоне. Дополнительно был проведен отбор признаков (feature selection) на основе корреляционного анализа и метода рекурсивного исключения признаков (Recursive Feature Elimination, Sharp) для выявления наиболее информативных переменных, что повысило эффективность модели и уменьшило влияние статистического шума. Для построения прогнозной модели были выбраны несколько алгоритмов машинного обучения: Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM),

Gradient Boosting, Naïve Bayes, Decision Tree, Multi-Layer Perceptron (MLP), Transformer-based Model и Kolmogorov–Arnold Networks (KAN). Эти алгоритмы были выбраны ввиду их высокой эффективности в задачах много классовой классификации. Обучение моделей проводилось с использованием стратифицированной 10-кратной кросс-валидации, что обеспечивало сбалансированное распределение классов во всех обучающих подмножествах и повышало способность моделей к обобщению на новых данных. Оценка качества моделей выполнялась с использованием нескольких метрик: accuracy, precision, recall, F1-score и ROC–AUC, что позволило комплексно оценить точность предсказаний, устойчивость и надежность каждой модели. Все этапы обработки данных, построения и оценки моделей выполнялись на языке программирования Python с использованием широко признанных библиотек Scikit-learn, TensorFlow и Pandas. Для повышения эффективности моделей и минимизации риска переобучения применялись методы Grid Search и Random Search для оптимизации гиперпараметров. Окончательная оценка производительности проводилась на независимом тестовом наборе данных (holdout dataset), что позволило проверить способность моделей к обобщению за пределами обучающей выборки. Во всех этапах исследования строго соблюдались этические принципы. Все данные студентов были анонимизированные для обеспечения конфиденциальности, а сбор данных проводился с разрешения соответствующих университетских структур. Исследование осуществлялось в соответствии с международными нормами этики в области защиты персональных данных и использования данных человека в научных целях, включая получение информированного согласия от всех участников. Таким образом, представленная методологическая структура обеспечивает, чтобы разработанная система прогнозирования карьерных траекторий на основе искусственного интеллекта была не только технически обоснованной и надежной, но и этически ответственной, что делает её применимой в реальных условиях академического консультирования.

Научная новизна: Научная новизна диссертационного исследования заключается в разработке и оптимизации слоистой архитектуры сети Колмогорова – Арнольда (Kolmogorov – Arnold Network, KAN), специально адаптированной для прогнозирования карьерных траекторий студентов. Предложенная архитектура обеспечивает более высокую точность прогнозирования, повышенную вычислительную скорость и сниженное потребление памяти по сравнению с традиционными и базовыми моделями KAN. Данное исследование представляет собой первое применение сети Колмогорова – Арнольда в области образовательных систем карьерных

рекомендаций, что открывает новые перспективы для использования методов искусственного интеллекта в академическом консультировании.

Научная, практическая и теоретическая значимость исследования: Создавая гибридную предсказательную модель, объединяющую академические, поведенческие и мотивационные данные, данная диссертация вносит вклад в развивающиеся междисциплинарные направления — *искусственный интеллект в образовании* (AIED) и *интеллектуальный анализ образовательных данных* (EDM). Именно в этом заключается её теоретическая значимость.

Исследование расширяет спектр методологических инструментов для учёных, работающих со структурированными и маловыборочными наборами данных, типичными для образовательной среды, благодаря применению *сетей Колмогорова–Арнольда (KANs)* к задаче прогнозирования карьерных траекторий.

Посредством единого аналитического подхода работа углубляет теоретические представления о том, как модели искусственного интеллекта могут понимать и количественно оценивать сложные данные, связанные со студентами, включая психологические характеристики, показатели успеваемости и внеучебную активность. Кроме того, исследование укрепляет теоретическую базу для внедрения *объяснимого искусственного интеллекта (XAI)* в образовательные системы, предоставляя интерпретируемые результаты моделей, которыми могут пользоваться академические консультанты без необходимости обладать высокой технической квалификацией.

Предложенный подход демонстрирует, как методы машинного обучения могут способствовать сокращению разрыва между анализом данных и ориентированным на человека процессом принятия решений в высшем образовании, обеспечивая теоретическую основу для персонализированных и адаптивных систем карьерных рекомендаций.

Практическая ценность диссертации подтверждается внедрением предложенной модели в *Beta Career Platform* — университетскую цифровую экосистему, где студенты Университета имени Сулеймана Демиреля подают заявки на стажировки и ранние карьерные возможности. Разработанный для данного исследования модуль искусственного интеллекта обеспечивает автоматический анализ данных студентов и предоставляет персонализированные карьерные рекомендации в режиме реального времени.

