

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій

Бакалаврська кваліфікаційна робота на тему:

«ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ
СТРЕСУ У СТУДЕНТІВ ЗАКЛАДІВ ВИЩОЇ ОСВІТИ»

Виконала: студентка 4 курсу, групи 2ІСТ-216
спеціальності 126 – «Інформаційні системи та
технології»

 Марина ВЕРЧИНСЬКА

Керівник: Ph.D., ст. викладач каф. САІТ

 Ольга ВОЙЦЕХОВСЬКА

« 05 » 06 2025 р.

Рецензент: к.т.н., асистент каф. КН

 Ілона БОГАЧ

« 10 » 06 2025 р.

Допущено до захисту

Завідувач кафедри САІТ

 д.т.н., проф. Віталій МОКІН

« 06 » 06 2025 р.

Вінниця ВНТУ – 2025 рік

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій
Рівень вищої освіти – перший (бакалаврський)
Галузь знань – 12 Інформаційні технології
Спеціальність 126 – «Інформаційні системи та технології»
Освітньо-професійна програма – Прикладні інформаційні технології

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

Мокін д.т.н., проф. Віталій МОКІН

“28” 05 2025 року

ЗАВДАННЯ
НА БАКАЛАВРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
Верчинській Марині Олександрівні

1. Тема роботи: «Інформаційна технологія аналізу та прогнозування стресу у студентів закладів вищої освіти»

керівник роботи: Войцеховська О.О., Ph.D., ст. викладач каф. САІТ

Затверджені наказом ВНТУ від «20» 05 2025 року № 97

2. Термін подання студентом роботи 31.05.25р.

3. Вихідні дані до роботи:

1) Kaggle Dataset «Student Stress Factors: A Comprehensive Analysis»

4. Зміст текстової частини:

1) Аналіз предметної області;

2) Вибір оптимального рішення та налаштувань для розв’язання поставленої задачі;

3) Розроблення інформаційної технології аналізу та прогнозування стресу у студентів.

5. Перелік ілюстративного матеріалу:

1) Кореляційна матриця факторів стресу студентів у закладів вищої освіти;

2) Гістограми розподілу даних;

3) Кількісний розподіл рівнів стресу у відсотках;

4) Матриці плутанини для моделей класифікації;

4) Оцінка точності роботи моделей;

5) Графік важливості ознак класифікації рівня стресу за моделлю Випадковий ліс.

6. Консультанти розділів роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання 28.05 25 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва та зміст етапу	Термін виконання		Примітка
		початок	закінчення	
1	Аналіз предметної області	01.04	15.04	вм
2	Аналіз наукових робіт дослідження стресу у студенті	15.04	30.04	вм
3	Вибір оптимального рішення та налаштувань для розв'язання поставленої задачі	01.05	10.05	вм
4	Розроблення інформаційної технології аналізу та прогнозування стресу у студентів	10.05	20.05	вм
5	Оформлення матеріалів до захисту БКР	20.05	30.05	вм

Студентка

Марина ВЕРЧИНСЬКА

Керівник роботи

Ольга ВОЙЦЕХОВСЬКА

АНОТАЦІЯ

Бакалаврська кваліфікаційна робота складається з 87 сторінок формату А4, на яких є 63 рисунків, список використаних джерел містить 26 найменувань.

Метою роботи є підвищення точності прогнозування стресу у студентів закладів вищої освіти шляхом створення інформаційної технології та використання методів машинного навчання.

В бакалаврській роботі проведено аналіз предметної області. Досліджено основні причини стресу студентів закладів вищої освіти та як стрес впливає на успішність студентів. Проведено аналіз методів прогнозування стресу у студентів закладів вищої освіти. Здійснено вибір оптимальних інформаційних технологій для інформаційної технології прогнозування стресу у студентів

В роботі наведено розроблення інформаційної технології на платформі Kaggle, на мові програмування Python. Проведено розвідувальний аналіз даних, було очищено та оптимізовано вхідні дані. Натреновано та обрано моделі машинного навчання для прогнозування рівню стресу студентів закладів вищої освіти. Проведено тестування створених моделей та обрано оптимальну модель машинного навчання для прогнозування рівня стресу студентів. За результатами тестування відібрано оптимальну модель машинного навчання для прогнозування рівня стресу студентів закладів вищої освіти.

Ключові слова: інформаційна технологія, стрес, студенти, машинне навчання, аналіз, прогнозування, заклад вищої освіти, Python.

ABSTRACT

The bachelor's thesis consists of 87 A4 pages, which contain 63 figures, the list of sources used contains 26 titles.

The aim of the work is to improve the accuracy of stress prediction in students of higher education institutions by creating an information technology and the use of machine learning methods.

The bachelor's thesis analyzes the subject area. The main causes of stress of students of higher education institutions and how stress affects student performance are investigated. The analysis of methods for predicting stress in students of higher education institutions is carried out. The choice of optimal information technologies for information technology for predicting stress in students is made.

The paper presents the development of information technology on the Kaggle platform, in the Python programming language. An exploratory data analysis was conducted, the input data was cleaned and optimized. Machine learning models were trained and selected to predict the level of stress among university students. The created models were tested and the optimal machine learning model for predicting the level of student stress was selected. Based on the test results, the optimal machine learning model for predicting the level of stress of students of higher education institutions is selected.

Keywords: information technology, stress, students, machine learning, analysis, forecasting, higher education institution, Python.

