УДК ?

О. М. Козачко
А. О. Лотоцький

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ РІВНЯ ЗНАНЬ ІНОЗЕМНОЇ МОВИ СТУДЕНТІВ ЗАКЛАДУ ВИЩОЇ ОСВІТИ

Вінницький національний технічний університет;

Анотація

Проведено аналіз даних та відібрано оптимальний набір ознак, за якими слід будувати модель. Розроблено комп'ютерну програму, яка прогнозує результат оцінювання з англійської мови та визначає які навчальні предмети найбільше впливають на оцінювання англійської мови.

Ключові слова: побудова моделей, аналіз, іноземна мова, виявлення закономірностей..

Abstract

An exploratory analysis of the data was performed and the optimal set of features on which to build the model was selected. An analysis of existing methods of analysis to solve the problem was done. A computer program has been developed, predicts the outcome of the assessment in English and determines which subjects have the greatest impact on the assessment of English.

Keywords *building models, analysis, foreign language, identification of regularities regarding influence factors*.

З стрімким розповсюдженням інформаційних технологій, збільшується і кількість збереженої інформації. Ця інформація може бути використана для проведення системного аналізу, тобто, виявлення загальних тенденцій змін, факторів, що впливають на них. За допомогою виявлених тенденцій та факторів можна оптимізувати роботу системи, підвищити продуктивність, скоротити кількість втрат. Також, сфера дослідження є однією з найбільш перспективних. Для будь-кого є очевидною важливість вивчення англійської мови в наш час. Англійська мова відкриває нескінченні можливості у повсякденному та професіональному житті. У статті «Прогнозування успішності навчання студентів – один із напрямів підвищення якості освіти» [1] розглянуто та доведено, що якість проведення навчальних занять має прямий вплив на успішність навчання студентів, тому проблема є актуальною.

Аналізованими даними є результати оцінювання студентів Вінницького національного медичного університету ім. М. І. Пирогова за 2, 4, 6 семестри навчання та результати першого етапу ЄДКІ. Спочатку дані було підготовлено до аналізу шляхом видалення пустих значень та непотрібних колонок. На рисунку 1 зображено приклад даних готових до прогнозування.

Військова гігієна	Загальна хірургія	Патоморфологія	Патофізіологія	Пропедевтика внутрішньої медицини	Пропедевтика педіатрії	Сестринська практика	Фармакологія	Іноземна мова_у	Тест
166.0	174.0	135.0	130.0	142.0	135.0	188.0	147.0	144.0	53
188.0	169.0	173.0	143.0	165.0	164.0	180.0	161.0	186.0	71
122.0	149.0	132.0	130.0	126.0	136.0	160.0	140.0	133.0	42
185.0	166.0	173.0	194.0	161.0	165.0	191.0	187.0	174.0	92
150.0	154.0	139.0	130.0	132.0	145.0	151.0	141.0	148.0	75

Рисунок 1 – приклад даних

Для кожного з датасетів було використоно 4 моделі: Lightgbm, Xgboost, Logistic Regeression та Random Forest. Першим було проаналізовно датасет з даними за 2 семестр навчання. На рисунку 2 показано діаграму важливості ознак lightgbm моделі.

