

ПРОГНОЗИРАНЕ НА УСПЕВАЕМОСТТА НА СТУДЕНТИТЕ: СЪВРЕМЕННИ МОДЕЛИ И ТЕХНИКИ

Доц. д-р Силвия Гафтанджиева,
проф. д-р Росица Донева

Пловдивски университет „Паисий Хилендарски“

Резюме. Успеваемостта на студентите е важен показател за измерване на качеството на образователния процес и допринася за издигане на висшето училище в рейтингови класации. Прогнозирането на успеваемостта на студентите позволява ранна превенция от отпадане на обучаеми от системата на висшето образование, поради което в последните години се наблюдава засилен интерес на изследователите към разработване на модели за прогнозиране на успеха на студентите. В статията са анализирани ползите от прогнозиране на успеваемостта за различни групи заинтересовани лица (студенти, преподаватели, лица на ръководни позиции), които мотивират нарастващото внимание към подобни модели. В сравнителен план са представени редица от вече разработените модели за прогнозиране на успеха на студентите, базирани на алгоритми за машинно обучение (Machine Learning – ML) и за „обясним изкуствен интелект“ (eXplainable Artificial Intelligence – XAI).

Ключови думи: успеваемост на студенти; прогнозиране; изкуствен интелект; Machine Learning; eXplainable Artificial Intelligence

1. Въведение

Високата успеваемост на студентите, навременното дипломиране и реализацията на пазара на труда са важни показатели за измерване на качеството на образователния процес и допринасят за висок ранг на университета в рейтинговите класации, който често е определящ фактор при избор на висше училище от бъдещи студенти. От друга страна ВУ получават финансиране според броя на обучаваните студенти. Ръководствата на ВУ търсят ефективни начини да идентифицират студенти с ниски академични постижения в ранни етапи на обучение (Jović et al. 2017), да насърчат студентите да повишат своя успех, и да намалят броя на прекъсналите студенти (Gaftandzhieva et al. 2022), в резултат на което разработват стратегии за повишаване на успеваемостта на

студентите (Ahmed 2024). За да могат да повлияят последователно на учебния процес и да поддържат студентите мотивирани, ангажирани и успешни, ръководните органи и преподавателите във ВУ се нуждаят от подкрепа при вземане на решения (Tiukhova et al. 2024).

Внедряването на софтуерни решения за прогнозиране на успеваемостта на студенти се благоприятства от процесите на дигитална трансформация на ВУ (Gaftandzhieva et al. 2022), в резултат на които се натрупват големи масиви от данни. Ръководствата на ВУ осъзнават потенциала на събраните данни за подобряване на качеството на обучението, задържането на студентите и тяхната успеваемост, което води до нарастване на интереса към решения, които извличат данни от използваните софтуерните системи, предвиждат оценките от изпити и позволяват на ръководството / академичния персонал да взема решения, базирани на данни, и по този начин да подобри качеството на предлаганите услуги и да насърчи студентите да постигат по-високи резултати. Разработените модели за прогнозирането на успеваемостта намират приложение при изграждане на системи за ранно предупреждение за студенти, които са изложени на риск от отпадане, и персонализирани системи за генериране на препоръки за подобряване на учебното изживяване на студентите (Alamri & Alharbi 2021). Подобни изследвания, базирани на данни, са много важни по отношение на установяването на рамка за анализ на обучението във висшето образование и допринасят за процесите на вземане на решения (Yağcı 2022).

В статията са анализирани ползите от прогнозиране на успеваемостта за различни групи заинтересовани лица (студенти, преподаватели, лица на ръководни позиции), които мотивират нарастващото внимание към подобни модели. В сравнителен план са представени редица от вече разработените модели за прогнозиране на успеха на студентите, базирани на алгоритми за машинно обучение (Machine Learning – ML) и за „обясним изкуствен интелект“ (eXplainable Artificial Intelligence – XAI).