Согласно внутренним опросам университета, внедрение системы привело к количественно измеримым улучшениям результатов консультирования, включая рост удовлетворённости студентов на 27% в период между 2023–2024 и 2024–2025 учебными годами. Кроме того,

разработанная методика помогает преподавателям и администраторам университета за счет предоставления аналитических панелей для выявления дефицитов компетенций, повышения соответствия учебных программ требованиям рынка и оптимизации карьерных сервисов.

С прикладной точки зрения предложенный фреймворк может быть масштабирован и адаптирован для использования в других вузах Казахстана, способствуя цифровизации академического консультирования и согласованию образовательных результатов с потребностями отечественного и международного рынка труда.

Научные гипотезы, основные результаты и значимость исследования:Основная идея данной диссертации заключается в том, что точность и персонализация карьерных рекомендаций для студентов IT-направлений могут быть значительно повышены за счёт внедрения технологий искусственного интеллекта в анализ академических и поведенческих данных. Четыре основные гипотезы послужили ориентиром исследования и в совокупности сформировали его аналитическую и экспериментальную основу.

Согласно первой гипотезе (H1), существует статистически значимая взаимосвязь между идеальными карьерными направлениями студентов в IT-сфере и их академической успеваемостью, включая оценки по дисциплинам «Программирование», «Алгоритмы» и «Операционные системы». Иными словами, устойчивые академические результаты, как правило, отражают как техническую компетентность, так и когнитивную готовность к профессиональной специализации, что делает академическую успеваемость надёжным предиктором соответствия карьеры.

Вторая гипотеза (H2) учитывает поведенческие характеристики, такие как участие в хакатонах, групповых проектах, лидерских активностях и конкурсах инноваций, наряду с академическими показателями. Она утверждает, что использование поведенческих данных наряду с академическими признаками повышает точность прогнозирования карьерных рекомендаций. Это предположение согласуется с предыдущими исследованиями в области Educational Data Mining, которые подчеркивают важность интеграции когнитивных и некогнитивных факторов для комплексной оценки потенциала учащихся.

В противоположность традиционным ручным системам консультирования, третья гипотеза (H3) утверждает, что применение искусственного интеллекта и методов машинного обучения — особенно моделей на основе Transformer, нейронных сетей и Gradient Boosting — позволяет обеспечить более точные и контекстно релевантные карьерные рекомендации. Алгоритмы машинного обучения обеспечивают более глубокий и персонализированный подход к прогнозированию карьеры,

выявляя скрытые зависимости и нелинейные взаимодействия между факторами в больших многомерных массивах данных.

Четвертая гипотеза (H4) объединяет предыдущие три, выдвигая предположение, что гибридная модель, одновременно учитывающая поведенческие и академические характеристики, обладает более высокой предсказательной способностью и практической применимостью в реальных академических условиях. Она исходит из того, что комплексное представление о способностях, интересах и готовности студента к определенным карьерным направлениям в IT можно получить, сочетая структурированные академические показатели с неструктурированными поведенческими данными.

Концептуальная структура исследования включает три основных компонента: интеграцию системы, обучение модели и сбор данных. В набор данных, полученный с платформы Beta Career, вошли академические показатели (например, GPA, оценки по курсам), поведенческие характеристики (например, участие в клубах, лидерские качества) и внеучебная активность (например, сертификаты и участие в хакатонах). Для обеспечения надежности и устойчивости модели использовались методы стратифицированной 10-кратной кросс-валидации после этапов нормализации, кодирования и отбора признаков.

Согласно сравнительному анализу алгоритмов — Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, MLP и TabTransformer — наилучшие результаты показал метод Gradient Boosting. Однако на более позднем этапе исследований сети Колмогорова–Арнольда (KANs) продемонстрировали еще более высокую интерпретируемость и способность к обобщению, особенно при работе с нелинейными зависимостями и ограниченными наборами данных.

На заключительном этапе предложенная модель была интегрирована в платформу Beta Career, что позволило превратить её из исследовательского прототипа в полноценный инструмент академического консультирования. Такая интеграция обеспечивает анализ данных студентов в режиме реального времени, формирование персонализированных карьерных рекомендаций и сбор обратной связи для оценки эффективности модели.

Таким образом, разработанный фреймворк создаёт самообучающуюся адаптивную систему карьерного консультирования, замыкая цикл между прогнозированием, применением и оценкой. Это способствует согласованию образовательных результатов с требованиями рынка труда цифровой экономики Казахстана.

Публикации и вклад автора: Результаты данной диссертации отражены в серии научных статей, демонстрирующих поступательное развитие исследований автора в области методов машинного обучения для

образовательной аналитики, прогнозирования поведенческих характеристик и прикладного моделирования данных. Каждая публикация внесла вклад в формирование методологической и экспериментальной основы диссертационного исследования.