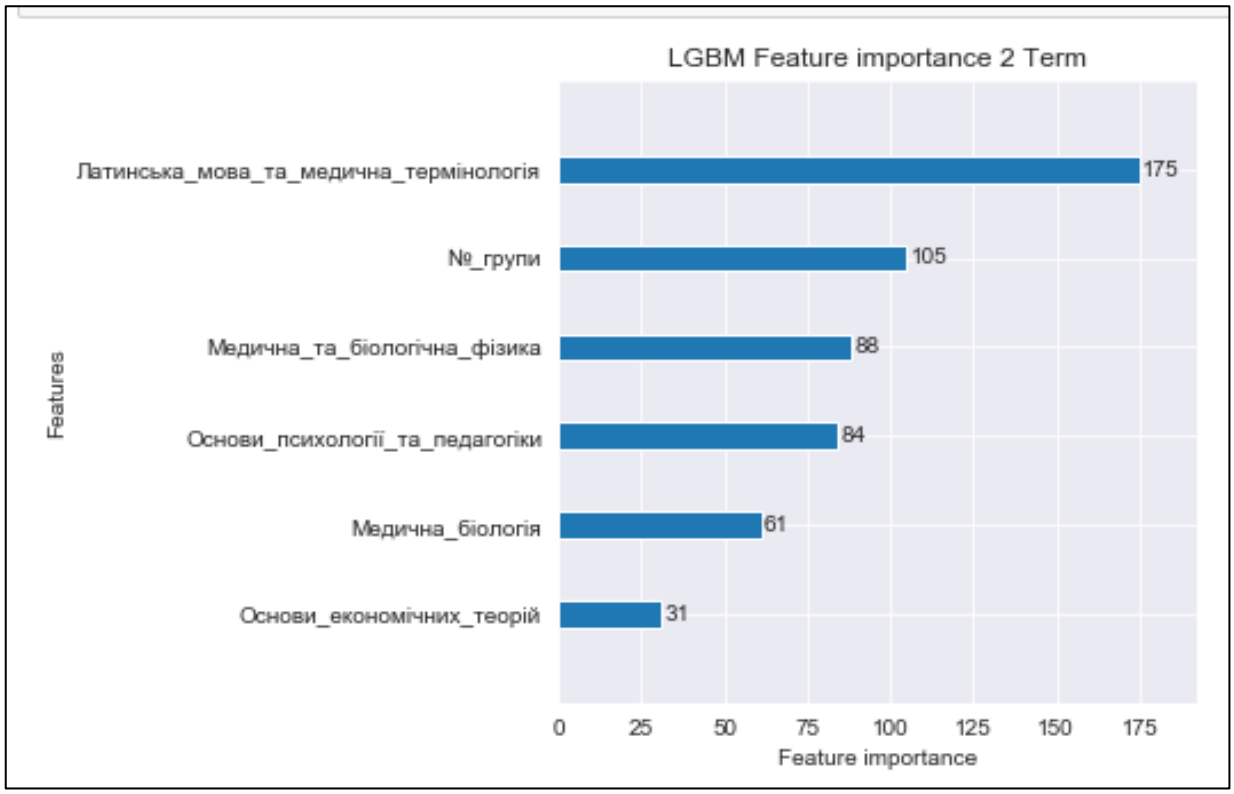

Рисунок 2 – Діаграма важливості ознак lightgbm моделі

На рисунку 3 показано діаграму важливості ознак xgboost моделі.