2. Ползи за различни групи заинтересовани лица от прогнозиране на успеваемостта на студентите

Прогнозирането на успеха на студентите е от полза за различни групи заинтересовани лица, сред които ръководни органи във ВУ, преподаватели и студенти.

Поради факта, че високите академични резултати подобряват класирането на университета в класации и увеличават възможностите за успешна професионална реализация на студентите, силно заинтересована група лица са ръководните органи във ВУ. Прогнозирането на успеваемостта на студентите позволява на ръководните органи да получат ценна информация за потенциалните резултати на студентите (Al Shibli et al. 2022) и да вземат решения за приемане на мерки за подобряване на академичните постижения на студен-

тите (Gaftandzhieva et al. 2022), да разработят ефективни планове и стратегии за подкрепа на студентите (Ahmed 2024; Ofori et al. 2020; Bujiang et al. 2021) и оптимално разпределяне на ресурсите (Ahmed 2024), които да доведат до увеличаване на процента дипломирани студенти.

Ранното прогнозиране на успеваемостта на студентите помага на преподавателите да идентифицират студенти с ниски постижения преди финалния изпит на базата на текущи резултати и дейности (Ofori et al. 2020; Paddalwar et al. 2022; Masangu et al. 2021; Gaftandzhieva et al. 2022; Al Shibli et al. 2022; Ujkani et al. 2024). Прогнозите, базирани на данни от средата за е-обучение, показват на преподавателите кои видове учебни ресурси и дейности оказват влияние върху крайната оценка (Gaftandzhieva et al. 2022). Такива прогнози дават навременни прозрения и насърчават преподавателите да предприемат мерки за намаляване на процента на студентите със слаби резултати и подобряване на техните постижения. Подобни мерки могат да бъдат предоставяне на допълнителни ресурси и подкрепа на студенти, които не постигат добри резултати, както и актуализиране на учебните материали.

Прогнозирането на оценките е полезно и за студентите, тъй като предоставя значима обратна връзка (Ofori et al. 2020), която им помага да направят планове за постигане на своите цели (Gaftandzhieva et al. 2022).

Разбирайки факторите, които влияят върху представянето на студентите, ВУ могат да разработят стратегии за подобряване на резултатите от обучението. Определянето на тези фактори позволява на преподавателите да предприемат мерки, които да доведат до подобряване на академичните постижения на студентите и увеличаване на ефективността на обучението. Като важни фактори за повишаване на успеваемостта се открояват институционалната подкрепа, надзорните практики, уменията за самоуправление, присъствието и пълноценното участие на студентите в часовете (Nguyen & Nguyen 2023), социално-икономически и демографски фактори.

3. Модели за прогнозиране на успеваемостта, базирани на машинно обучение (ML)

Алгоритмите за машинно обучение (като Logistic regression – LR, Support Vector Machines – SVM, Random Forest – RF, Perceptron classification – PC, Linear Discriminant Analysis – LDA, Gradient boosting – GB, Naive Bayes – NB, Neural network – NN, Decision Trees – DT, K-nearest neighbor – kNN, Stacking Regressor – SR, XGBoost, LightGBM) се открояват като мощна техника за прогнозиране на резултатите на студентите, която показва как различните фактори взаимодействат, за да повлияят на резултатите от обучението. Важно е да се отбележи, че за да се постигнат висока надеждност и точност на прогнозите, прогнозирането трябва да се извърши върху достатъчно голям набор от данни, да се използват различни алгоритми за машинно обучение и да се изследват множе-

ство параметри. Ефикасността на използваните алгоритми се оценява чрез показатели за оценка на производителността, като точност, прецизност, припомняне, F-мярка, средна абсолютна грешка (MAE), средна квадратична грешка (RMSE), R-квадрат, средна квадратична логаритмична грешка (MSLE) и среден абсолютен процент (MAP), а за валидиране се прилагат техники за кръстосано валидиране. Често използван подход за преодоляване на небалансиран набор от данни с множество класификации и избягване на пренастройване на моделите е прилагането на техники като SMOTE и FS (Bujang et al. 2021; Gaftandzhieva et al. 2022; Arévalo Cordovilla & Peña Carrera 2024).

