



ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – СОФИЯ

**Факултет Компютърни Системи и Технологии
Катедра Програмиране и Компютърни Технологии**

Маг. инж. Берат Тефик Уйкани

**АЛГОРИТМИ И ТЕХНОЛОГИИ ЗА ИЗКУСТВЕН ИНТЕЛЕКТ
ЗА ПРОГНОЗИРАНЕ ЕФЕКТИВНОСТТА ВЪВ ВИСШЕТО
ОБРАЗОВАНИЕ**

А В Т О Р Е Ф Е Р А Т

на дисертация за придобиване на образователна и научна степен

"ДОКТОР"

Област: 5. Технически науки

Професионално направление: 5.3. Комуникационна и компютърна техника

Научна специалност: Автоматизирани системи за обработка на информация и управление

Научн ръководители: доц. д-р инж. Даниела Минковска

проф. д-р инж. Огнян Наков

СОФИЯ, 2024 г.

Дисертационният труд е обсъден и насочен за защита от Катедрения съвет на катедра „Програмиране и компютърни технологии “ към Факултет Компютърни системи и технологии на ТУ-София на редовно заседание, проведено на 07.03.2024 г.

Публичната защита на дисертационния труд ще се състои на 20.05.2024 от 13:00 часа в Конферентната зала на БИЦ на Технически университет – София на открито заседание на научното жури, определено със заповед № ОЖ-5.3-13/19.03.2024 г. на Ректора на ТУ-София в състав:

1. доц. д-р Валентин Христов – председател
2. Доц. д-р Веска Ганчева – научен секретар
3. Проф. д-р. Десислава Петрова - Антонова
4. Проф. д-р Александър Бекярски
5. Доц. д-р Йорданка Анастасова

Рецензенти:

1. доц. д-р инж. Валентин Христов
2. доц. д-р инж. Йорданка Анастасова

Материалите по защитата са на разположение на интересуващите се в канцеларията на Факултет Компютърни системи и технологии на ТУ-София, блок № 1, кабинет 1443А.

Дисертантът е задочен докторант към катедра „Програмиране и компютърни технологии“ на факултет Компютърни системи и технологии. Изследванията по дисертационната разработка са направени от автора, като някои от тях са подкрепени от научноизследователски проекти.

Автор: маг. инж. Берат Уйкани

Заглавие: Алгоритми и технологии за изкуствен интелект за прогнозиране ефективността във висшето образование

Тираж: 25 броя

Отпечатано в ИПК на Технически университет – София

I. ОБЩИ ХАРАКТЕРИСТИКИ НА ДИСЕРТАЦИОННИЯ ТРУД

Актуалност на проблема

В бързо развиващата се област на висшето образование, характеризираща се с цифрова трансформация и все по-голямо разчитане на вземането на решения, основано на данни, дисертационният труд изследва потенциала на изкуствения интелект (ИИ) при прогнозирането на резултатите на студентите. Актуалността на изследването се подчертава от навременния му отговор на належащите нужди, пред които са изправени образователните институции днес: повишаване на ангажираността на студентите, оптимизиране на процесите на записване, намаляване на процента на отпадане и осигуряване на качество на фона на глобалната промяна към онлайн и смесена учебна среда. Използвайки машинно обучение и усъвършенствани алгоритми за дълбоко обучение, дисертационният труд има за цел да предложи точни и разбираеми прогнози, осигурявайки тяхното прозрачно прилагане. Освен това включването на система за осигуряване на качеството, базирана на обработка на естествен език, допълнително демонстрира приложното значение, демонстрирайки потенциала на ИИ за рационализиране на администрацията на образованието. Този двоен фокус върху прогностичния анализ и осигуряването на качеството, подчертава уместността на дисертационния труд за настоящите предизвикателства и възможности в рамките на висшето образование, което го прави изключително уместен и значителен принос в областта.

Цел на дисертационния труд, основни задачи и методи за изследване

Основната цел на дисертационния труд е, да се използва трансформираният потенциал на алгоритмите и техниките на ИИ в областта на висшето образование, като се съсредоточи особено върху прогностичните способности на машинното обучение и алгоритмите за дълбоко обучение.

Целите са предназначени да се приведат в съответствие с общата цел на дисертационния труд и са формулирани, както следва:

- **Изграждане на модел за прогнозиране на отпадането на студенти:** Използване на данни от развиваща се държава с един от най-добрите алгоритми за машинно обучение за двоични класификации.
- **Разработване на прогностични модели за записване на студенти:** Разработване и усъвършенстване на модели за машинно обучение, които могат точно да предскажат статуса на записване на студенти. Това включва анализиране на данни за идентифициране на успешното прогнозиране на записването на студенти на университетско ниво и провеждане на сравнителен анализ на ефективността и продуктивността на алгоритмите за машинно обучение.
- **Идентифициране и анализ на студенти, изложени на риск:** Използване на възможностите на алгоритмите за дълбоко обучение, за да бъдат идентифицирани студенти, които са изложени на риск от отпадане. Това включва анализ на поведенческите модели, нивата на ангажираност и други съответни фактори, които допринасят за процента на отпадане на студентите.
- **Прилагане на обясним ИИ за тълкуване на модели:** Осигуряване на прозрачност на моделите с ИИ чрез прилагане на нов обясним метод с ИИ. Това ще осигури ясно разбиране за вътрешната работа на модела и ще помогне да се разкрият факторите, допринасящи за успеха на студентите.
- **Предлагане на система, базирана на обработка на естествен език** за улесняване и подобряване на процесите за осигуряване на качеството във висшите учебни заведения.
- **Принос към областта на ИИ във висшето образование:** Крайната цел включва проучване на теоретичния аспект на прилагането на ИИ във висшето образование и развиващия се пейзаж на прогнозиране на резултатите, основано на ИИ, чрез анализ на литературата от платформата Web of Science.

За постигане на целите е използвана многостранна методология, обхващаща машинно обучение и алгоритми за дълбоко обучение за разработване на прогнозни модели, свързани с отпадането, записването и успеха на студентите. Тези модели бяха обучени и валидирани с помощта на образователни данни, като се гарантира тяхната точност и надеждност. Освен това, дисертационният труд включва обясними техники на ИИ, по-специално метода SHAP, за да се разберат процесите на вземане на решения на моделите на ИИ, като по този начин се повишава тяхната прозрачност. Обработката на естествен език се използва за анализ на образователното съдържание, извличане и оценка на резултатите от обучението, за да се гарантира привеждане в съответствие със стандартите на учебната програма.

Научна новост

Научната новост се крие в уникалната интеграция на съвременни техники за изкуствен интелект във висшето образование, по-специално чрез разработването на прогнозни модели, използващи машинно обучение и дълбоко обучение за отпадане, записване и успех на студенти. С него се въвежда прилагането на метода SHAP за подобряване на възможността за интерпретиране на решенията на ИИ в образователните среди. Освен това дисертационният труд е иновативен, като предлага система, базирана на обработка на естествен език за осигуряване на качеството на образователното съдържание. Този труд преодолява пропастта между теоретичния напредък в областта на ИИ и практическото му приложение във висшето образование, като предоставя значителен принос както за академичните изследвания, така и за образователната практика.

Практическа приложимост

От практическа гледна точка дисертационният труд превежда научните си прозрения в осезаеми инструменти и стратегии със значителни ползи за висшите учебни заведения. Прогнозните модели, които разработва, дават възможност за прогнозиране на отпадането, записването и успеха на студентите, улеснявайки стратегическото планиране и целенасочените интервенции. Прилагайки ХАІ, дисертационният труд гарантира, че решенията, основани на ИИ, са по-разбираеми и изпълними. Тези практически приложения подчертават влиянието на дисертационния труд върху подобряването на процесите на вземане на решения и качеството на образованието във висшите учебни заведения, демонстрирайки неговата стойност извън академичната сфера.

Апробация

Приносът на дисертационния труд е отчетен и валидиран в различни конферентни презентации, по-специално на Международна научна конференция по компютърни науки през 2023 г., XXXI Международна научна конференция по електроника през 2022 г., XXX Международна научна конференция по електроника през 2021 г. и IV Международна конференция по високи технологии за устойчиво развитие в София, България през 2021 г. Данните от Системата за управление на Университета в Митровица и набора от данни на Open University Learning Analytics са от основно значение за разработването на моделите на ИИ, като гарантират, че констатациите на дисертационния труд се основават на реални образователни контексти.

Публикации

Основните изводи от дисертационния труд са разпространени чрез четири научни труда, представени на международни конференции и индексирани в Scopus. Колективно тези публикации понастоящем имат 13 цитирания, отразяващи академичното въздействие и признанието на научните изследвания в рамките на научната общност.

Структура и обем на дисертационния труд

Дисертационният труд е в обем от 182 страници, включващ въведение, 6 глави за решаване на формулираните основни задачи, списък на основните приноси, списък на публикациите по дисертационния труд и използвана литература. Цитирани са общо 174 библиометрични източника. Трудът включва общо 70 фигури и 25 таблици. Номерата на фигурите и таблиците в автореферата съответстват на тези в дисертационния труд.

ГЛАВА 1: ПРЕГЛЕД НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ И НЕГОВОТО ПРИЛОЖЕНИЕ ВЪВ ВИСШЕТО ОБРАЗОВАНИЕ

Първоначалната глава разглежда задълбочено навлизане в света на изкуствения интелект и неговите трансформиращи приложения в сектора на висшето образование, като дава следните всеобхватни прозрения:

- Еволюцията на ИИ се очертава от създаването му като символна система до ерата на машинното обучение и дълбокото обучение. Тези фази постепенно дават възможност на машините да имитират подобни на човека способности за решаване на проблеми и вземане на решения, които сега са от основно значение за различни приложения във висшето образование.
- В центъра на темата е изследването на алгоритмите за машинно обучение и основите за обработка на естествен език, като се подчертават техните роли за целите на прогнозното моделиране и осигуряване на качеството.
- В главата се разработва задълбоченото въздействие на прогнозните анализи на ИИ в образователните среди, като се подчертава способността му да персонализира учебния процес, да оптимизира административните процеси и значително да подобри вземането на решения в институциите. Това отразява целта на дисертационния труд за използване на ИИ за осъществими, основани на данни изводи.
- Ролята на ИИ е свързана с бъдещето на висшето образование. Това включва предвиждането на етични предизвикателства и разработването на системи с ИИ, които дават приоритет на прозрачността и справедливостта. В главата се подчертава императивът за етично използване на ИИ, като се гарантира, че той служи като инструмент за приобщаване и справедливост в рамките на академичната сфера.
- И накрая, темата синтезира текущото състояние на ИИ във висшето образование с поглед в бъдещето. Тя подчертава необходимостта институциите да се адаптират към нарастващото присъствие на ИИ, като разгледат необходимостта от прозрачност.

Като цяло, първата глава задава цялостен фон за дисертационния труд, интегрирайки историческото развитие на ИИ с настоящите му приложения във висшето образование. Тя установява привиждането на способностите на ИИ в съответствие с целта на дисертационния труд за подобряване на образователните парадигми, като същевременно полага основите за подробно проучване на възможностите за прогнозиране на ИИ и неговите последващи по-широки резултати в образователната сфера.

ГЛАВА 2: БИБЛИОМЕТРИЧЕН АНАЛИЗ НА ОСНОВАНО НА ИИ ПРОГНОЗИРАНЕ НА РЕЗУЛТАТИТЕ ВЪВ ВИСШЕТО ОБРАЗОВАНИЕ

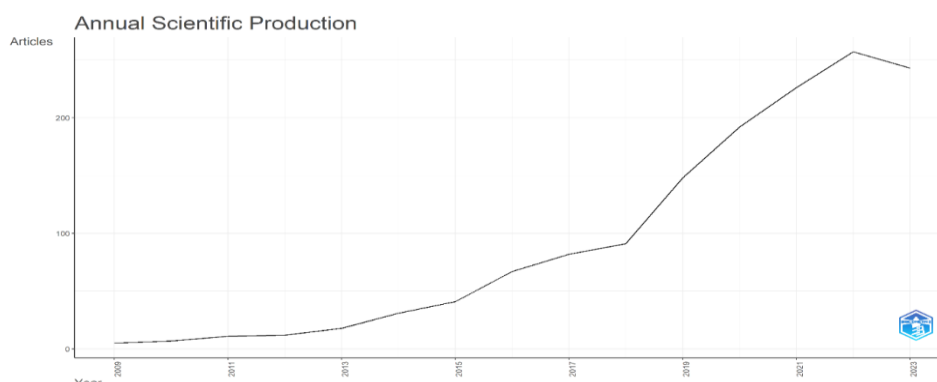
Темата представя изчерпателен библиометричен анализ на използването на изкуствен интелект при прогнозиране на резултатите на студентите във висшето образование, използвайки данни от основната колекция на Web of Science. Чрез използването на усъвършенствани библиометрични инструменти като VOSViewer и R-based Biblioshiny, изследването се задълбочава в еволюцията, тенденциите и въздействието на методологиите, движени от ИИ, в образователните среди.