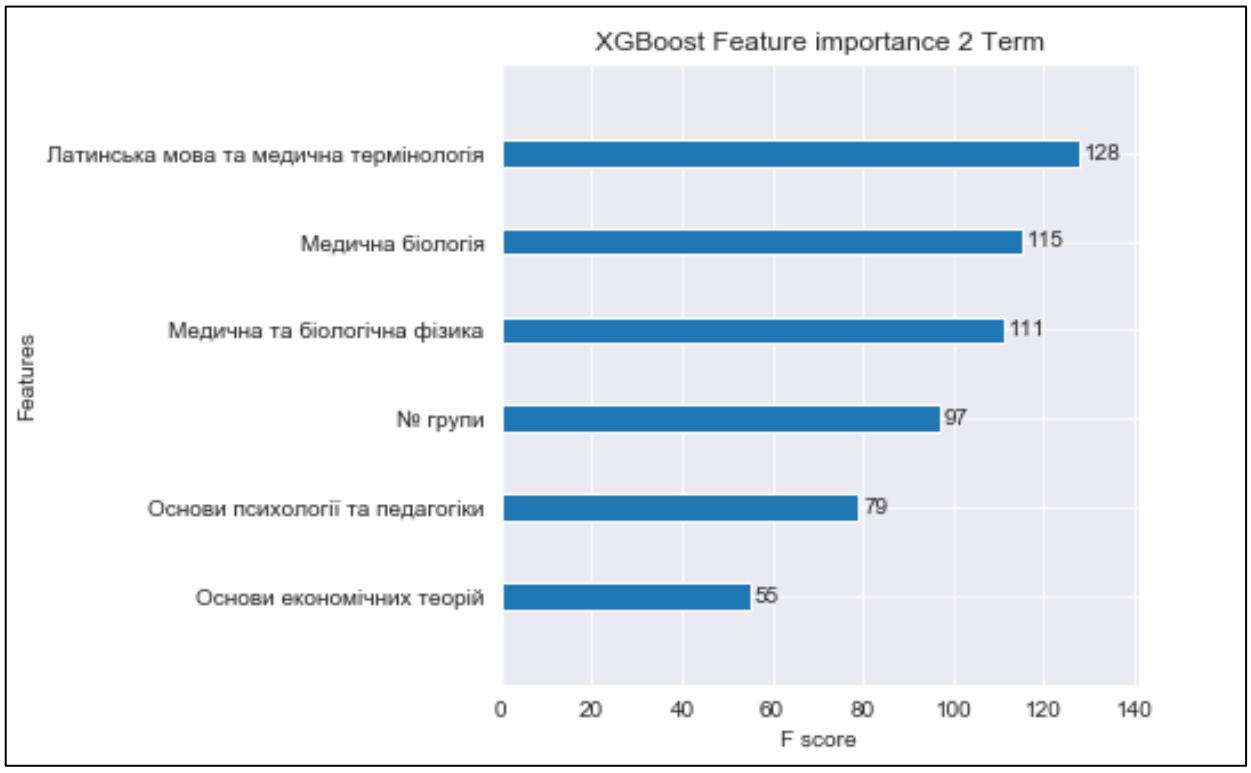

Рисунок 3 – Діаграма важливості ознак xgboost моделі

Основною ознакою lightgbm та xgboost моделей стала колонка «Латинська мова та медична термінологія». На основі цього можна зробити висновок, що «Латинська мова та медична термінологія» є найбільш впливовою на рівень вивчення іноземної мови студентами 2 семестру

навчання. Також, було побудовано матриці невідповідностей для моделей Logistic Regeression (рис. 4) та Random Forest (рис. 5)

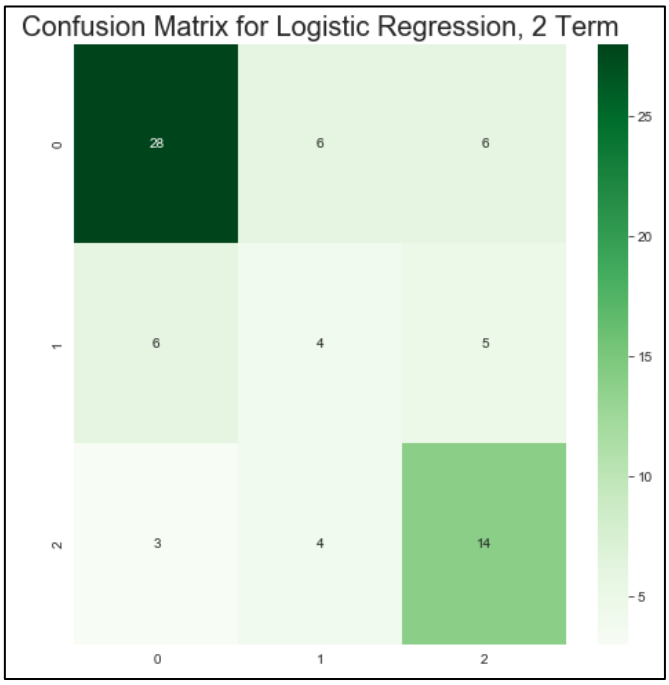

Рисунок 4 – Матриця невідповідностей Logistic Regression

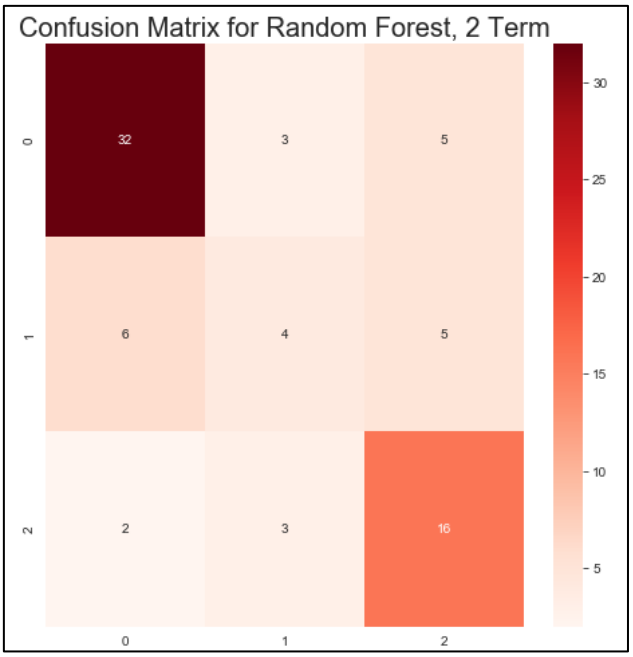

Рисунок 5 – Матриця невідповідностей Random Forest

Наступним було проаналізовано датасет з даними за 4 семестр навчання. На рисунку 6 показано діаграму важливості ознак lightgbm моделі.

