

УДК 004.8

ПРОГНОЗУВАННЯ АКАДЕМІЧНОЇ УСПІШНОСТІ СТУДЕНТІВ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Пилипенко В.І. – ДФКН-22, аспірант, pylypenko.vi@knutd.edu.ua

Стаценко В.В. – д.т.н., проф., statsenko.v@knutd.edu.ua

Київський національний університет технологій та дизайну

Метою роботи є розробка моделі для прогнозування академічної успішності студентів із використанням методів машинного навчання, зокрема алгоритму випадкового лісу. У рамках дослідження було проаналізовано наявні освітні дані. Із бази даних Moodle взято оцінки, результати тестів, оцінок за лабораторні та практичні роботи, відвідуваність на лекціях, лабораторних та практичних заняттях. А також дані взаємодії студентів з відеоматеріалами із розробленого плагіну який встановлено в систему Moodle [1]. Загальна кількість даних для обробки склала 1710 записів вибірок користувачів, які були розподілені у відношенні 342/1368. З яких тренувальна вибірка містила – 1368, а тестова – 342. Створення моделей для прогнозування виконано в середовищі розробки PyCharm на мові програмування Python. Для побудови моделей використано бібліотеку scikit-learn.

Випадковий ліс (Random Forest) – це ансамблевий алгоритм класифікації, який формує велику кількість дерев рішень і обирає найбільш ймовірний результат на основі "голосування" усіх дерев [2]. Математично процес голосування серед усіх дерев у випадковому лісі для визначення фінального класу \hat{y} для нового спостереження x описується так:

$$\hat{y} = \arg \max_{y \in Y} \sum_{j=1}^T I(T_j(x) = y) \quad (1)$$

де:

\hat{y} – фінальний прогноз, який є класом, до якого, ймовірно, належить спостереження x ;

$y \in Y$ – ймовірні класи, до яких може належати спостереження x , а Y – множина всіх можливих класів;

$T_j(x)$ – прогноз j -го дерева для об'єкта x , тобто клас або значення, яке це дерево передбачає для цього спостереження.

На рис. 1 зображено графік ROC-кривої із розрахованою площею під кривою (AUC) для побудованої моделі прогнозування успішності студентів.

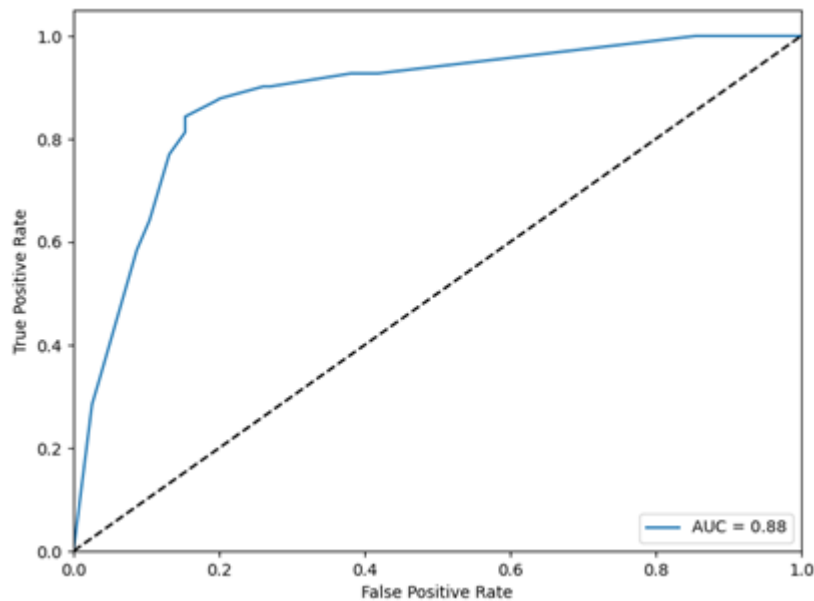


Рисунок 1 – Графік ROC-кривої для побудованої моделі прогнозування

Розрахунок значення загальної точності прогнозування склав – 87%, чутливість склала – 91%, збалансована точність – 82%, специфічність – 78%, F1 Score – 0,89.

Висновки. Прогнозування академічної успішності студентів досить важливе для навчального процесу, а точний результат прогнозування дає можливість реагувати на проблеми та вчасно допомагати студентам які цього потребують. Загальна точність прогнозування моделі склала – 87% на наявних даних. Отримані розрахунки говорять про збалансованість моделі та гарну здатність виявляти позитивні випадки. Отримані результати будуть використані для подальших досліджень, а отримана модель буде перевірена на більшій кількості даних.

Л і т е р а т у р а

1. Пилипенко, В., & Стаценко, В. (2025). ПЛАГІН ДЛЯ ЗБОРУ ДАНИХ ВЗАЄМОДІЇ КОРИСТУВАЧІВ MOODLE З ВІДЕО МАТЕРІАЛАМИ. Наука і техніка сьогодні, 1(42), 1318-1330. [https://doi.org/10.52058/2786-6025-2025-1\(42\)-1318-1330](https://doi.org/10.52058/2786-6025-2025-1(42)-1318-1330)
2. Benaya, D. (2024). Implementasi Random Forest dalam klasifikasi kanker paru-paru. JOINTER: Journal of Informatics Engineering, 5(01), 27-31.