Статья **«Development of Method to Predict Career Choice of IT Students in Kazakhstan by Applying Machine Learning Methods»** (Journal of Robotics and Control, 2025, Scopus Q2) представляет комплексное исследование автора по прогнозированию карьерных траекторий студентов IT-направлений. Автор диссертации осуществил сбор, структурирование и анонимизацию данных 692 студентов, реализовал полную методику обучения и сравнения алгоритмов машинного обучения (Random Forest, KNN, SVM, Gradient Boosting, Decision Tree). Им выполнен статистический анализ, визуализация результатов и интерпретация взаимосвязей между академическими и поведенческими факторами. Данная работа отражена в Главе 4 («Экспериментальные результаты и обсуждение») диссертации, где приведены сравнительные метрики и аналитические выводы.

Статья **«Development of a Hybrid Machine Learning Model for Classification of Soil Types Based on Geophysical Parameters»** (International Journal of Innovative Research and Scientific Studies, 2025, Scopus Q2) посвящена разработке гибридной модели машинного обучения. Автор диссертации занимался очисткой, предварительной обработкой и оптимизацией параметров данных, обеспечивая их корректность для обучения модели. Полученный опыт был использован при создании индивидуального дата-пайплайна, описанного в Главе 3 («Методы исследования и предварительная обработка данных»).

В статье **«Intelligent Career Path Recommendations: Leveraging Blockchain and Machine Learning»** (IEEE SIST, 2025) Автор диссертации разработал архитектуру интеграции модуля машинного обучения с блокчейн-системой с целью повышения прозрачности и достоверности данных в карьерных рекомендациях. Он принимал участие в тестировании модели и создании прототипа распределенной платформы. Результаты и технические решения данного исследования отражены в Главе 5 («Интеграция и практическая реализация модели на платформе Beta Career»).

Статья **«Identification of Key Features for Career Prediction through Recursive Feature Elimination»** (Вестник Абай атындағы ҚазҰПУ, 2025) представляет исследование по отбору признаков. Автор диссертации применил метод *Recursive Feature Elimination (RFE)* для выявления наиболее информативных предикторов, провёл корреляционный анализ академических и поведенческих характеристик, а также подготовил аналитические таблицы и графические материалы. Результаты данной работы включены в Главу 3 («Методы исследования и обработка данных»),

где подробно описан процесс отбора признаков и оптимизации параметров модели.

В статье **«Detecting Anxiety and Depression from Social Media Text by Applying Machine Learning Methods»** (Вестник АГА, 2025) Автор диссертации занимался предобработкой текстовых данных, включая токенизацию, очистку и формирование признакового пространства, а также участвовал в обучении моделей классификации на основе методов машинного обучения. Разработанные приёмы анализа неструктурированных текстовых данных были интегрированы в Главу 3 диссертации при построении пайплайна обработки поведенческих признаков.

Наконец, работа **«Comprehensive Evaluation of Real-Time Object Detection Algorithm Based on Extended Criteria»** (Вестник КазАТК, 2024) была посвящена оценке алгоритмов компьютерного зрения с расширенными критериями точности и скорости обработки. Автор диссертации участвовал в проектировании и тестировании данных алгоритмов, а также в анализе вычислительной эффективности (времени обучения и использования памяти). Полученный опыт был адаптирован в Главе 5 («Сравнительный анализ производительности моделей»), что укрепило методологическую основу диссертации в части оценки эффективности AI-моделей.

Объем и структура диссертации: Диссертация логически структурирована и состоит из восьми основных глав, за которыми следуют приложения, список литературы и заключение. Такая структура обеспечивает последовательное изложение теоретических основ, методологического дизайна, эмпирических результатов и практических рекомендаций исследования. Во введении раскрывается контекст исследования в системе высшего образования Казахстана, определяется его актуальность, формулируются тема, цели, задачи, гипотезы и инновационные аспекты диссертации. Обзор литературы содержит критический анализ предыдущих исследований, выявляет существующие научные пробелы, которые данная работа стремится восполнить, и охватывает труды, посвященные методам машинного обучения, моделям прогнозирования карьеры и применению искусственного интеллекта в образовании. Методология описывает набор данных, включающий 692 анонимизированные записи студентов с платформы Beta Career Platform. В этой главе подробно изложены процессы отбора признаков, предварительной обработки данных и стратегий моделирования. Также представлено сравнительное тестирование восьми алгоритмов машинного обучения Random Forest, KNN, SVM, Gradient Boosting, Naïve Bayes, Decision Tree, MLP и TabTransformer. В качестве усовершенствования представлена сеть Колмогорова–Арнольда (Kolmogorov–Arnold Network,

KAN), повышающая производительность и интерпретируемость моделей. Общий объем диссертации составляет около 123 страниц, включает 41 рисунок, 6 таблиц и 230 библиографических источников, что обеспечивает баланс между аналитической строгостью и наглядностью представления результатов. Такая структура гарантирует ясность, логичность изложения и всестороннее раскрытие поставленных целей и итогов исследования.