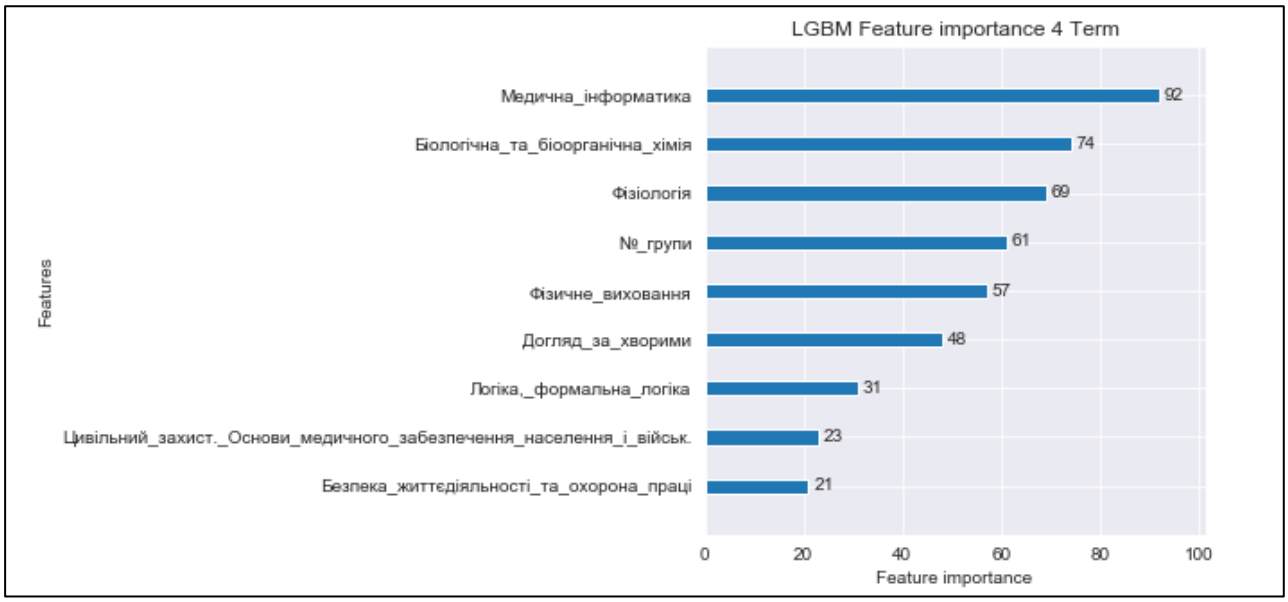

Рисунок 6 – Діаграма важливості ознак lightgbm моделі

На рисунку 7 показано діаграму важливості ознак xgboost моделі.