Jović, Kisić, Milić, Domazet и Chandra (Jović et al. 2017) предлагат модел за прогнозиране на представянето на студентите на база на средния успех, оценки от тестове, домашни, проекти и участие в клас, посещаемост на часовете, брой неуспешни опити за полагане на финалния изпит и получена окончателна оценка. Анализираният набор от данни съдържа данни за 1696 студенти от 4 курса, извлечени от системите за е-обучение и управление на образованието. Поради постигнатата точност от 88.5% за обучение на модела е избран класификаторът SVM. Други изследвани алгоритми дават по-ниска точност – LR (87,6%), LDA (84,6%), kNN (86,5%), DT (84,4%) и NB (85,7%). Резултатите показват, че оценките от изпълнението на проекти оказват най-голямо влияние върху крайната оценка.

Masangu, Jadhav и Ajoodha (Masangu et al. 2021) изследват кои фактори влияят върху представянето на студентите и ефективността на алгоритми за машинно обучение (SVM, DT, PC, LR, RF) за прогнозиране на резултатите на студентите, и предлагат модел, който да помогне преподавателите да идентифицират студенти със слаби оценки на ранен етап. Прогнозирането се извършва на базата на данни за ангажираността на студентите в платформата за електронно обучение (честота на преглед на съобщения, брой прегледи на форуми и ресурси), текущи оценки и демографска информация. Като недостатък на изследването могат да се посочат фактите, че изследваният набор от данни се състои от 480 записа и не е изследвано влиянието на предприети действия и психологически фактори. Резултатите показват, че най-висока точност за обработвания набор данни е постигната от SVM – 70.8%. Получените резултати показват, че отсъствията на студентите влияят на техните резултати, докато получените текущи оценки не оказват влияние върху крайните резултати.

Bujang, Selamat, Krejcar (Bujang et al. 2021) предлагат модел за прогнозиране на финалната оценка на 489 студенти от университета Malaysian Polytechnics въз основа на исторически данни за академичните постижения, използващ алгоритмите за машинно обучение – DT, RF, SVM и LR. Констатациите от проведените тестове показват, че с постигнатата най-висока точност (99,6%) DT може да помогне да бъдат идентифицирани студенти, които е вероятно да от-

паднат на ранен етап, и предприемат мерки за подобряване на представянето на студентите. Моделът може да се използва както като система за ранно предупреждение за идентифициране на слаби студенти от преподавателите, така и за вземане на стратегически решения от ръководството за подобряване на представянето на студентите. В друго свое изследване Bujang (Bujang et al. 2021) изследват ефективността на алгоритмите DT, SVM, NB, kNN и LR за прогнозиране на оценки на студенти в курсовете през семестър на базата на средния успех на студентите. Моделите са тествани с набор от данни за 1282 студенти. Резултатите показват че всички прогнозни модели постигат по-висока производителност след прилагане на SMOTE към небалансиран набор от данни, като най-висока точност е получена с kNN (99,3%), DT (99,2%) и NB (98,3%). В допълнение, изследователите предлагат мултикласификационен модел за прогнозиране за небалансиран набор от данни, който използва два типа решения за подобряване точността на прогнозиране – SMOTE и F.S., който се интегрира с kNN и подобрява точността на прогнозата (99,6%). Като ограничение на проведените изследвания могат да бъдат посочени малкият набор от данни (за една и две дисциплини) и изборът на изследвани характеристики от небалансиран набор от данни, който може да повлияе на точните резултати от прогнозирането.