Анализът обхваща 15-годишен период от 2009 г. до 2023 г., включващ набор от данни от 1431 документа, които подчертават значителен годишен темп на растеж от 31,97% в публикациите. Този растеж, особено ясно отбелязан по време на пандемията от COVID-19, подчертава нарастващото значение на ИИ в образованието на фона на предизвикателствата на дистанционното обучение. Основните констатации от изследването, обобщени в Таблица 2.2, разкриват средна честота на цитиране от 9,573% на документ, което показва влиятелния характер на изследването в тази област.

Таблица 2.2. Основна информация за извлечените документи

ОСНОВНА ИНФОРМАЦИЯ ЗА ДАННИТЕ	
Описание	Резултати
Период	2009 г. – 2023 г.
Източници (списания, книги и др.)	802
Документи	1431
Годишен темп на растеж %	31,97
Средна възраст на документа	4,15
Средни цитирания на документ	9,573
Позовавания	40 188

Фигура 2.1 илюстрира годишното производство на документи, показващи пик в публикациите през 2022 г., отразяващи динамичния характер на изследванията в тази област.



Фиг. 2.1. Годишно изготвяне на документи във времето.

Анализът се простира до идентифициране на водещи участници в областта, като в таблица 2.3 се подчертават петте най-големи институции, ръководени от Tecnológico De Monterrey, които са били от основно значение за напредъка на образователните изследвания, основани на ИИ.

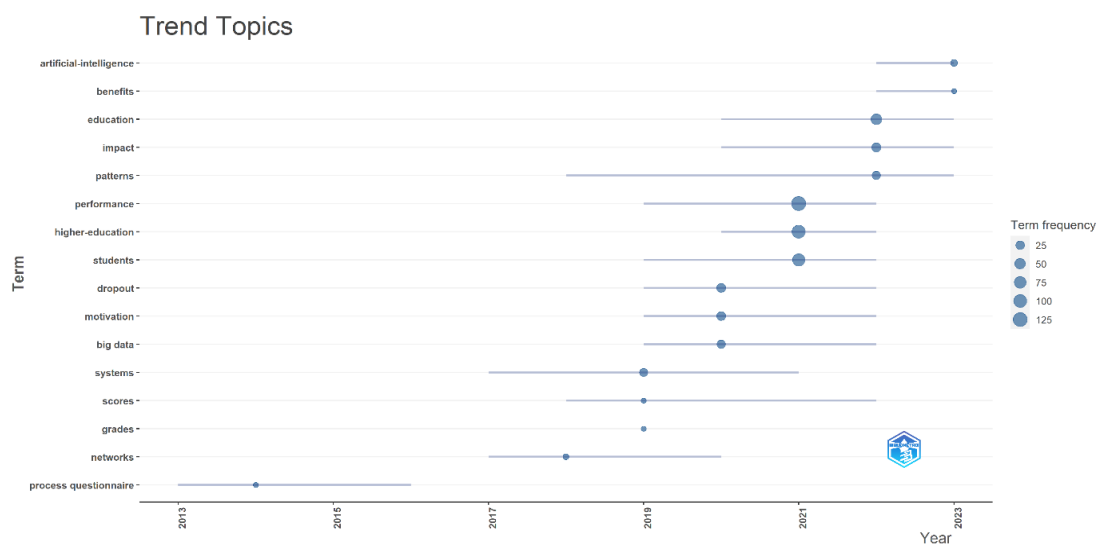
Таблица 2.3. Петте най-големи институции с най-голям брой публикации

Партньорство	Статии
Tecnologico De Monterrey	40
Университет „Крал Абдулазиз“	32
Египетска банка на знанието	18
Калифорнийски щатски университет	14
Университет на Патра	13

Освен това изследването използва визуализация на WordCloud (Фигура 2.4) и анализ на популярните теми (Фигура 2.5), за да подчертае преобладаващите теми и терминологии в корпуса, като „резултати“ и „висше образование“ са най-често срещаните термини.

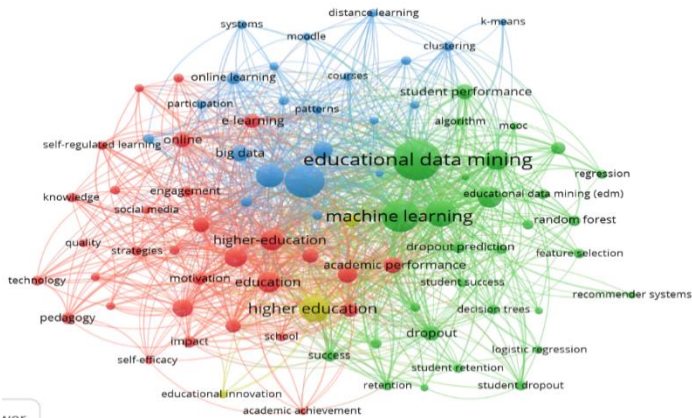


Фиг. 2.4. Картата на WordCloud

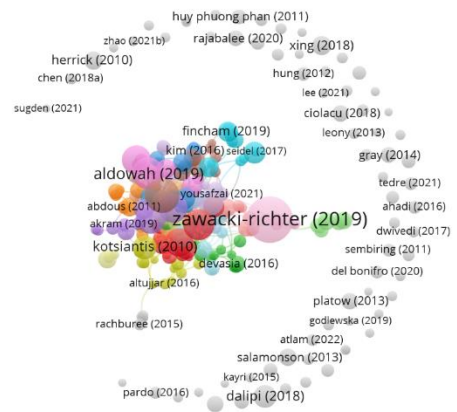


Фиг.2.5. Популярни теми във времето

Мрежовите визуализации играят решаваща роля за разбирането на взаимосвързаността в областта. Фигура 2.7 предлага анализ на съвместно появяващите се ключови думи, илюстриращ тематичните съвкупности, които определят картината на текущото изследване. По подобен начин фигура 2.8 разглежда мрежите за цитиране, като подчертава най-влиятелните документи и източници, които оформят дискурса за ИИ във висшето образование.

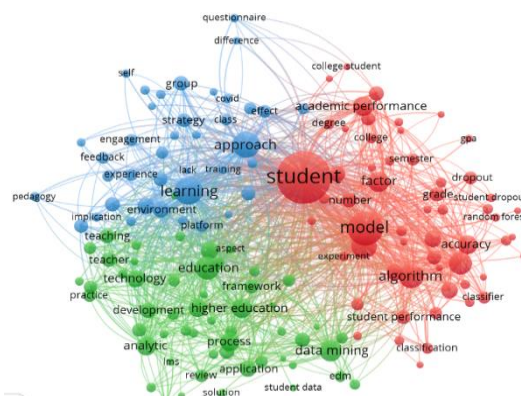


Фиг.2.7. Мрежова визуализация на съвместно появяващите се термини



Фиг.2.8. Визуализация на мрежата от цитирани документи

И накрая, картата за съпътстващо появяване на ключови думи, изобразена на фигура 2.18, предоставя графично представяне на тематичните концентрации и нововъзникващите тенденции в литературата, предлагайки общ поглед върху точките на фокусиране в областта.



Фиг.2.18. Визуализация в мрежа на ключови думи

Въпреки че проучването включва множество таблици и цифри в подкрепа на своите констатации, само избрани визуализации са графично представени, за да се поддържа яснота и фокус. Пропуснатите таблици и цифри, въпреки че не са визуално представени, допринасят за дълбочината и всеобхватността на анализа, като предлагат подробен преглед на приноса, сътрудничеството и въздействието на цитирането в областта.

В заключение, този библиометричен анализ хвърля светлина върху значителната роля на ИИ за трансформирането на висшето образование чрез прогнозиране на резултатите, като подчертава ключовите участници, институции и държави, които движат тази иновация. Констатациите не само илюстрират текущото състояние на научните изследвания, но и показват бъдещи посоки, подчертавайки потенциала на ИИ в прогнозирането на резултатите във висшето образование.

ТЕМА 3: ПРОГНОЗИРАНЕ НА ОТПАДАНЕТО НА СТУДЕНТИ: ПОДХОД С ЛОГИСТИЧНА РЕГРЕСИЯ

Тази тема разглежда критичния въпрос за отпадането на студенти във висшето образование, като използва логистична регресия, за да анализира данни от държавен университет в Косово. Това новаторско изследване в контекста на Косово има за цел да идентифицира факторите, допринасящи за процента на отпадане, и точно да предскаже статуса на отпадане на студентите.

Данни

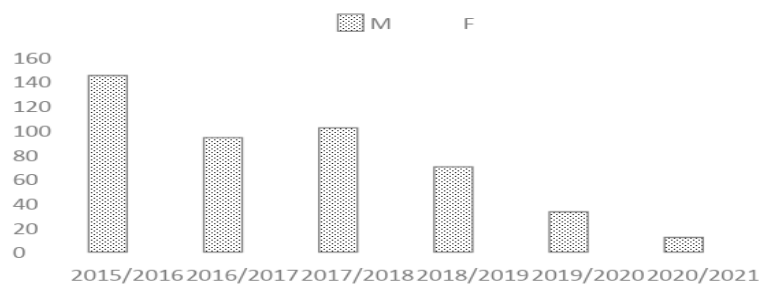
Изследването използва набор от данни, извлечени от системата за управление на студентите на държавния университет в Митровица, обхващащ шест академични години от 2015/2016 г. до 2020/2021 г., първоначално включващи 4818 записа. След почистването на данните окончателният набор от данни включва 4697 записа, обхващащи атрибути като пол, възраст, град, дата на записване, ниво на обучение, изпити, които трябва да се положат, издържани изпити, средна оценка и текущ статус, с добавен атрибут на двоичен клас за статус на отпадане, както е описано подробно в таблица 3.1.

Таблица 3.1. Атрибути, видове и възможни стойности в набор от данни

Атрибут	Вид	Стойност
Ид. №	числов	число
Град	име	низ
Катедра	име	0 = Компютърни науки и инженерство 1 = Производствено инженерство 2 = Икономическо инженерство
Записване	дата	01.10.2015 г. – 30.09.2021 г.
Ниво на обучение	име	0= Бакалавър 1 = Магистър
Възраст	числов	18–35
Пол	име	0= мъж 1 = жена
Брой изпити	числов	15 – 30
Издържани изпити	числов	1 – 30
Средна оценка	числов	6 – 10
Статус	име	0= Завършил 1 = Активен 2 = Неактивен
Клас	име	0 = Отпадане 1 = Няма отпадане

Проучване на данни

Статистическият анализ на данните разкрива значителни резултати. Фигура 3.1 илюстрира процента на задържане и отпадане по пол в различните випуски, като разкрива забележителен пик в отпадането на студенти жени през учебната 2017/2018 година.



Фиг. 3.1. Процент на задържане и отпадане по випуски

Освен това в таблица 3.2 се подчертават процентите на отпадане във връзка с броя на положените изпити, което разкрива значителна корелация между по-малко положени изпити и по-високи проценти на отпадане, особено през учебната 2017/2018 година. Общият процент на отпадане е изчислен на 23,74%.

Таблица 3.2. Брой отпадания по випуск

Учебна година	Успешно положени изпити		
	<5	5–10	над 10
2015/2016	86	70	40
2016/2017	88	37	20
2017/2018	122	22	11
2018/2019	33	3	13
2019/2020	10	7	4

Методология

Проучването възприема двуетапен подход: първоначално събиране и предварителна обработка на данни, последвано от прилагане на логистична регресия за прогнозиране на отпадането на студентите. Моделът на двоична класификация е обучен на характеристики като резултати на изпити и средни оценки, като логистичната регресия е избрана заради нейната пригодност за категорични резултати. Отпадането се определя като напускане на образователната система без получаване на степен, с изключение на студенти, които сменят области или видове обучение.