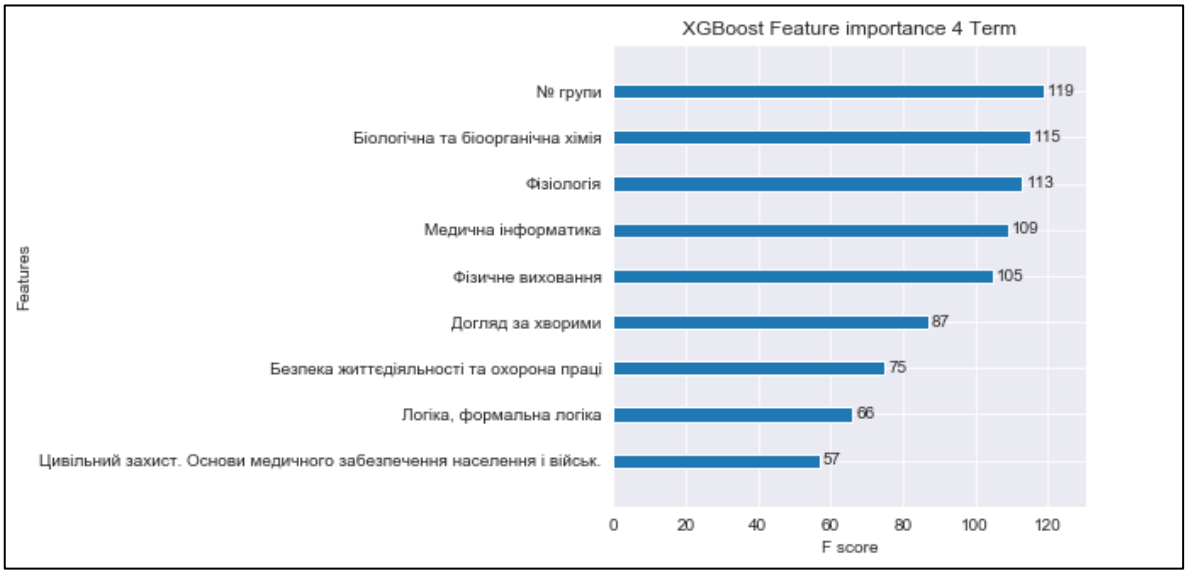

Рисунок 7 – Діаграма важливості ознак xgboost моделі

Основною ознакою lightgbm стала колонка «Медична інформатика». Але на діаграмі ознак xgboost вона зайняла 5 місце, проте перші 5 ознак мають дуже близькі значення важливості. На основі цього можна зробити висновок, що «Медична інформатика» є найбільш впливовою на рівень вивчення іноземної мови студентами 4 семестру навчання. Матриці невідповідностей для моделей Logistic Regeression та Random Forest зображено на рисунках 8 та 9 відповідно.

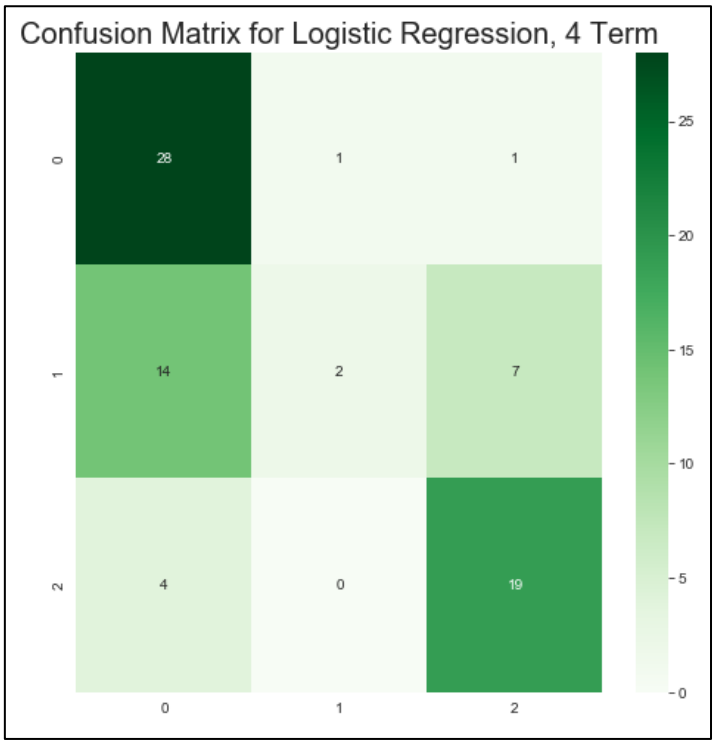

Рисунок 8 – Матриця невідповідностей Logistic Regression

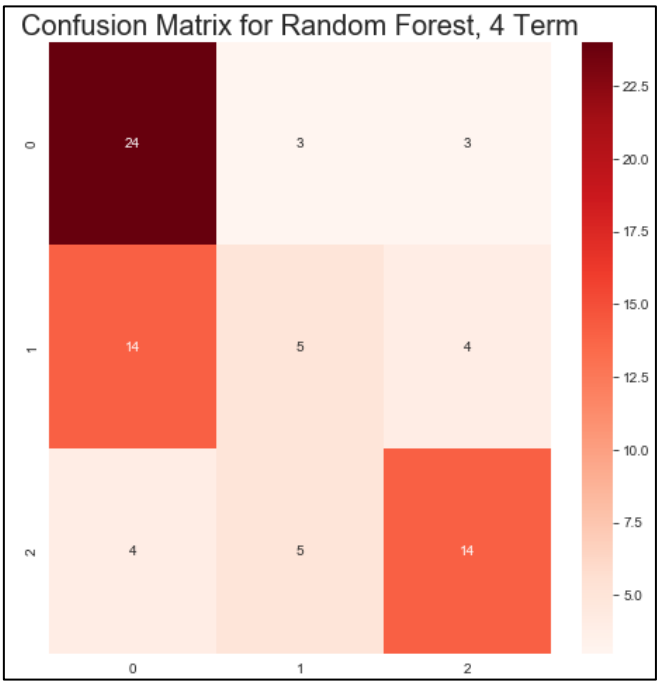

Рисунок 9 – Матриця невідповідностей Random Forest

Останнім було проаналізовано датасет з результатами 6 семестру начання. На рисунку 10 показано діаграму важливості ознак lightgbm моделі.