Paramita и Tjahjono (Paramita & Tjahjono 2021) предлагат модел за категоризиране на представянето на студенти на базата на демографски данни, ангажираност на студенти в средата за е-обучение и оценки по дисциплини. Изследваният набор от данни съдържа 32 593 записа за студенти от The Open University. Експерименталните резултати показват, че kNN алгоритъмът се представя добре в ситуации, включващи ангажираност, оценки и комбинация от двата фактора със степен на точност съответно от 0,9979, 0,9907 и 0,9918. В ситуации, които включват демографски данни, демографски данни и оценки, ангажираност и оценки и трите изследвани фактора, ANN превъзхожда kNN със степен на точност съответно 0,7902, 0,9952, 0,9834 и 0,9956. Други изследвани алгоритми са NB, RF и SVM. Тези резултати показват, че за точно класифициране на представянето в различни сценарии може да е необходимо да се комбинират няколко алгоритъма.

Други изследователи прилагат статистически методи (Хи-квадрат тест) и алгоритми за машинно обучение (DT, SVM, NB, kNN, LR и RF) за прогнозиране на финалните оценки на студентите по учебна дисциплина на базата на техните дейности в системата за е-обучение Moodle и присъствието на онлайн лекции (Gaftandzhieva et al. 2022). Наборът от данни се състои от финалните оценки на 105 студенти от ПУ „Паисий Хилендарски“, 7075 записа за тяхната активност в онлайн курса и 738 записа за присъствие на лекции. Направените прогнози са базирани на 46 атрибута. Резултатите показват корелации между крайните оценки на студентите и активността им в онлайн курса и между крайните оцен-

ки на студентите и присъствието на лекции. Всички изследвани алгоритми се представят умерено добре при прогнозиране на финалната оценка на студентите, като RF постига най-висока точност – 78%. За намаляване на грешките, причинени от небалансирана мултикласификация, базирана на SMOTE, е предложен мултикласификационен модел за прогнозиране, който подобрява точността на RF до 99,5%. Прогнозираните от модела резултати позволяват на преподавателите да предприемат мерки за намаляване броя на студентите със слаби оценки и да идентифицират кои видове учебни ресурси и дейности са по-добри предиктори за академичното представяне на студентите.

Yağcı (Yağcı 2022) предлага модел за прогнозиране на оценките от финалните изпити на студенти и ранно прогнозиране на студенти с висок риск от отпадане на базата на оценки от междинен изпит и данни от факултета. Използваният набор от данни включва оценки на 1854 студенти. Резултатите показват, че най-висока точност дава RF (74,6%), следван от NN (74,6%), SVM (73,5%), LR (71,7%), NB (71,3%) и kNN (69,9%). Като ограничение на изследването можем да посочим малкия брой изследвани характеристики.

Nguyen & Nguyen (Nguyen & Nguyen 2023) прилагат алгоритми за машинно обучение за прогнозиране на резултатите при дипломиране на 7837 студенти от университета Can Tho на базата на демографски и академични данни. Констатациите от изследването показват, че RF дава най-надеждни прогнози (98.2%), следван от DT (96.8%), GB (94.6%), kNN (84.4%) и NN (80.2%). Факторите, които оказват значително влияние върху оценките при завършване са среден успех, адрес на живеене, колеж, специалност и пол. Несъответствието между нивата на точност на използваните два метода за валидация е незначително, което показва, че предсказуемите модели, генерирани от тези алгоритми, са много надеждни.

Badal & Sungkur (Badal & Sungkur 2023) предлагат модел за прогнозиране на крайната оценка и нивото на ангажираност на студентите в средата за е-обучение на базата на данни за профила на обучаемия и активност в системата. Моделът е тестван върху набор от данни за 1074 студенти. Експерименталните резултати показват, че RF превъзхожда останалите алгоритми (LR, kNN, NB, DT, SVM и Deep Learning) – точност от 85% и 83% за прогнозиране на оценката и ангажираността. Изследователите разработват уеб приложение, което изпълнява алгоритми за машинно обучение и позволява на преподавателите да предвиждат представянето и ангажираността, анализирайки данни от качен файл, да идентифицират обучаемите, които са изложени на риск, и предприемат навременни коригиращи мерки. Като ограничение на това изследване можем да посочим факта, че не е оценено въздействието на външни фактори при участие на студентите в дискусии, напр. производителност на устройството.