Логистична регресия

Логистичната регресия се използва за прогнозиране на отпадналите студенти въз основа на анализирания атрибут от набора от данни, като се използва Python заедно с библиотеки като NumPy и Scikit-Learn за обработка на данни и разработване на модели. Наборът от данни е разделен на съотношение 70:30 за обучение и тестване, с категорични етикети, кодирани за обозначаване на статуси отпаднали и неотпаднали.

Логистичното регресионно уравнение, от централно значение за модела, е следното:

$$\hat{y} = \frac{e^{xW}}{\sum_j e^{xW}} \quad (0.1)$$

където \hat{y} са прогнозите, X входните данни и W теглата.

Резултатите под формата на получени прогнози трябва да бъдат сравнени с целевите стойности, като се използва обективната функция за загуба на кръстосана ентропия.

$$J(\theta) = -\sum_i \ln(\hat{y}_i) = -\sum_i \ln\left(\frac{e^{x_i W}}{\sum_j e^{x_j W}}\right) \quad (0.2)$$

Оценка на модела

Важна стъпка при изпълнението на прогноза в машинното обучение е процесът на оценка и проверка дали използваната техника се представя по подходящ начин. За измерване на ефективността и производителността на модела са използвани следните показатели: матрица на объркване, точност, пълнота, прецизност и F_1 метрика. Кратко обяснение на всеки от показателите, като в [133], е дадено в редовете по-долу.

Процентът на истинските положителни резултати (TP) представлява броя на правилно прогнозираните проби, докато процентът на фалшивите положителни резултати (FP) представлява броя на фалшивите положителни резултати и те се определят както е посочено по-долу.

$$TP = \frac{TP}{TP+FN} \quad (0.3)$$

$$FP = \frac{FP}{FP+TN} \quad (0.4)$$

Точност представлява общата точност на прогнозата и се изчислява като съотношение между правилно класифицираните прогнози и общия брой случаи.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (0.5)$$

Пълнотата се фокусира само върху правилно класифицирани прогнози. Тя измерва вероятността отпадането на студент да бъде класифицирано правилно измежду всички правилно положителни и отрицателни класификации.

$$Recall = \frac{TP}{TP+FN} \quad (0.6)$$

Условната вероятност студент, който е класифициран като отпаднал, да е точно класифициран, се измерва с помощта на метриката точност. Този измерител също е много полезен, за да се избегнат неправилни прогнози.

$$Precision = \frac{TP}{TP+FP} \quad (0.7)$$

Последният измерител е оценка F_1 , която намира балансираните мерки от резултатите от модела. Той дава 0, когато моделът основно не работи, до 1, когато моделът работи перфектно.

$$F_1 = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (0.8)$$

Наблюдавайки показателите за оценка в таблица 3.3, правилно класифицираните стойности могат да се видят в диагонала, а предложеният модел постига точност от 90%.

Таблица 2.3. Ефективност на модела

	Точност	Прецизност	Пълнота	F1
%	0,90	0,89	0,81	0,85

Обсъждане

Констатациите подчертават потенциала на логистичната регресия при точното прогнозиране на отпадането на студентите, предлагайки ценен инструмент за образователните институции за проактивно идентифициране на студентите, изложени на риск. Високата точност и F_1 резултатът, постигнат от модела, подчертават неговата ефективност в образователния контекст, полагайки основите за бъдещо изследване, което да се разшири чрез включване на повече променливи и задълбочаване на причините за отпадането на студентите.

Като цяло това новаторско проучване не само предоставя проницателни анализи за процента на отпадане на студенти в Косово, но и създава прецедент за прилагането на логистична регресия в образователните изследвания, с потенциал да повлияе на разработването на политики и стратегическото планиране във висшите учебни заведения.

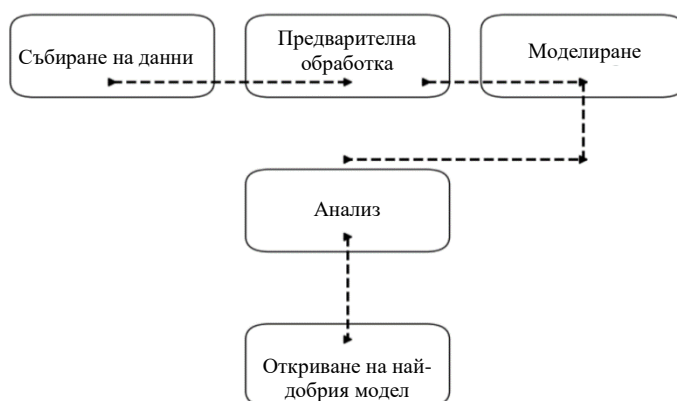
ГЛАВА 4: ПРОГНОЗИРАНЕ НА ЗАПИСВАНЕТО НА СТУДЕНТИ В УНИВЕРСИТЕТА: СРАВНИТЕЛЕН АНАЛИЗ НА АЛГОРИТМИТЕ ЗА МАШИННО ОБУЧЕНИЕ

Тук е разгледана в дълбочина друга важна тема, свързана със записването на студенти в университета. Използват се различни техники за машинно обучение, за да се предскаже статусът на записване в университет. Използвайки набор от данни от университета „Иса Болетини“ в Митровица, Косово, това изследване има за цел да изгради точен и надежден прогнозен модел чрез сравняване на работата на различни алгоритми.

Данни и методология

Изследването е методично разделено на две основни фази: Компиляция и анализ на набори от данни и Сравнение на модели. Първоначално данните се събират и обработват предварително, за да се осигури чист и надежден набор от данни за анализ. Впоследствие проучването използва софтуера Weka за анализ и оценка на ефективността на избрани алгоритми за машинно обучение: наивен Байес, логистична регресия, дървета за решения и К-най-близки съседи (KNN).

Чрез тези фази методологията използва силата на контролираните техники за обучение и усъвършенстваните възможности на софтуера Weka за справяне с проблема с прогнозирането на записването в университет. Този структуриран подход гарантира задълбочеността на нашия анализ, но също така допринася за разработването на високо-надежден прогнозен модел. Фигура 4.1., по-долу, описва петте фази на изследването.



Фиг. 4.1. Петте фази на това изследване

Събиране и предварителна обработка на данни

Данните се събират от Системата за управление на студентите на университета, обхващаща три академични години и четири академични звена. Фазата на предварителна обработка включва почистване на данните, за да се премахнат непълните записи, което води до 2338 пълни и валидни записи за анализ. Таблица 4.1. предоставя преглед на атрибутите на набора от данни.

Таблица 3.1. Общ преглед на набора от данни

Атрибути	Вид	Описание
10-ти клас	число	Оценка за учебни резултати в 10-ти клас, по скала от 2 (най-ниска) до 5 (най-висока).
11-ти клас	число	Оценка за учебни резултати в 11 клас, по скала от 2 до 5.
12-ти клас	число	Оценка за учебните резултати в 12 клас, по скала от 2 до 5.

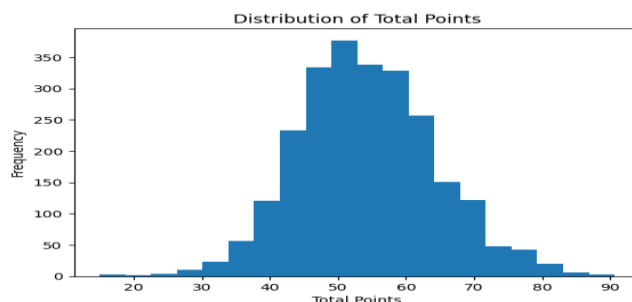
13-ти клас	число	Оценка на учебните резултати в 13-ти клас, обикновено последната година от гимназията, по скала от 2 до 5.
Успех в гимназията	число	Обща оценка за успех в гимназията, обобщаващ резултати по различни предмети, по скала от 6 (най-ниска) до 20 (най-висока).
Точки на държавен зрелостен изпит	число	Точки, получени на матура – стандартизиран тест за гимназисти, с резултати, вариращи от 0 (най-ниски) до 40 (най-високи).
Точки на приемния изпит	число	Точки, отбелязани на приемния изпит в университета, по скала от 0 до 40.
Общо	число	Кумулативен резултат, съчетаващ успех в гимназията, точки на държавен зрелостен изпит и точки от приемния изпит, по скала от 0 (най-нисък) до 100 (най-висок).
Клас	име	Статус на записване на студента, като се посочва дали студентът е бил записан (Успешно) или не (Неуспешно) въз основа на общите точки.

Надеждността на набора от данни се измерва с Алфа коефициента на Кронбах, като се получава стойност 0,83, което показва добро ниво на съгласуваност.

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum V_i^2}{V_T^2} \right)$$

Анализ на разпределението

Разпределението на атрибутите дава ценна представа за обхвата и сходството на стойностите в набора от данни. Например, при визуално изследване чрез хистограма се наблюдава, че разпределението на „Общите“ точки се доближава до нормално разпределение, като средната стойност е центрирана около 50. Това предполага, че резултатите са симетрично разпределени около тази централна стойност, като честотата на резултатите намалява еднакво в двете посоки от средната стойност. В нормално разпределен набор от данни повечето от точките от данни са близо до средната стойност, със стандартно отклонение, което дава представа за променливостта на набора от данни.



Фигура 4.2. Разпределение на точки „Общо“

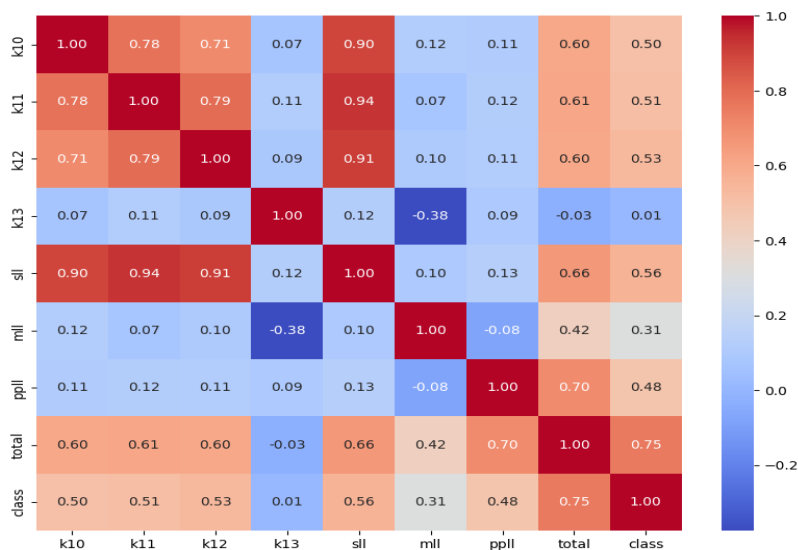
Корелационен анализ

Във връзка с анализа на разпределението, оценката на корелацията между различните атрибути е от съществено значение. Корелационната матрица може да служи като мощен инструмент в този аспект. Тя количествено определя линейните връзки между двойки атрибути, като предлага ясен поглед върху това как една променлива може да предскаже или да бъде свързана с друга. За прогнозното моделиране е необходимо разбиране на тези корелации, тъй като те могат да повлияят на избора на характеристики, които да бъдат включени в модела. Силните корелации между независимите променливи (мултиколинеарност) могат да повлияят на производителността и интерпретацията на някои модели, докато силната корелация между независима променлива и зависимата променлива може да показва добър предиктор.

Корелационната матрица, количествено представяне на тези взаимоотношения, показва корелационните коефициенти между всички атрибути. Особено забележима е корелацията между точките „Общо“ и всеки от учебните класове от 10-ти до 13-ти клас, както и точките за матура и входящ изпит. Тези корелации са от съществено значение, тъй като предлагат прозрения за това кои фактори са най-предсказуеми за статуса на записване на студента.

Интересното е, че докато повечето атрибути показват положителна корелация с точките „Общо“, има един атрибут (13-ти клас), който се откроява със значително по-ниска корелация. Това предполага, че за разлика от другите, този атрибут има по-слаба линейна връзка с точките „Общо“. Идентифицирането на такъв атрибут е от решаващо значение, тъй като може да сочи към области, в които прогнозният модел трябва да бъде нюансиран или където може да са необходими допълнителни данни за цялостен анализ.

Освен това корелациите между оценките за 10-ти (K10), 11-ти (K11) и 12-ти (K12) класове са изключително силни. Това показва значително ниво на последователност в учебните постижения през тези години, което предполага, че учениците, които се представят добре през една година, вероятно ще се представят добре през следващите години. Такава последователност засилва предсказващата сила на тези атрибути, тъй като те не само индивидуално корелират с точките „Общо“, но и взаимно корелират, осигурявайки многоизмерен поглед върху представянето на учениците във времето.