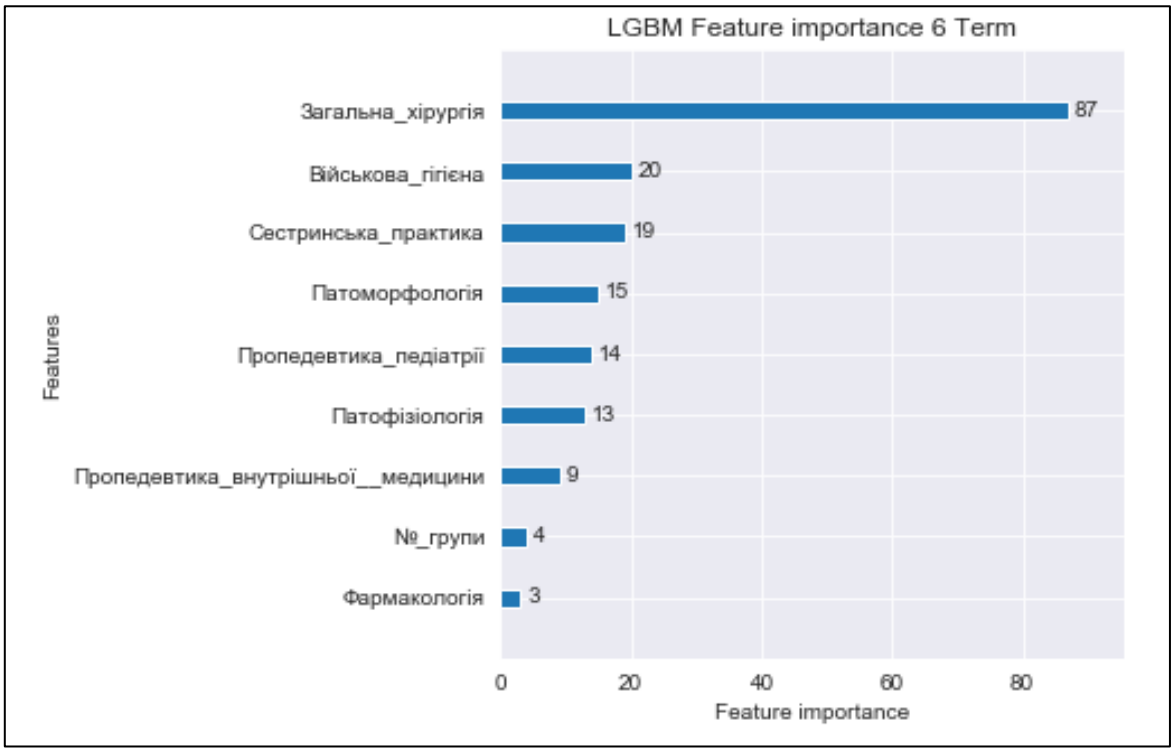

Рисунок 10 – Діаграма важливості ознак lightgbm моделі

На рисунку 11 показано діаграму важливості ознак xgboost моделі.