Е. Ahmed изследва как интегрирането на алгоритми за машинно обучение в традиционните методи на преподаване може да подобри резултатите

от обучението (Ahmed 2024). Използваният за изследването набор от данни съдържа записи за 32 582 студенти (пол, регион, входящи резултати, брой предходни опити за явяване, натрупани кредити, инвалидност, краен резултат). Сравнението на получените резултати показва, че алгоритъмът SVM дава най-точни резултати (96%), следван от DT (93.4%), kNN и NB (83.3%). За да осигури стабилни и надеждни прогнозни резултати, Ahmed използва строги методологии, като многократно k-кратно кръстосано валидиране и оптимизация на хиперпараметър. Предложеният модел се откроява с иновативната си техника за групиране, цялостен сравнителен анализ и практическо приложение при прогнозиране на резултатите. Получените резултати позволяват на ВУ да използват иновативни технологии, за да се справят с предизвикателствата и да вземат информирани решения за предприемане мерки за подобряване на резултатите, персонализирани интервенции, оптимално разпределяне на ресурсите и др.

В резултат на проведено изследване за подобряване на механизмите за подкрепа на студентите Cordovilla и Carrera (Cordovilla & Carrera 2024) предлагат метод за ранно прогнозиране на резултатите на студентите на база на данни от системата за е-обучение Moodle и външни фактори. Изследваният набор от данни включва данни за 591 студенти от Universidad Estatal de Milagro (UNEMI), които са предварително обработени със SMOTE. Експерименталните резултати показват, че невронната мрежа MLP постига най-висок AUC-ROC (0.9991), следвана от RF (0.988), SVM (0.983) и LR (0.9584). Изследователите стигат до заключение, че LR е практичен и интерпретируем инструмент за ВУ, които целят да внедрят системи за ранно предупреждение. Опростеността и ефективността на модела улесняват навременното идентифициране на учениците в риск, позволявайки прилагането на персонализирани интервенции за подобряване на резултатите и степента на задържане на студенти. Според получените резултати значими фактори за успеха на студентите са резултатите от ранното оценяване и ангажираността в средата за е-обучение, измервана с различни показатели (като брой прегледани ресурси и изпратени задания). Въпреки че останалите модели също показват силна прогнозна производителност, тяхната сложност и намалена интерпретируемост в сравнение с тези на LR може да представляват предизвикателства за практическо прилагане в образователни среди. Като ограничение на изследването могат да бъдат посочени фактите, че данните са събрани от конкретен курс, което може да е ограничило възможността за обобщаване на констатациите.

4. Модели за прогнозиране на успеваемостта, базирани на обясним изкуствен интелект (XAI)

Алгоритмите за машинно обучение, обучени върху големи набори от данни, могат с голяма точност да предскажат дали студентът е изложен на риск от

провал, но високата точност не е единственият критичен фактор за успеха на такива модели. За да се доверят на модела, заинтересованите лица трябва да разбират причините, поради които са направени прогнозите (Alsubhi et al. 2023; Alamri & Alharbi 2021). Сложният характер на тези алгоритми често замъглява процесите на вземане на решения в „черна кутия“, което затруднява разбирането на конкретните причини зад техните прогнози (Ujkani et al. 2024). През последните години eXplainable Artificial Intelligence (XAI) се утвърждава като популярен подход за тълкуване и обяснение на причините за решенията на алгоритмите за машинно обучение (Hoq et al. 2023). XAI използва техники, които подобряват прозрачността на моделите и позволяват на заинтересованите лица да разберат защо даден модел прави определени прогнози, а не само какъв е процентът на точност (Ujkani et al. 2024). Това позволява на всички групи заинтересовани лица (студенти, преподаватели и лица на ръководни позиции) да разберат факторите, които стоят зад прогнозните резултати, и да адаптират своите интервенции към конкретни проблеми. Използването на обясними модели е полезно и за създателите на системи с изкуствен интелект, тъй като им позволява да гарантират качеството на обучението модел (Alamri & Alharbi 2021). Използването на XAI за интерпретация на прогнозите за успеваемостта на студентите е сравнително нова, но бързо развиваща се област на изследване.