Фигура 4.3. Топлинна карта на корелационната матрица

Намиране на най-добрия модел

За да се определи най-добрият модел, се използва следната стратегия:

1. **Обучете множество модели**, като използвате различни алгоритми и хиперпараметри.

2. **Оценете всеки модел**, като използвате описаните по-горе показатели, като използвате кръстосано валидиране, за да се уверите, че производителността на модела е последователна в различните подгрупи от данни.
3. **Сравнете моделите въз основа на тези показатели**. Най-добрият модел не винаги е този с най-висок реален положителен процент или прецизност; това зависи от специфичните нужди на проекта. Например, в медицинската диагноза, пълнотата може да бъде приоритизирана, за да се гарантира възможно най-малко пропуснати диагнози.
4. **Коригиране на прага за вземане на решение**, както е необходимо, което може да премести компромиса между фалшиви положителни и фалшиви отрицателни резултати, за да приспособи ефективността на модела към конкретното приложение.
5. **Помислете за контекста** на проблема, разходите, свързани с фалшиви положителни и фалшиви отрицателни резултати, и баланса между различните показатели.

Моделиране в Weka

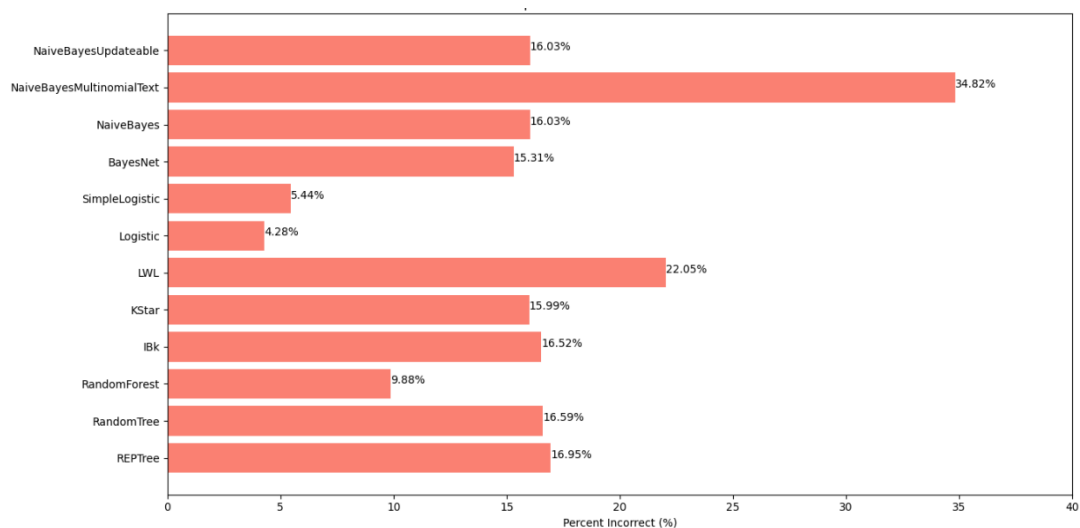
Интерфейсът на софтуера Weka се използва за обучение и анализ на модели, като се извършва настройка на хиперпараметъра, за да се оптимизират прогнозните резултати на всеки модел. Различни алгоритми се подлагат на оптимизация на хиперпараметрите, за да се подобри тяхната производителност. Например, точността на алгоритъма на BayesNet е значително подобрена с помощта на алгоритъма за търсене LAGDHillClimber, постигайки точност от 93,1138%, както е показано в таблица 4.2.

Таблица 4.2. Точност на BayesNet в различните алгоритми за търсене

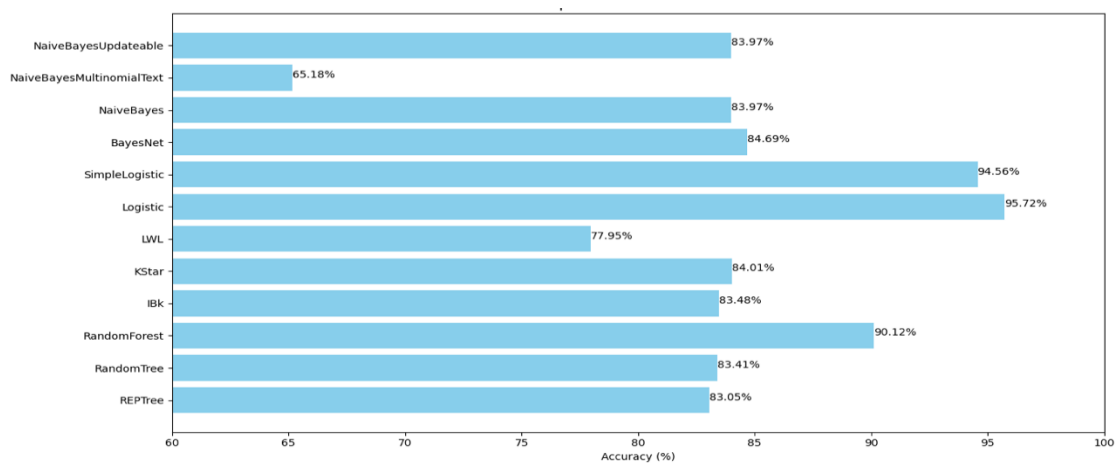
Алгоритъм за търсене	Точност
K2	84,8161 %
GeneticSearch	89,2216%
LAGDHillClimber	93,1138%
HillClimber	84,8161 %
симулирано Annealing	91,6595 %

Резултати и дискусия

Изследването прилага 12 алгоритъма в четири техники за машинно обучение, анализирайки тяхното представяне при прогнозиране на записването на студенти. Логистичният модел се очертава като най-точен, с 95,72% точност, докато моделът NaiveBayesMultinomialText показва най-ниска точност – 65,18%. Представянето на всички модели се сравнява изчерпателно, като моделите Logistic и SimpleLogistic демонстрират изключителна производителност по различни показатели, както е илюстрирано на фигури 4.16 до 4.21 и обобщено в таблица 4.4.



Фигура 4.16. Сравнение на неправилен процент на моделите



Фигура 4.17. Сравнение на точността на модела

Таблица 4.4. Сравнение на производителността на моделите

Модел	Точност	Процентът неправилни	F-стойност	IR прецизност	IR пълнота
REPTree	83,05	16,95	0,77	0,74	0,80
RandomTree	83,41	16,59	0,77	0,75	0,80
RandomForest	90,12	9,88	0,86	0,87	0,85
IBk	83,48	16,52	0,77	0,74	0,81
KStar	84,01	15,99	0,77	0,78	0,76
LWL	77,95	22,05	0,63	0,76	0,54
Logistic	95,72	4,28	0,94	0,92	0,96

SimpleLogistic	94,56	5,44	0,92	0,93	0,92
BayesNet	84,69	15,31	0,80	0,74	0,87
NaiveBayes	83,97	16,03	0,79	0,73	0,87
NaiveBayesMultinomial Text	65,18	34,82	0,79	0,73	0,00
NaiveBayesUpdateable	83,97	16,03	0,79	0,73	0,87

Заклучение

Това изследване изследва възможността за прогнозиране на записването на студентите в университета, като се вземат предвид точки на приемните изпити, резултатите от зрелостните изпити и предишните гимназиални оценки.

Предложеният модел беше обучен и тестван с данни от предходните три академични години, използвайки четири добре познати техники за контролирано машинно обучение и общо 12 алгоритъма като логистичен, K-най-близки и др.

В крайна сметка резултатите бяха анализирани и сравнени. Всяка техника се изпълнява по различен начин, което води до тесен диапазон от точности. RepTree от логистични алгоритми надмина всички други алгоритми демонстрира най-добър резултат, т.е. най-висока точност на прогнозиране от над 95% и най-нисък процент на грешки. И обратно, алгоритъмът NaiveMulti се оказа най-зле представящият се, тъй като излезе с прогнози с под 80% точност.

ГЛАВА 5: ПРОГНОЗИРАНЕ НА УСПЕХА ОТ КУРСА И РАННО ИДЕНТИФИЦИРАНЕ НА ИЗЛОЖЕНИТЕ НА РИСК СТУДЕНТИ С ПОМОЩТА НА ОБЯСНИМ ИЗКУСТВЕН ИНТЕЛЕКТ

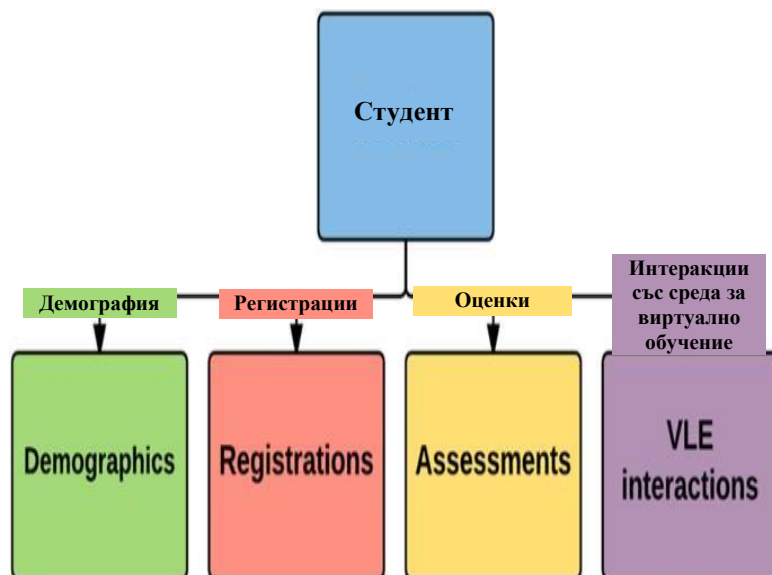
Пета тема се впуска в нюансирано проучване на използването на машинно обучение и алгоритми за дълбоко обучение за прогнозиране на представянето на студентите в контекста на онлайн учебната среда. Отличителният белег на това изследване е акцентът му върху интерпретируемостта на предсказуемите модели чрез прилагането на SHAP (SHapley Additive exPlanations), като се обръща внимание на критичната нужда от прозрачност в образователните анализи, задвижвани от ИИ.

Методология

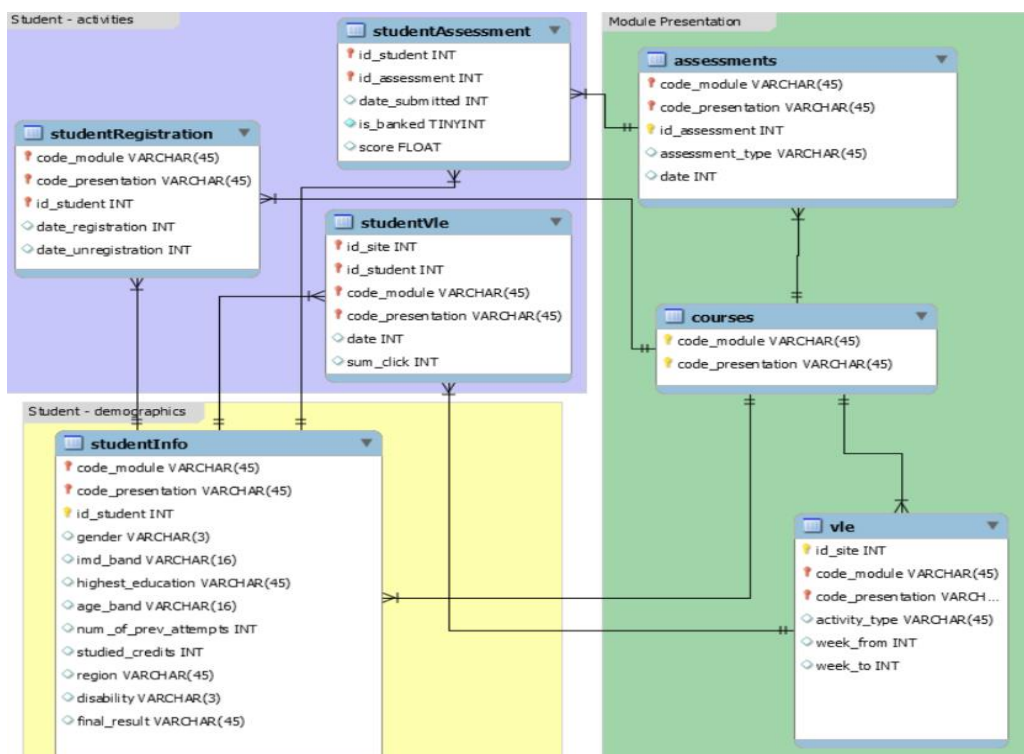
Изследването приема изчерпателна методология, която включва избор на набор от данни, обширна предварителна обработка на данни, сложно разработване на модели, настройка на хиперпараметри, цялостна оценка на модела и иновативно прилагане на метода SHAP за интерпретация на модели.