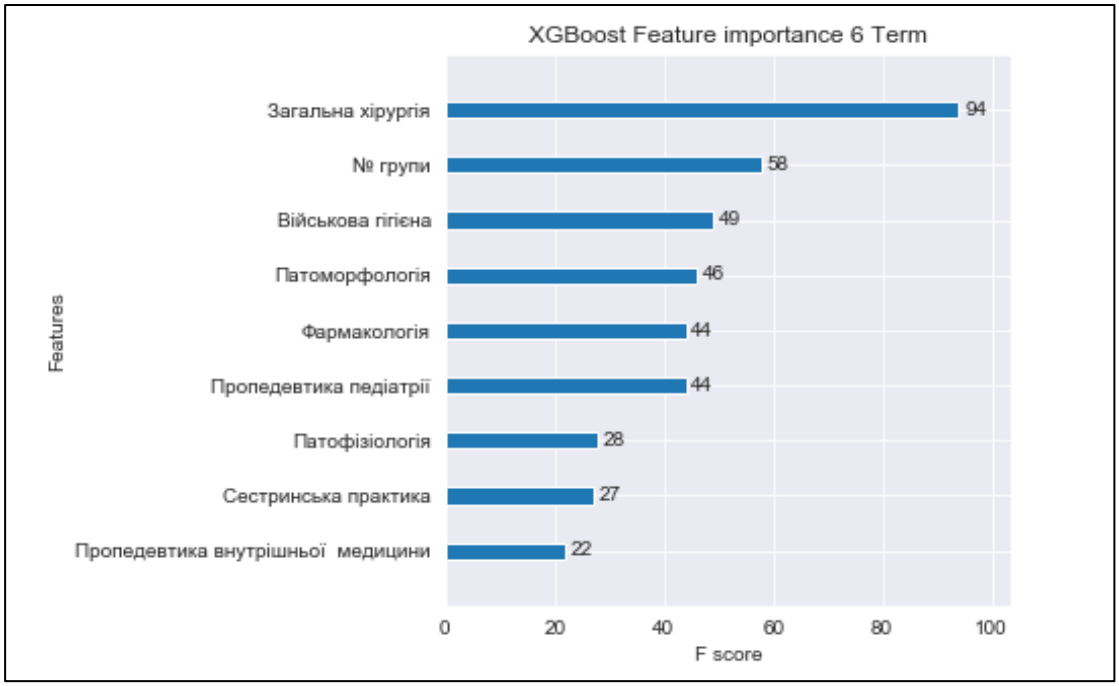

Рисунок 12 – Діаграма важливості ознак xgboost моделі

Основною ознакою lightgbm та xgboost моделей однозначно стала колонка «Загальна хірургія. На основі цього можна зробити висновок, що «Загальна хірургія» є найбільш впливовою на рівень вивчення іноземної мови студентами 6 семестру навчання.

Матриці невідповідностей для моделей Logistic Regeression та Random Forest зображено на рисунках 13 та 14 відповідно.