Група изследователи, ръководени от Jang (Jang et al. 2022), предлагат практически метод за прогнозиране на представянето на студентите. Използвайки XAI, те предлагат метод за предоставяне на информация на преподавателите, която да подпомогне тълкуването на резултатите от класификацията и визуално представяне на информация. Този метод позволява на преподавателите да интерпретират резултатите от класификацията на студентите в риск за предоставяне на персонализирано обучение. Като недостатък на изследването може да бъде посочен фактът, че експериментите са проведени с данни от два курса в една предметна област.

Hoq, Brusilovsky и Akram (Hoq et al. 2023) предлагат XAI модел за прогнозиране на финалните оценки на студентите на база на информация за преподаване на задачи и обясняване на направените от модела прогнози чрез алгоритъма SHAP. Обясненията на взетите от модела решения се представени на две нива – индивидуално за всеки студент и на ниво курс. Моделът може да помогне на преподавателите да си обяснят поведението на студентите, което позволява ефективна намеса и предоставяне на адаптивна подкрепа за студентите.

Други изследователи (Alsubhi et al. 2023) предлагат модел за категоризиране на оценките на студенти CORELS, който превъзхожда съществуващите модели BB и WB. Направените изводи предоставят важна информация за най-надеждните модели за прогнозиране на оценки в рамките на определен диапазон – напр. полезност на RF при прогнозиране на оценка C и на SVM при

прогнозиране на оценка D. Получените резултати могат да бъдат приложени за подобряване точността на системите за прогнозиране на оценки във ВУ.

Други изследователи предлагат подход за прилагане на ХАИ за изследване на стабилността на моделите за прогнозиране на успеха на студентите (Tiukhova et al. 2024), който съчетава традиционни техники за изследване на стабилността на моделите за прогнозиране на успеха на студентите във времето и ХАИ за изследване стабилността на класациите за важност на извлечените характеристики, генерирани за тези модели. В резултат се генерират нови прозрения от стабилни характеристики, което позволява на преподавателя да дава съвети за обучение. Получените резултати имат принос към изследванията в областта на образованието, увеличават интерпретируемостта и обяснимостта на алгоритмите за прогнозиране и осигуряват тяхната приложимост в променящ се контекст.

Özkurt (Özkurt 2024) предлага подход за разясняване на прогнозите на алгоритмите за машинно обучение чрез прилагане на методологиите SHAP и LIME. Постигнатите резултати са обещаващи и могат да допринесат за развитието на процеси на вземане на решения, базирани на данни, и по-ефективно планирани интервенции за подобряване успеха на студентите.

Ujkani, Minkovska и Hinov (Ujkani et al. 2024) използват техники за машинно обучение и дълбоко обучение за прогнозиране успеха на студентите и ХАИ техниката SHapley (SHAP), за да прогнозират оценките на студентите, да идентифицират студенти, изложени на риск, и ключовите фактори, които допринасят за успеха или провала. Използваният набор от данни включва данни за 32 593 студенти – информация, свързана с демографията на студентите, записвания в курсове, оценяване и действия в среда за е-обучение. За разлика от традиционните статистически методи, които изследват променливи връзки, предложеният подход използва усъвършенствани техники за дълбоко обучение за идентифициране на модели и прозрения, което позволява по-добро разбиране на факторите, влияещи върху успеха на студентите, и позволява целенасочени интервенции в подкрепа на техния успех. Резултатите показват, че ангажираността на студентите и сроковете за регистрация са критични фактори, които влияят върху тяхното представяне.