Избор на набор от данни

Изследването използва набора от данни на Open University Learning Analytics, който съдържа подробни записи на 32 593 студенти за 9-месечен период на обучение от 2014 до 2015 г. Илюстриран на фигура 5.1, този набор от данни е богат на демографски данни за студентите, записвания на курсове, резултати от оценяването и взаимодействия с виртуалната среда за обучение (VLE), осигурявайки солидна основа за прогнозно моделиране.



Фигура 5.1. Структура на набора от данни



Фигура 5.2. Схема на базата данни на Open University Learning Analytics

Обработка на данни

Фазата на предварителна обработка е от решаващо значение за подготовката на набора от данни за анализ, включващ поредица от стъпки, предназначени за обединяване на таблици, обработка на липсващи стойности, инженерни функции и ефективно разделяне на данните.

Предварителна обработка за прогнозиране на успеха на курсове

Това води до цялостен процес на сливане за интегриране на информацията за студентите с техните данни за оценки и виртуална среда за обучение, допълнително обогатени чрез инженеринг на функциите и кодиране на категорични променливи. Наборът от данни след предварителната обработка, подробно описан в Таблица 5.1, се появява с 68 характеристики, показателни за различни фактори, които потенциално влияят върху успеха на студентите в курсовете.

Таблица 5.1. Окончателен набор от функции

Номер на характеристика	Име на характеристиката	Не-нулев брой	Тип данни	Описание
0	code_module	180662	обект	Идентификатор за модула на курса.
1	code_presentation	180662	обект	Идентификатор за представяне на курса (семестър).
2	id_student	180662	int64	Уникален идентификатор за всеки студент.
3	gender	180662	обект	Пол на студента.
4	region	180662	обект	Географски район на пребиваване на студента.
5	highest_education	180662	обект	Най-високото ниво на образование, постигнато от студента, преди да вземе модула.
6	imd_band	172852	обект	Индекс на групата на множествените лишения за пребиваването на студента.
7	age_band	180662	обект	Възрастова категория на студента.
8	num_of_prev_attempts	180662	int64	Колко пъти студентът е опитвал модула преди.
9	studied_credits	180662	int64	Общ брой кредити, които студентът изучава през този семестър.
10	disability	180662	обект	Показва дали студентът е декларирал увреждане.
11	final_result	180662	обект	Окончателен резултат в модула (Успешно, Неуспешно, Оттеглил се и т.н.).
12	date_registration	180614	float64	Брой дни, в които студентът се е регистрирал преди началната дата на модула.
13	date_unregistration	18454	float64	Брой дни, след които студентът се е deregистриран от модула след началната дата (ако е приложимо).

14	id_valuation	173912	float64	Уникален идентификатор за всяка оценка.
15	date_submitted	173912	float64	Брой дни, в които оценката е подадена преди/след крайния срок.
16	is_banked	173912	float64	Показва дали оценката е натрупана за бъдещи опити.
17	score	173739	float64	Оценка, постигната от студента при оценяването.
18	assessment_type	173912	обект	Вид на оценката (оценка от преподавател (ТМА), оценка от компютър (СМА), изпит).
19	date	171047	float64	Насрочена дата на оценката.
20	weight	173912	float64	Тежест на оценката при изчисляване на крайния резултат от модула.
21	total_clicks	177179	float64	Общ брой кликания, направени от студента във виртуалната среда за обучение.
22	module_presentation_length	180662	int64	Дължина на представянето на модула в дни.

Предварителна обработка за ранно прогнозиране на цялостния успех на изследването

Насочена към ранните показатели за ангажираност и изпълнение, тази стратегия за предварителна обработка е пригодена да прогнозира дългосрочните академични постижения. Тя отразява предварителната обработка на прогнозата за успех на курса при сливането, инженеринга на функциите и разделянето на данните, макар и съобразени с дългосрочните показатели за успех.

Изграждане на прогнозен модел

За точното прогнозиране на резултатите на учениците бяха използвани редица модели за машинно обучение и дълбоко обучение:

- **Множествени методи:** Използвани са класификатор случайна гора и XGBClassifier заради способността им да се справят със сложни структури от данни и да предоставят стабилни прогнози.
- **Последователни модели:** Приложен е класификатор градиентно усилване заради неговата ефективност при работа с таблични данни.
- **Базирано на инстанции обучение:** k-най-близки съседи (k-NN) беше използван за изследване на влиянието на мерките за сходство върху точността на прогнозите.
- **Архитектури на невронни мрежи:** Разработени модели, използващи многослоев Персептрон, персонализирани невронни мрежи, съобразени с характеристиките на нашия набор от данни, дълга краткосрочна памет (LSTM) и

конволюционни невронни мрежи (CNN) за улавяне на времеви и пространствени зависимости при взаимодействия с виртуалната среда за обучение.

Моделиране и настройка на хиперпараметъра

Проучването изследва редица прогнозни модели, включително ансамблови методи, последователни модели, учене, базирано на инстанции, и разработени по поръчка архитектури на невронни мрежи. Проведена е настройка на хиперпараметъра, описана подробно в Таблица 5.2 и Таблица 5.3, за да се прецизират характеристиките на всеки модел, като се оптимизира въз основа на точността, прецизността, припомнянето и резултата F1.

Първи персонализиран модел

Таблица 5.2. Резултати от фина настройка на първия модел

Брой слоеве	Максимален брой филтри	Брой епохи	Точност на обучение	Точност на валидиране
2	32	10	85%	74%
2	32	20	87%	77%
2	32	30	88%	76%
3	64	10	92	84%
3	64	20	95%	90%
3	64	30	96%	89%
4	128	10	93%	85%
4	128	20	96%	91%
4	128	30	97	93%

Ранно прогнозиране на цялостния успех на изследването

Втори персонализиран модел

Таблица 5.3. Резултати от фината настройка на втория модел

Брой слоеве	Максимален брой филтри	Брой епохи	Точност на обучение	Точност на валидиране
2	32	10	81%	81%
2	32	20	82%	81%
2	32	30	83%	82%
3	64	10	84%	82%
3	64	20	87%	83%
3	64	30	89%	85%
4	128	10	89%	85%
4	128	20	94%	88%
4	128	30	96%	93%

Оценка на резултатите от прогнозирането

Моделите се оценяват с помощта на стандартни показатели, за да се установи тяхната ефикасност при прогнозирането на успеха на студентите и идентифицирането на студентите, изложени на риск. Тази оценка, обобщена в таблици 5.4 до 5.7, улеснява избора на най-ефективния модел за целите на изследването.

Таблица 5.4. Точност на обучените модели

Модел	Точност
първи персонализиран модел на невронна мрежа	94%
Класификатор случайна гора	84%
Класификатор градиентно усилване	65%
Класификатор k-най-близки съседни (k-NN)	79%
Класификатор многослоен перцептрон	85%

Таблица 5.5. Прецизност

Клас\модел	първи персонализиран модел на невронна мрежа	Случаен	Класификатор градиентно усилване	Класификатор k-най-близки съседни (k-NN)	Класификатор многослоен перцептрон
Неуспешно полагане	0,89	0,89	0,41	0,62	0,55
Оттегляне	0,85	0,88	0,48	0,62	0,54
Успешно полагане	0,97	0,90	0,79	0,95	0,89
С отличие	0,99	1,0	1,0	0,88	0,99

Таблица 5.6. Пълнота

Клас\модел	първи персонализиран модел на невронна мрежа	Случаен	Класификатор градиентно усилване	Класификатор k-най-близки съседни (k-NN)	Класификатор многослоен перцептрон
Неуспешно полагане	0,95	0,69	0,66	0,93	0,81
Оттегляне	0,92	0,66	0,66	0,86	0,81
Успешно полагане	0,93	0,91	0,59	0,72	0,66
С отличие	0,99	0,99	0,99	0,91	0,99

Таблица 5.7. F1 резултат

Клас\модел	първи персонализиран модел на невронна мрежа	Случаен	Класификатор градиентно усилване	Класификатор k-най-близки съседни (k-NN)	Класификатор многослоен перцептрон
Неуспешно полагане	0,92	0,78	0,51	0,74	0,66
Оттегляне	0,88	0,75	0,56	0,72	0,65
Успешно полагане	0,95	0,91	0,67	0,82	0,76
С отличие	0,99	0,99	0,99	0,90	0,99

Ранно прогнозиране на цялостния успех на изследването

Използвани са осем модела за оценка на тяхната ефективност върху набора от данни. Всеки модел беше оценен въз основа на прецизност, пълнота и F1 резултат, със специфичен акцент върху способността им да класифицират инстанциите като „Успешно“ или „Неуспешно“. Матриците за объркване предоставиха визуална представа за възможностите за прогнозиране на всеки модел.

Таблица 5.8. Точност на обучените модели

Модел	Точност
Класификатор случайна гора	86%
Класификатор градиентно усилване	81%
k-най-близки съседни (k-NN)	82%
Многослоен перцептрон (невронна мрежа)	85%
Втори персонализиран модел	93%
Convolutional Neural Network (CNN)	86%
Дълга краткосрочна памет (LSTM)	81%
XGBoost	86%

Таблица 5.9. Прецизност

	Класификатор случайна гора	Класификатор градиентно усилване	k-NN	MLP	Втори персонализиран модел	CNN	LSTM	XGBoost
Успешно полагане	0,85	0,72	0,7	0,74	0,88	0,78	0,63	0,80

Неуспешно полагане	0,86	0,83	0,84	0,88	0,95	0,87	0,84	0,87
--------------------	------	------	------	------	------	------	------	------