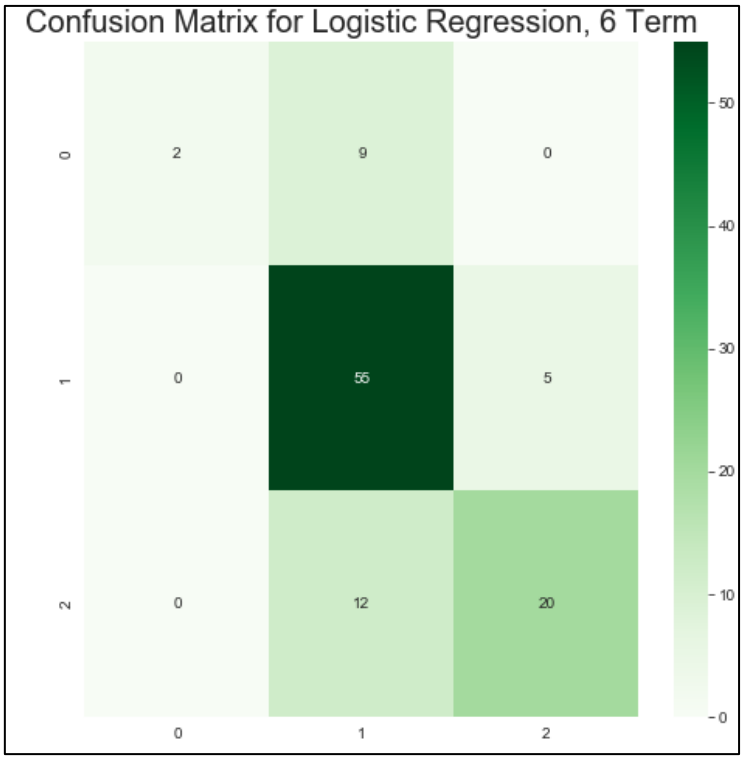

Рисунок 13 – Матриця невідповідностей

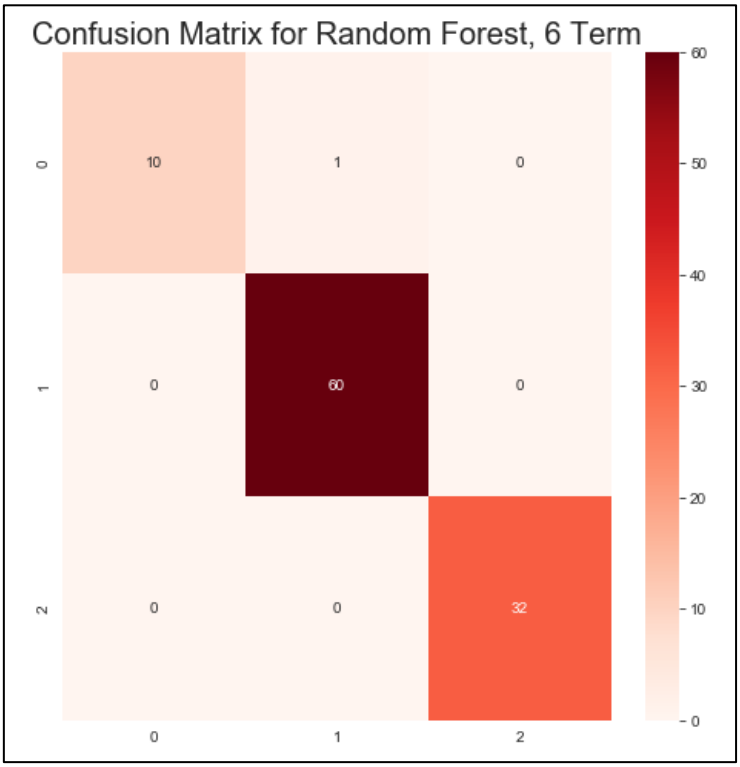

Рисунок 14 – Матриця невідповідностей

На рисунку 15 зображено точності передбачення усіх використаних моделей.

Точність моделей (Accuracy)			
	2 семестр	4 семестр	6 семестр
LGBM	0.70	0.65	0.60
XGBoost	0.80	0.72	0.88
Logistic regression	0.60	0.64	0.75
Random Forest	0.70	0.60	0.99

Рисунок 15 – Точності моделей

В середньому найвищу точність має модель градієнтного бустингу Xgboost. Але найточнішою в одиночному випадку є модель Random Forest з точністю прогнозу 99%, використана на датасеті з даними 6 семестру навчання.

Висновки

Розроблений програмний код, прогнозує оцінку з іноземної мови.

Аналіз даних показав, що:

– «Латинська мова та медична термінологія» є найбільш впливовою на рівень вивчення іноземної мови студентами 2 семестру навчання;

– «Медична інформатика» є найбільш впливовою на рівень вивчення іноземної мови студентами 4 семестру навчання;

– «Загальна хірургія» є найбільш впливовою на рівень вивчення іноземної мови студентами 6 семестру навчання;

– «Патофізіологія», «Фізіологія» та «Іноземна мова» за 6 семестр навчання є найбільш впливовими на результат здачі ЄДКІ з іноземної мови.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. ПРОГНОЗУВАННЯ УСПІШНОСТІ НАВЧАННЯ СТУДЕНТІВ – ОДИН ІЗ НАПРЯМІВ ПІДВИЩЕННЯ ЯКОСТІ ОСВІТИ. Збірник науково-методичних праць «Удосконалення освітньо-виховного процесу в закладі вищої освіти». 2020. № 23. – Режим доступу до ресурсу: http://elar.tsatu.edu.ua/bitstream/123456789/10547/1/%d0%a1%d0%b1%d0%be%d1%80%d0%bd%d0%b8%d0%ba%20%d1%81%d1%82%d0%b0%d1%82%d0%b5%d0%b9%202020%208.04-58-65.pdf

Лотоцький Андрій Олександрович — студент групи 2ІСТ-20м, Кафедра системного аналізу, комп'ютерного моніторингу та інженерної графіки, Вінницький національний технічний університет, Вінниця; e-mail: and.lototskyi@gmail.com;

Науковий керівник: ***Козачко Олексій Миколайович —*** к.т.н., доцент, доцент кафедри системного аналізу комп'ютерного моніторингу та інженерної графіки, Вінницький національний технічний університет,Вінниця, e-mail: lekoz80@gmail.com.

Lototskyu Andrii O. – Department of system analysis, computer monitoring and engineering graphics, , Vinnytsia National Technical University, Vinnitsa;

*Supervisor****: Kozachko Oleksiy M. —*** Cand. Sc. (Eng), Assistant Professor of Department of system analysis, computer monitoring and engineering graphics, Vinnytsia National Technical University, Vinnytsia, lekoz80@gmail.com.