5. Заключение

На база на анализирания изследвания може да се констатира, че повечето модели се фокусират върху прогнозирането на оценките на студентите на базата на ограничен набор от характеристики, без да вземат предвид механизми за подобряване на резултатите от обучението. Малка част от изследванията анализират влиянието на активността на студентите при изучаване на учебни материали и изпълнение на дейности в среди за е-обучение. Като ограничение на всички изследвания можем да посочим факта, че моделите са тествани

върху данни от една образователна институция и анализират статичен набор с данни. Получените резултати в 7 изследвания са обнадеждаващи, като демонстрират степен на точност над 95%. RF постига силна прогнозна производителност в редица изследвания, но неговата сложност може да доведе до затруднения при практическо прилагане. Поради своята опростеност и ефективност LR е подходящ за внедряване във ВУ. По-сложните модели (като kNN, NN и SVM) превъзхождат по-простите модели в малка част от изследванията, което показва, че увеличената сложност на модела не води непременно до по-добра точност на прогнозиране. Към настоящия момент използването на ХАИ за обясняване на причините зад направените прогнози е в ранен етап на изследване и не са предложени метрики за оценка за обяснимостта на модела, които да позволяват сравняване нивата на обяснимост на различни модели.

Интересна насока за развитие на изследванията в областта е да се създадат модели, които да предсказват крайната оценка на студентите по избрана дисциплина на база на тяхната активност в среда за е-обучение и/или присъствия на лекции и упражнения, текущите оценки по дисциплината, както и модели за предсказване на успеха при дипломиране на студенти на база на оценките по дисциплините, среден успех, данни за приема във ВУ, демографски данни. Данните, върху които подобни модели ще осъществяват анализ, могат да бъдат извлечени в реално време от средите за е-обучение и/или студентските информационни системи на ВУ. Интегрирането на избрани модели за прогнозиране на успеваемостта в съответни софтуерни системи, предоставящи услуги за ръководни органи, преподаватели и студенти, биха позволили, съгласно нуждите на различните заинтересовани лица, да бъдат анализирани причините, които стоят зад направените прогнози, и на подходящ етап и ниво да бъдат предприети мерки за подобряване на успеха, за отстраняване на причините за слабо представяне, за намаляване на броя отпаднали студенти и др.

Благодарности и финансиране

Изследването е финансирано от Европейския съюз – NextGenerationEU, чрез Националния план за възстановяване и устойчивост на Република България, проект № BG-RRP-2.004-0001-C01.

Acknowledgements & funding

This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project No. BG-RRP-2.004-0001-C01.

REFERENCES

- AHMED, E., 2024. Student Performance Prediction Using Machine Learning Algorithms. *Applied Computational Intelligence and Soft Computing*, vol. 1. ISSN:1687-9732.
- AL SHIBLI, K. et al., 2022. Model for Prediction of Student Grades using Data Mining Algorithms. *European Journal of Information Technologies and Computer Science*, vol. 2, no. 2, pp. 1 – 6. ISSN: 2736-5492.
- ALAMRI, R. & ALHARBI, B., 2021. Explainable student performance prediction models: a systematic review. *IEEE Access*, vol. 9, pp. 33132 – 33143. ISSN: 2169-3536.
- ALSUBHI, B., et al., 2023. Effective Feature Prediction Models for Student Performance. *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11937 – 11944. ISSN 1792-8036.
- ARÉVALO CORDOVILLA, F. & PEÑA CARRERA, M., 2024. Comparative analysis of machine learning models for predicting student success in online programming courses: a study based on LMS data and external factors. *Mathematics*, vol. 12, no. 20. ISSN 2227-7390.
- BADAL, Y. & SUNGKUR, R., 2023. Predictive modelling and analytics of students' grades using machine learning algorithms. *Education and information technologies*, vol. 28, no. 3, pp. 3027 – 3057. ISSN 1573-7608.
- BUJANG, S., et al., 2021. A predictive analytics model for students' grade prediction by supervised machine learning. *IOP Conference Series: Materials Science and Engineering*, vol. 1051, no. 1. ISSN: 1757-899X.
- BUJANG, S., et al., 2021. Multiclass prediction model for student grade pre-diction using machine learning. *IEEE Access*, vol. 9, pp. 95608 – 95621. ISSN: 2169-3536.
- GAFTANDZHIEVA, S. et al., 2022. Exploring online activities to predict the final grade of student. *Mathematics*, vol. 10, no. 20. ISSN 2227-7390.
- HOQ, M., et al., 2023. Analysis of an Explainable Student Performance Prediction Model in an Introductory Programming Course. *Proceedings of the 16th International Conference on Educational Data Mining*, Bengaluru: International Educational Data Mining Society.
- JANG, Y., et al., 2022. Practical early prediction of students' performance using machine learning and eXplainable AI. *Education and Information Technologies*, vol. 27, no. 9, pp. 12855 – 12889. ISSN 1573-7608.
- JOVIĆ, J., et al., 2017. Prediction of student academic performance using machine learning algorithms. *13th International Conference on e-Learning*, pp. 333 – 338. ISBN 979-8-3503-8103-0.