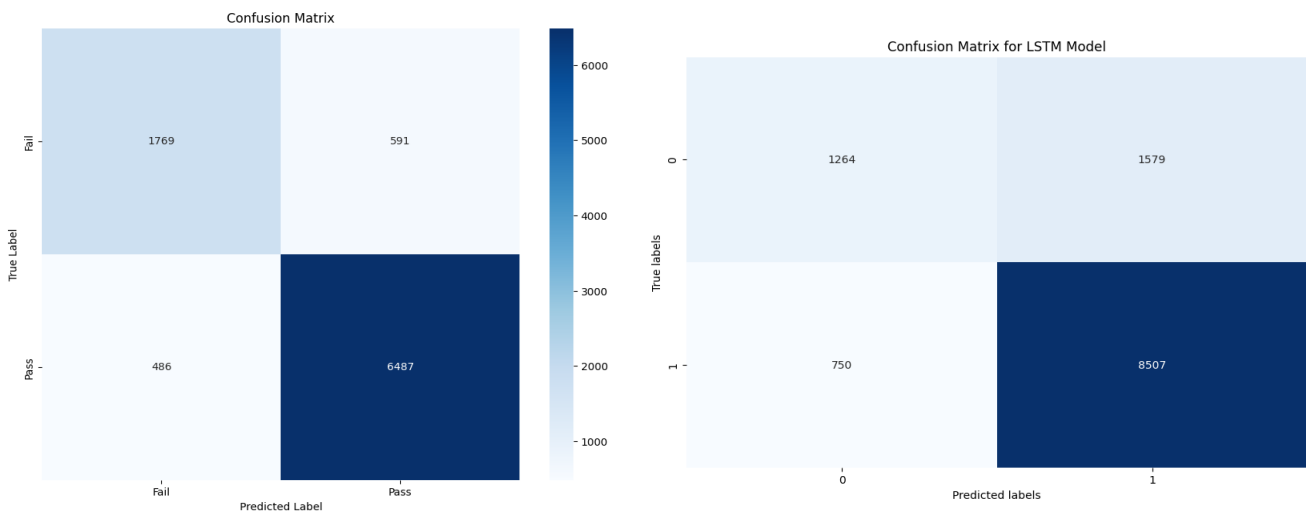
Таблица 5.10. Пълнота

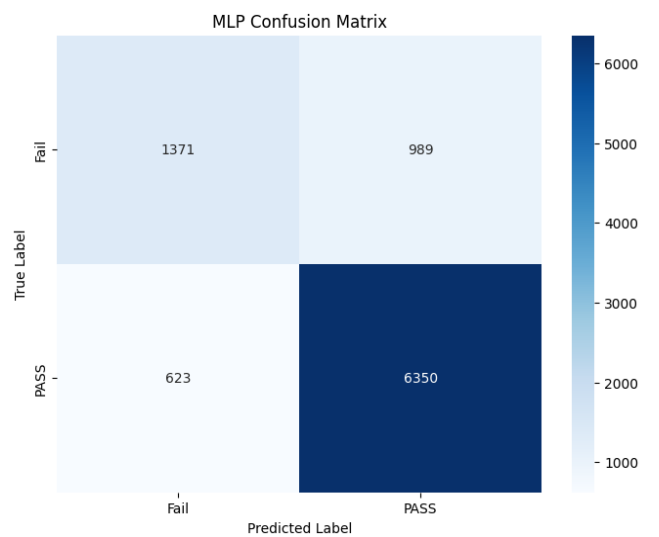
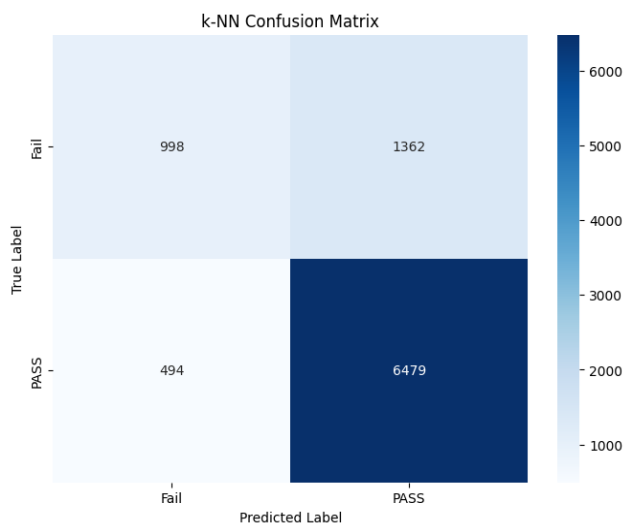
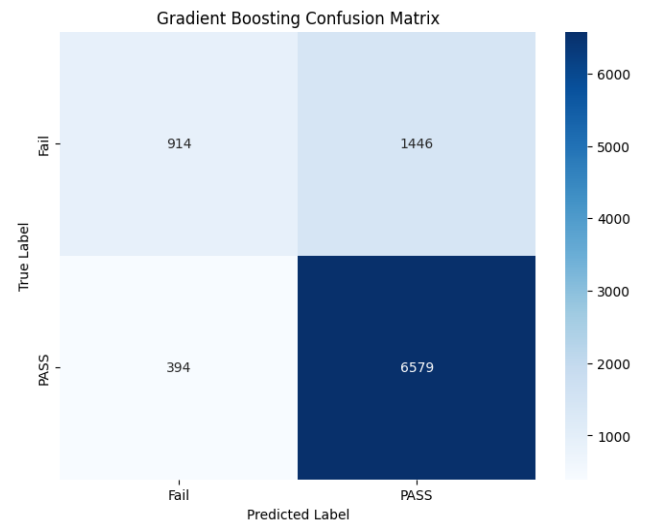
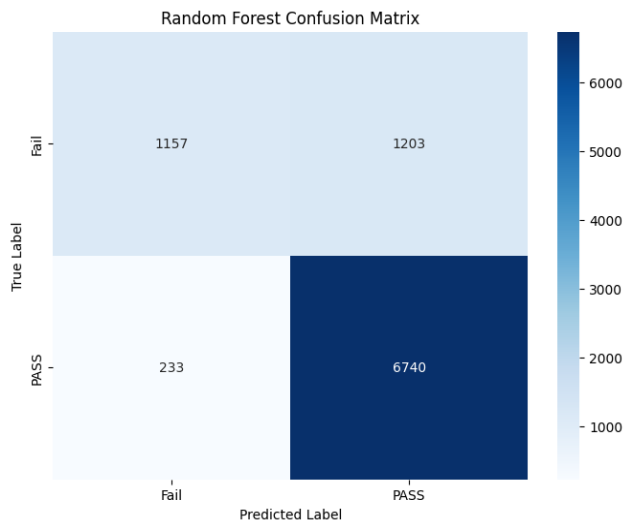
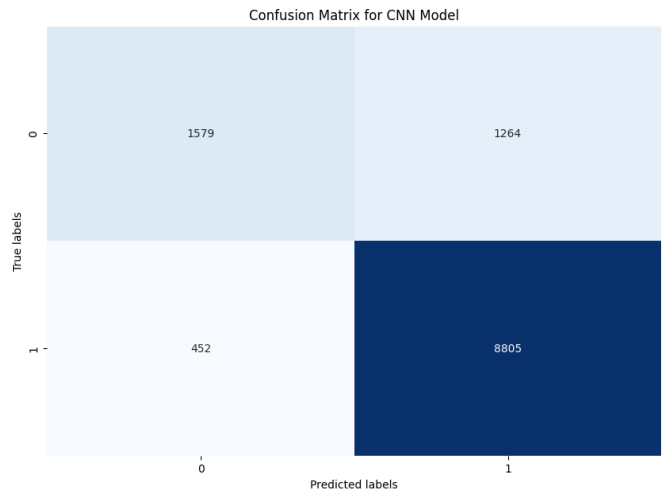
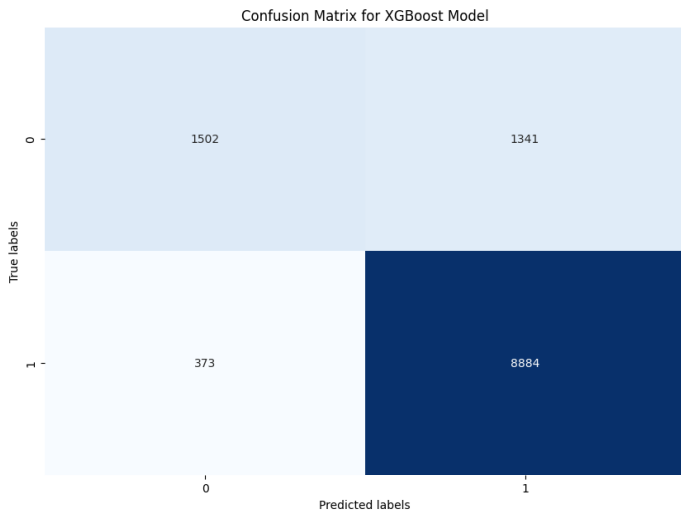
	Класификатор случайна гора	Класификатор градиентно усилване	k-NN	MLP	Втори персонализиран модел	CNN	STM	XGBoost
Успешно полагане	0,47	0,35	0,43	0,57	0,82	0,56	0,44	0,53
Неуспешно полагане	0,97	0,96	0,94	0,94	0,97	0,95	0,92	0,96

Таблица 5.11. F1 резултат

	Класификатор случайна гора	Класификатор градиентно усилване	k-NN	MLP	Втори персонализиран модел	CNN	LSTM	XGBoost
Успешно полагане	0,61	0,47	0,54	0,65	0,85	0,65	0,52	0,64
Неуспешно полагане	0,91	0,89	0,89	0,91	0,96	0,91	0,88	0,91

Очевидно е, че вторият персонализиран модел има най-добра производителност по различни показатели. С точност от 94%, превъзходната производителност на персонализираните модели може да бъде приписана на тяхната персонализирана настройка, за да отговаря на наличния набор от данни. За разлика от стандартните модели като класификатор случайна гора и класификатор градиентно усилване, персонализираните модели са проектирани и оптимизирани специално за характеристиките и тънкостите на набора от данни OULA.





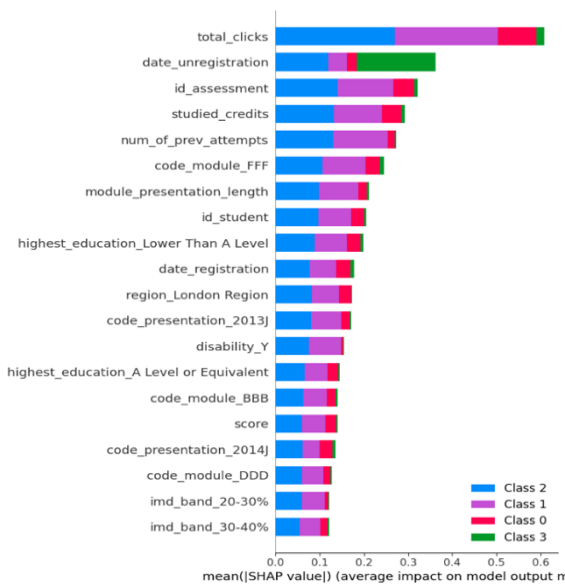
Фигура.5.15. Объркващи матрици за всички модели

Прилагане на метода SHAP

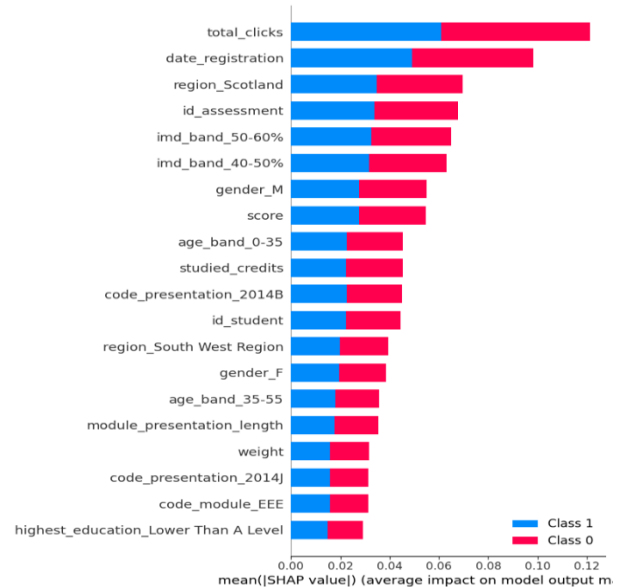
SHAP е техника, която се корени в теорията на игрите и се основава на ценностите на Shapley [6], която има за цел да обясни резултатите, генерирани от моделите за прогнозно машинно обучение. Целта му е да даде обяснение за една прогноза, като определи приноса на всяка характеристика към тази прогноза. Формално стойностите на Shapley се изчисляват, както следва:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]$$

където \overline{F} е наборът от всички функции, $f_{S \cup \{i\}}$ е моделът, обучен с функциите на под-набора и $f_S(x_S)$ е моделът, обучен без функцията. По същество формулата изчислява разликата в прогнозите чрез сравняване на резултатите от обучението на модел със и без конкретна характеристика, като се има предвид всяка възможна комбинация от характеристики. Методът SHAP се използва, за да осигури интерпретация на прогнозите на модела, като предлага анализ на значението и въздействието на различните характеристики върху прогнозираните резултати, както е визуализирано на фигури 5.16 и 5.17. Тази интерпретативност е от ключово значение за разбирането на движещите сили на представянето на студентите.