- MASANGU, L. et al., 2021. Predicting student academic performance using data mining techniques. *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 1, pp. 153 – 163. ISSN 2415-6698.
- NGUYEN, M., & NGUYEN, H., 2023. Predicting graduation grades using Machine Learning: A case study of Can Tho University students. *CTU Journal of Innovation and Sustainable Development*, vol. 15. ISDS, pp. 83 – 92. ISSN 2588-1418.
- OFORI, F., et al., 2020. Using Machine Learning Algorithm to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review. *Journal of Information and Technology*, vol. 4, no. 1, pp. 23 – 45. ISSN 2617-3573.
- ÖZKURT, C. 2024. Assessing Student Success: The Impact of Machine Learning and XAI-BBO Approach. *Journal of Smart Systems Research*, vol. 5, no. 1, pp. 40 – 54. ISSN 2757-6787.
- PADDALWAR, S., et al., 2022. Predicting students' academic grades using machine learning algorithms with hybrid feature selection approach. *ITM Web of Conferences*, vol. 44. ISSN 2271-2097.
- PARAMITA, A. & TIAHJONO, L., 2021. Implementing machine learning techniques for predicting student performance in an E-learning environment. *International Journal of Informatics and Information Systems*, vol. 4, no. 2, pp. 149 – 156. ISSN 2579-7069.
- TIUKHOVA, E., et al., 2024. Explainable Learning Analytics: Assessing the stability of student success prediction models by means of explainable AI. *Decision Support Systems*, vol. 182, Article 114229. ISSN 0167-9236.
- UJKANI, B., et al., 2024. Course Success Prediction and Early Identification of At-Risk Students Using Explainable Artificial Intelligence. *Electronics*, vol. 13, no. 21. ISSN 2079-9292.
- YAĞCI, M., 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, vol. 9, no. 1, Article 11. ISSN 2196-7091.

PREDICTING STUDENT ACADEMIC PERFORMANCE: MODERN APPROACHES AND TECHNIQUES

Abstract. Student academic performance is an important indicator for measuring the quality of the educational process and contributes to the advancement of higher education institutions in rankings. Predicting student success allows for early prevention of dropout of students from the higher education system. Due to this, researchers have been interested in developing models for predicting student success in recent years. The article analyses the benefits of predicting student success for different stakeholder groups (students, teachers, leadership), which motivate the growing attention to such models. In a comparative perspective, several already developed models for predicting student success, based on machine learning (Machine Learning – ML) and “explainable artificial intelligence” (eXplainable Artificial Intelligence – XAI), are presented.

Keywords: student academic performance; prediction, artificial intelligence; machine learning; eXplainable Artificial Intelligence

✉ **Dr. Silvia Gaftandzhieva, Assoc. Prof.**

ORCID iD: 0000-0002-0569-9776

✉ **Prof. Dr. Rositsa Doneva**

ORCID iD: 0000-0003-0296-1297

University of Plovdiv “Paisii Hilendarski”

Plovdiv, Bulgaria

E-mail: sissiy88@uni-plovdiv.bg

E-mail: rosi@uni-plovdiv.bg