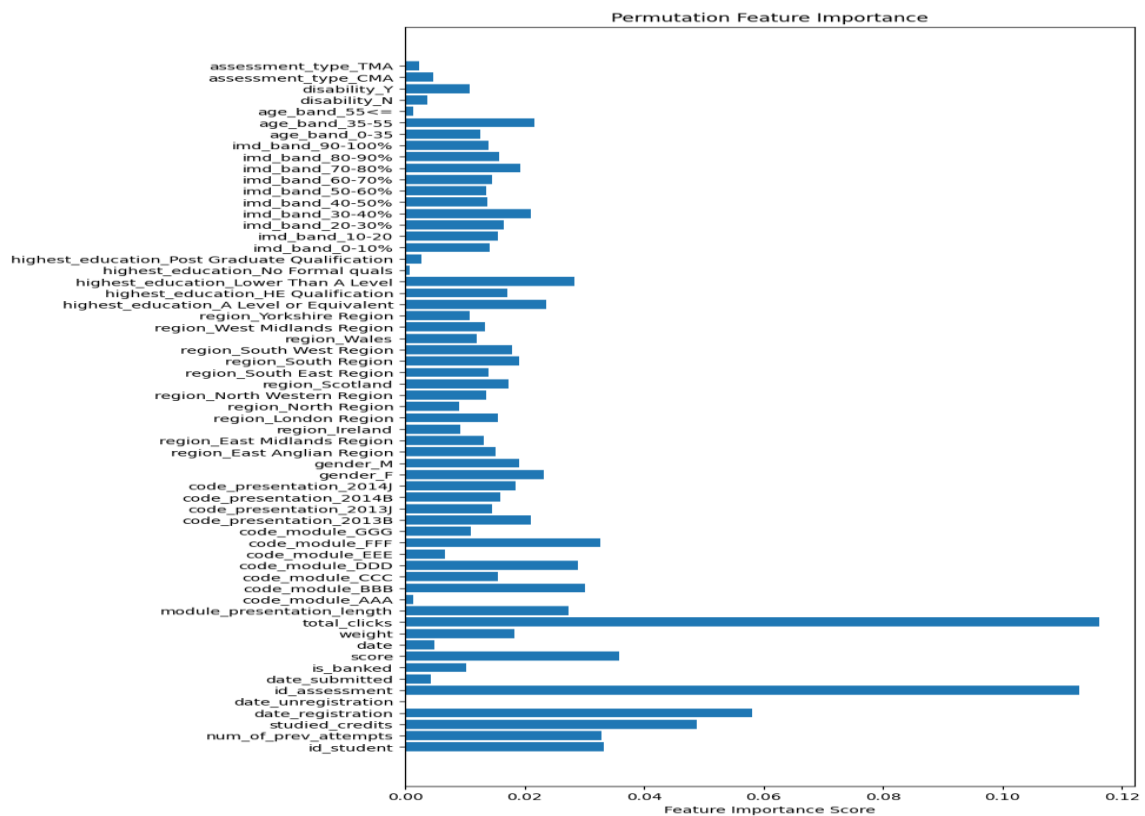


Фиг. 5.16 Обобщен график на SHAP за прогнозиране на успеха на курса



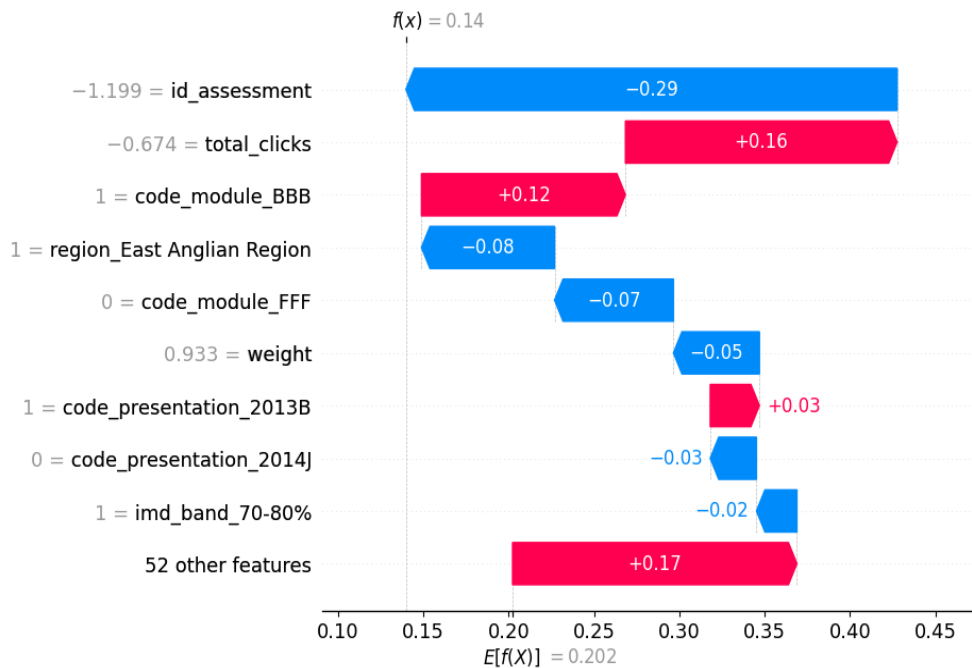
Фиг. 5.17. Обобщена графика на SHAP за ранно прогнозиране на цялостния успех на изследването

Анализът на важността на характеристиките на пермутацията и обобщените графики на SHAP допълнително навлизат в дълбочина в ключовите характеристики, влияещи върху прогнозите на моделите, като идентифицират „Общите кликания“ и „Дата на отписване“ като значими прогнози за успеха на студентите, както е показано на фигури 5.18 до 5.20.

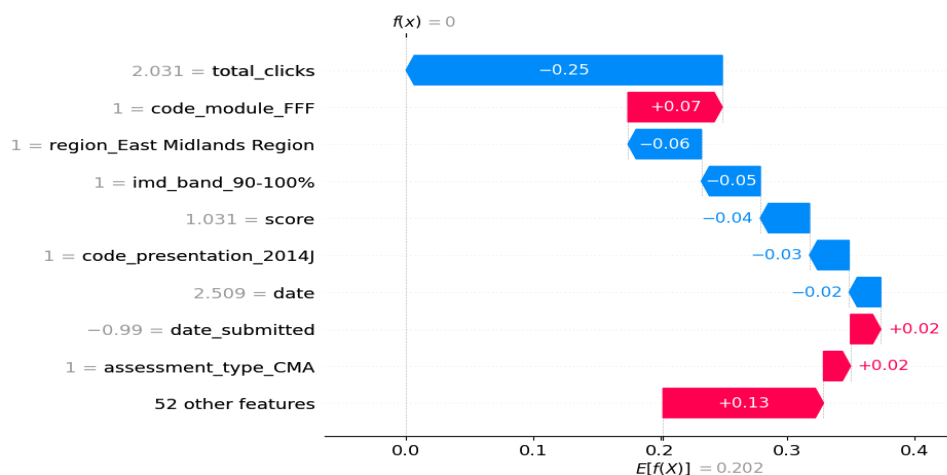


Фиг.5.18. Важност на характеристиката за пермутация

Следните каскадни диаграми представят стойностите на SHAP за два различни случая в набора от данни, илюстриращи приноса на всяка функция за прогнозирането на модела. В първата диаграма функцията „id_valuation“ значително намалява продукцията на модела, докато „total_clicks“ също я намалява, но в по-малка степен.



Фиг.5.19. SHAP стойности за първи случай



Фиг.5.20. SHAP стойности за втори случай

Дискусии

Изследването подчертава значението на показателите за ангажираност и навременната регистрация като основни предиктори на академичния успех, като се застъпва за персонализирани интервенции за повишаване на ангажираността и успеха на студентите. То подчертава стойността на обяснимия ИИ в образователните анализи, което дава възможност за практическа информация за целенасочени стратегии за подкрепа.

В заключение, това всеобхватно изследване обогатява областта на науката за образователни данни, като предоставя задълбочени познания за факторите, влияещи върху успеха на студентите в онлайн учебна среда. Чрез прилагането на техники с обясним ИИ, то предлага ценни насоки за преподавателите, администраторите и създателите на политики, целящи да използват стратегии, основани на данни, за подобряване на образователните резултати и изживяването на студентите в контекста на онлайн обучението. Внимателният подход на изследването, от предварителната обработка на данните до интерпретацията на модела, определя еталон за бъдещи изследвания в тази област, като гарантира, че прогнозните анализи в образованието са ефективни и разбираеми.

ГЛАВА 7: ИЗПОЛЗВАНЕ НА ОБРАБОТКАТА НА ЕСТЕСТВЕН ЕЗИК ЗА ОСИГУРЯВАНЕ НА КАЧЕСТВОТО ВЪВ ВИСШЕТО ОБРАЗОВАНИЕ

Темата изследва интегрирането на осигуряването на качество (QA) във висшето образование с напредъка в областта на изкуствения интелект (ИИ) и обработката на естествен език (NLP), за да се гарантира привеждането на учебните планове в съответствие с учебните програми, като по този начин се повиши качеството на образованието. Тя подчертава значението на добре структурираните резултати от обучението, като използва таксономията на Блум за определяне на образователните цели и въвежда нова система, базирана на обработка на естествен език, насочена към поддържане на съгласуваност между учебните планове и учебните програми.

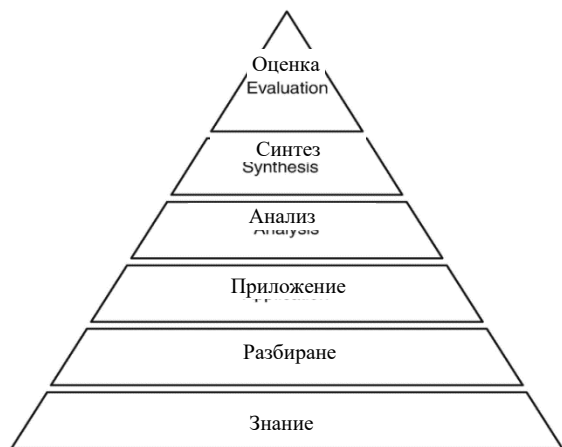
Въпроси и отговори във висшето образование, ИИ и обработка на естествен език

Осигуряването на качеството във висшето образование е от основно значение за поддържането на високи образователни стандарти и посрещането на променящите се нужди на обществото и студентската маса. Разделът изяснява значението на осигуряването на качеството за привличане на качествени студенти и персонал, повишаване на удовлетвореността от работата и поддържане на конкурентоспособността на висшите учебни заведения. Процесът от Болоня е подчертан като ключова инициатива за интегриране на принципите за осигуряване на качество в цяла Европа, осигурявайки прозрачни процеси и стабилни системи за акредитация. Дискусията се простира до присъщите и външни качества на висшето образование, като подчертава необходимостта

от привеждане на образователните предложения в съответствие с обществените очаквания. Появата на технологиите за изкуствен интелект и обработка на естествен език предлага обещаващи възможности за увеличаване на процесите за осигуряване на качеството, особено чрез анализ и подобряване на учебните програми и резултатите от обучението.

Резултати от обучението и таксономията на Блум

Традиционният подход, ориентиран към преподавателя, към изграждането на учебната програма е в контраст със съвременния модел, ориентиран към студентите, който се фокусира върху постиженията, очаквани от тях до края на курса. Таксономията на Блум е представена като решаваща рамка за формулиране на резултатите от обучението, категоризиране на когнитивните процеси в йерархични нива и осигуряване на структуриран подход към образователните цели. Подчертава се ползата от таксономията за създаване на ясни и измерими резултати от обучението, като се подчертава нейното значение в контекста на процесите за осигуряване на качеството, подобрени чрез ИИ, в образованието.



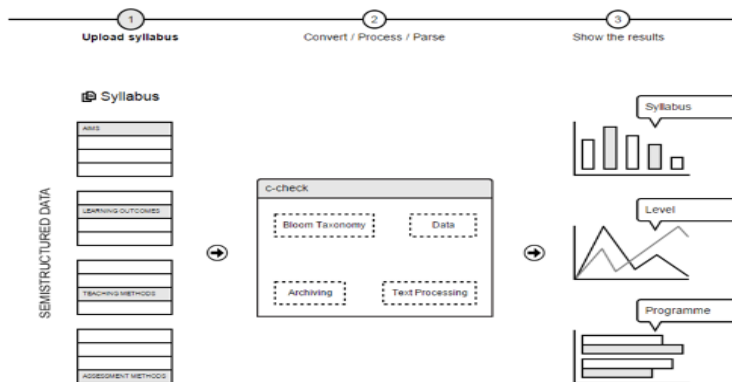
Фиг.6.1. Когнитивни нива според таксономията на Блум

Definitions	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Bloom's Definition	Remember previously learned information.	Demonstrate an understanding of the facts.	Apply knowledge to actual situations.	Break down objects or ideas into simpler parts and find evidence to support generalizations.	Compile component ideas into a new whole or propose alternative solutions.	Make and defend judgments based on internal evidence or external criteria.
Verbs	<ul style="list-style-type: none"> Arrange Define Describe Duplicate Identify Label List Match Memorize Name Order Outline Recognize Relate Recall Repeat Reproduce Select State 	<ul style="list-style-type: none"> Classify Convert Defend Describe Discuss Distinguish Estimate Explain Express Extend Generalized Give examples) Identify Indicate Infer Locate Paraphrase Predict Recognize Rewrite Review Select Summarize Translate 	<ul style="list-style-type: none"> Apply Change Choose Compute Demonstrate Discover Dramatize Employ Illustrate Interpret Manipulate Modify Operate Practice Predict Prepare Produce Relate Schedule Show Sketch Solve Use Write 	<ul style="list-style-type: none"> Analyze Appraise Breakdown Calculate Categorize Compare Contrast Criticize Create Differentiate Discriminate Distiguish Examine Experiment Identify Illustrate Infer Model Outline Point out Question Relate Select Separate Subdivide Test 	<ul style="list-style-type: none"> Arrange Assemble Categorize Collect Combine Comply Compose Construct Create Design Develop Devise Explain Formulate Generate Plan Prepare Rearrange Reconstruct Relate Reorganize Revise Rewrite Set up Summarize Synthesize Tell Write 	<ul style="list-style-type: none"> Appraise Argue Assess Attach Choose Compare Conclude Contrast Defend Describe Discriminate Estimate Evaluate Explain Judge Justify Rate Relate Select Predict Support Value

Фиг. 6.2. Таксономия на Блум

Предложение за система

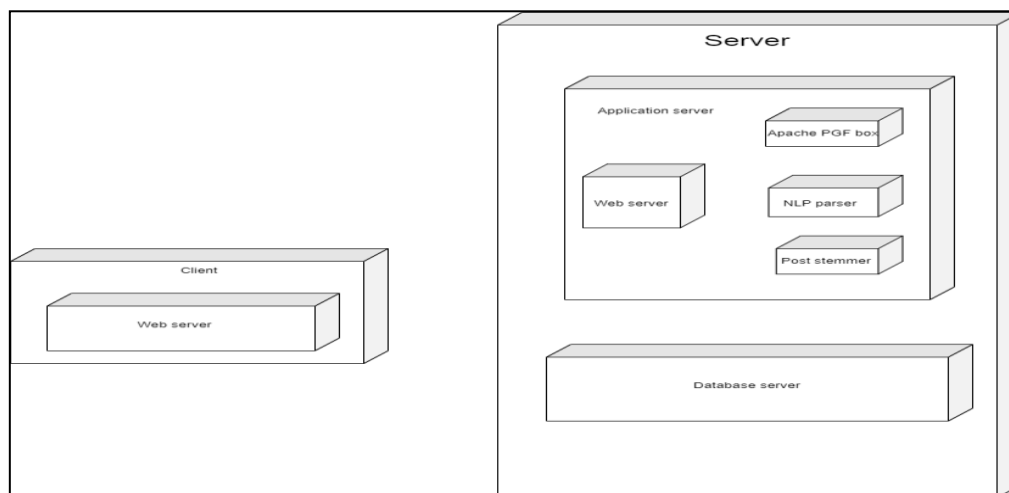
Разделът въвежда нова система, използваща обработка на естествен език, за да се преодолее пропастта между съдържанието на учебните програми и резултатите от учебните програми. Чрез обработка на учебните програми чрез техники с обработка на естествен език, системата извлича и оценява резултатите от обучението спрямо таксономията на Блум, осигурявайки образователно уравниване и последователност. Общ преглед на работния процес на системата, от приноса на учебните програми до оценката на резултатите от обучението, е илюстриран на фигура 6.6.4, показвайки ефективността на процеса и потенциала за рационализиране на осигуряването на качеството във висшето образование.



Фиг. 6.3. Преглед на предложената система

Системен анализ и архитектура

Предложената система е разработена с акцент върху удобството и ефективността на използване, като отговаря на нуждите на служителите по осигуряване на качеството и на преподавателите. Тя подчертава използваемостта, надеждността и функциите за поддръжка на системата, като има за цел да осигури интуитивен и ефективен инструмент за осигуряване на качеството. Архитектурата на системата, подробно описана на фигура 6.6.7, включва сложни компоненти като Stanford NLP Parser и алгоритъма Porter Stemmer за задълбочен анализ на текста на учебната програма, осигуряващ точно извличане и оценка на резултатите от обучението.



Фиг.6.6. Архитектура на системата

Заклучение

Предложената система, базирана на обработка на естествен език, бележи значителен напредък в областта на осигуряване на качеството за висшето образование. Тя обещава не само да рационализира процеса на осигуряване на качеството, като осигури уравниване на учебните планове по учебни програми, но и отваря пътища за по-нататъшно развитие и персонализиране, за да отговори на специфичните институционални нужди. Потенциалът на системата за повишаване на качеството на образованието я прави безценен актив за висшите учебни заведения, които се стремят да подобрят своите академични предложения и да се адаптират към променящия се образователен пейзаж.

ЗАКЛЮЧЕНИЕ

Този дисертационен труд излага целта да се използва ИИ за подобряване на способностите за прогнозиране във висшето образование, като се съсредоточи върху записването на студенти, отпадането, успеха и ранното идентифициране на изложените на риск студенти. Развитието на изследването доведе до значителни констатации и представи нова система за оценка на качеството на образованието. Основните констатации и заключения включват:

- **Прогностични модели за записване и успех:** Успешно внедрени усъвършенствани техники за машинно обучение и дълбоко обучение за прогнозиране на записването на студенти, отпадането на студенти и успеха. Тези модели предоставят надеждни анализи, които могат да послужат за вземане на решения.
- **Идентифициране на студенти, изложени на риск:** Използвайки прогностичната сила на ИИ, дисертационният труд представи метод за ранно идентифициране на студенти, които може да са изложени на риск от отпадане. Това е важен механизъм за подобряване на системите за подкрепа в рамките на образователната среда и подобряване на процента на задържане на студентите.

- **Интеграция на обясним ИИ:** Добавяйки ценностите на SHAP в моделите на ИИ, дисертационният труд предприе важна стъпка за това да направи вземането на решения за ИИ във висшето образование по-прозрачно и интерпретируемо.
- **Система с ИИ за осигуряване на качеството във висшето образование:** Друг принос на този дисертационен труд е предложената система за осигуряване на качеството на учебни планове и учебни програми с използване на обработка на естествен език. Тази система обработва учебни планове за извличане на резултати от обучението, които след това се оценяват с помощта на глаголи от таксономията на Блум. Получените графики осигуряват визуално разпределение на резултатите от обучението, предлагайки аналитична перспектива за уравниване на образователното съдържание с общите цели на курса и програмата.
- **Принос към ИИ във висшето образование:** Този труд допринася за дискусиите относно ролята на ИИ във висшето образование, като показва как техниките и технологиите на ИИ могат да бъдат приложени на практика за подобряване на няколко процеса. Той служи за преодоляване на пропастта между потенциала на ИИ и неговите реални ползи във висшето образование.

ПРИНОСИ

Основните приноси на дисертационния труд могат да бъдат обобщени, както следва:

Научни приноси:

- Разработени са прогностични модели на ИИ за прогнозиране на записването на студенти, отпадането и успеха им, допълващи развитието на областта на анализа на образователни данни и допринасящи за теоретичното разбиране на прогнозния анализ в контекста на висшето образование.
- В моделите на ИИ е интегриран обясним ИИ, чрез въвеждането на стойностите на SHAP, допринасяйки за научния дискурс за тълкуването на модела и прозрачността на ИИ в образователните среди.

Научно-приложни приноси:

- Предложен е подход за прогнозиране на отпадането на студентите, използващ логистична регресия въз основа на демографски данни, академично представяне и данни за записване.
- Изследвани са техники за машинно обучение за прогнозиране на приема в университети, чрез анализиране на резултатите от приемните изпити, матурите и оценките от средното образование.
- Предложен е прогностичен анализ за идентифициране на ученици в риск, комбинирайки научни изследвания с практическо приложение, в подкрепа на средите за образование в реалния свят и повишаване на процента на задържане на студентите.
- Въведена е система за осигуряване на качеството на учебната програма, използваща обработка на естествен език за анализ на учебните програми, извличане резултатите от обучението и оценката им, с помощта на Bloom's таксономия на глаголите.

Приложни приноси:

- Анализирани са алгоритми за ИИ за прогнозиране на представянето на студентите във висшето образование, допринасящи за разбирането на въздействието на ИИ върху академичния успех.
Демонстрирано е практическото приложение на технологиите за ИИ за повишаване на качеството на обучение във висшето образование, които могат да бъдат приложени, за да се извлекат ползите от технологията с ИИ.

СПИСЪК НА ПУБЛИКАЦИИТЕ

- [1] **Berat Ujkani**, Daniela Minkovska, Ognyan Nakov, "*Understanding Student Success Prediction Using SHapley Additive exPlanations*", Международна научна конференция „Компютърни науки“ (COMSCI), Созопол, България, 2023, стр. 1–4, doi: 10.1109/COMSCI59259.2023.10315938.
- [2] **Berat Ujkani**, "*ChatGPT for Students eBook*", самостоятелна публикация, 2023, URL: https://www.researchgate.net/publication/368721139_ChatGPT_for_Students_eBook.
- [3] **Berat Ujkani**, Betim Drenica, Rinor Kurteshi, Muzafer Shala, "*Development and implementation of speech-to-text technology using AI in Albanian language*", Enlargement and Integration Workshop "Digital Transformation, Data and AI in the Western Balkans", Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53599-7, doi:10.2760/863985, JRC129903.
- [4] **Berat Ujkani**, Daniela Minkovska, Lyudmila Stoyanova, "*Application of Logistic Regression Technique for Predicting Student Dropout*", 2022 г. XXXI Международна научна конференция „Електроника“ (ET), Созопол, България, 2022 г., стр. 1–4, doi: 10.1109/ET55967.2022.9920280.
- [5] **Berat Ujkani**, Daniela Minkovska, Lyudmila Stoyanova, "*A Machine Learning Approach for Predicting Student Enrollment in the University*," XXX Международна научна конференция по електроника (ET), Созопол, България, 2021 г., стр. 1–4, doi: 10.1109/ET52713.2021.9579795.
- [6] **Berat Ujkani**, Daniela Minkovska, Lyudmila Stoyanova, "*Using Natural Language Processing for Quality Assurance Purposes in Higher Education*," IV Международна конференция по високи технологии за устойчиво развитие (HiTech), София, България, 2021 г., стр. 01–04, doi: 10.1109/HiTech53072.2021.9614206.
- [7] **Berat Ujkani**, Daniela Minkovska, Lyudmila Stoyanova, "*A recommender system for WordPress themes using item-based collaborative filtering technique*," XXIX Международна научна конференция по електроника (ET), Созопол, България, 2020 г., стр. 1–3, doi: 10.1109/ET50336.2020.9238305.

SUMMARY

In the realm of higher education, which is rapidly being reshaped by digital transformation and an increased reliance on data-driven decision-making, the thesis in question addresses the pressing need for leveraging Artificial Intelligence to predict student performance. Recognizing the challenges faced by educational institutions, such as enhancing student engagement, optimizing enrollment processes, and reducing dropout rates, the thesis proposes the utilization of machine learning and deep learning algorithms to provide precise and understandable predictive insights. This is particularly relevant in the context of a global shift towards online and blended learning modalities.

The thesis sets forth several objectives aimed at harnessing AI's potential in higher education. Among these are the development of models for predicting student dropout and enrollment, the identification of at-risk students, and the implementation of explainable AI to ensure the transparency of predictive models. Additionally, it proposes an NLP-based system to facilitate quality assurance processes, thereby underscoring the multifaceted application of AI in enhancing educational administration and student success.

In terms of scientific novelty, the thesis introduces advanced AI techniques to predict student dropout, enrollment, and success, marking a significant contribution to the field. It emphasizes the application of the SHAP method for enhancing AI decision interpretability in educational settings and proposes an NLP-based system for quality assurance, bridging the gap between theoretical AI advancements and practical applications in education.

The practical applicability of the thesis is demonstrated through the development of predictive models that enable educational institutions to make data-driven decisions to improve student engagement and success. Furthermore, the thesis highlights the successful implementation of its findings in the learning process at the University of Mitrovica, Kosovo, showcasing the real-world impact of the research.

The content of the dissertation is meticulously structured to cover a broad spectrum of topics relevant to AI's application in higher education. It begins with an overview of AI and its transformative role in education, followed by a bibliometric analysis of AI-driven performance prediction. Subsequent topics delve into predictive modeling for student dropout and enrollment, the utilization of explainable AI for course success prediction and early identification of at-risk students, and the use of NLP for quality assurance in educational content.

In conclusion, the thesis makes substantial contributions to the field of AI in higher education, offering innovative solutions for predicting student performance and enhancing quality assurance processes. Its practical applications, scientific novelty, and comprehensive analysis of AI's role in education make it a significant addition to academic research and educational practice alike. The findings not only advance the theoretical understanding of AI in education but also provide actionable insights for institutions aiming to leverage AI for improved educational outcomes.