ЗМІСТ

ВСТУП	3
1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	5
1.1 Опис об'єкта досліджень	5
1.2 Аналіз наукових робіт дослідження стресу у студентів	7
1.3 Аналіз методів прогнозування стресу у студентів	17
1.4 Висновки	21
2. ВИБІР ОПТИМАЛЬНОГО РІШЕННЯ ТА НАЛАШТУВАНЬ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ	23
2.1 Аналіз середовища та мови програмування для розробки інформаційної технології прогнозування стресу у студентів	23
2.2 Розвідувальний аналіз даних	25
2.3 Висновки	46
3. РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ СТРЕСУ У СТУДЕНТІВ.....	48
3.1 Розроблення інформаційної технології	48
3.2 Налаштування моделей	52
3.3 Використання моделей машинного навчання.....	56
3.4 Висновки	67
ВИСНОВКИ.....	68
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	70
Додаток А (обов'язковий) Технічне завдання	75
Додаток Б (обов'язковий) Протокол перевірки кваліфікаційної роботи.....	77
Додаток В (довідниковий) Фрагмент лістингу програми	78
Додаток Г (обов'язковий) Ілюстративна частина	83

ВСТУП

Актуальність теми. Проблема стресу серед студентської молоді є надзвичайно актуальною в умовах сучасного освітнього середовища, що характеризується високим рівнем навантаження, нестабільністю соціально-економічної ситуації та стрімкою диджиталізацією навчального процесу. Часте переживання стресових станів негативно впливає не лише на емоційне здоров'я студентів, але й на їхню успішність, мотивацію до навчання та загальну якість життя.

У зв'язку з цим виникає необхідність у впровадженні ефективних цифрових рішень для моніторингу, аналізу та прогнозування рівня стресу серед студентів. Інформаційні технології, зокрема методи машинного навчання, відкривають нові можливості для виявлення закономірностей, які недоступні традиційним методам аналізу.

Тому актуальною є задача по створенню інформаційної технології аналізу та прогнозування рівня стресу у студентів, яка дозволить не лише підвищити ефективність психологічної підтримки, а й своєчасно виявляти групи ризику, що сприятиме формуванню більш безпечного й сприятливого освітнього середовища.

Мета і задачі дослідження. Метою дослідження є підвищення точності прогнозування стресу у студентів закладів вищої освіти шляхом створення інформаційної технології та використання методів машинного навчання.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- провести аналіз предметної області;
- здійснити вибір оптимальних інформаційних технологій для інформаційної технології прогнозування стресу у студентів;
- здійснити розвідувальний аналіз даних;
- розробити інформаційну технологію для прогнозування стресу у студентів.

Об'єктом дослідження є процес розроблення інформаційної технології для прогнозування стресу у студентів закладів вищої освіти.

Предметом дослідження є методи машинного навчання і програмні засоби, які використовуються у процесі створення інформаційної технології для прогнозування стресу у студентів закладів вищої освіти.

Публікації: За результатами даної роботи були опубліковані тези «Розвідувальний аналіз даних для інформаційної технології аналізу та прогнозування стресу у студентів закладів вищої освіти методами машинного навчання» на Міжнародній науково-практичній інтернет-конференції студентів, аспірантів та молодих науковців «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2024-2025 рр.) [1].

1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Опис об'єкта досліджень

Об'єктом дослідження бакалаврської роботи виступає інформаційна технологія, спрямована на прогнозування рівня стресу у студентів закладів вищої освіти за допомогою методів машинного навчання. Предмет дослідження становлять результати використання цих методів, які дозволяють визначити основні фактори, що впливають на рівень стресу у студентів. Метою дослідження є виявлення взаємозв'язків між певними чинниками (наприклад: соціальна взаємодія, академічне навантаження) та рівнем стресу.

Стрес у студентському середовищі є поширеним явищем, яке виникає внаслідок різноманітних питань, пов'язаних з адаптацією до нових навчальних середовищ, високого рівня відповідальності, необхідності поєднання навчання з особистим життям, з роботою. Фізичні, емоційні та поведінкові зміни можуть відображатись у: виснаженні, страху, дратівливості. Студенти першого курсу, які досі намагаються пристосуватися до нової освітньої системи та соціального середовища, особливо вразливі до стресу.

До основних чинників стресу студентів відносять: важке академічне навантаження, часті іспити, фінансові труднощі, невизначеність щодо майбутнього, низька самооцінка та соціальне відчуження. Стрес може бути мотивуючим фактором у деяких випадках, але в більшості випадків це призводить до виснаження, зниження продуктивності та психічної нестабільності [2].

Під час проходження навчання студенти можуть втрачати інтерес до освітнього процесу, а саме: відсутність бажання відвідування занять, неактивна участь у житті закладу вищої освіти та також падіння успішності студента. Стрес – це негативна емоція на зовнішні подразники. Хоча певний стрес необхідний для особистісного зростання. Вплив стресу охоплює широкий спектр психологічних і фізіологічних проявів, серед яких порушення

когнітивних функцій, таких як: концентрація та пам'ять, емоційна дисрегуляція, а також різноманітні соматичні симптоми. Розуміння багатогранної природи стресу та використання передових моделей машинного навчання сприяє не лише виявленню груп ризику, але й розробці індивідуальних стратегій втручання. Ці стратегії спрямовані на зміцнення психологічної стійкості та полегшення академічної адаптації студентів.

З 2022 року Україна перебуває в стані війни, що супроводжується безперервним потоком інформації, необхідністю систематичного моніторингу потенційних конфліктів та підвищенням рівня усвідомлення щодо існуючих загроз. Така ситуація створює суттєве навантаження на нервову систему студентів, ставлячи перед ними додаткові психологічні виклики.

Більшість студентів змушені навчатися онлайн, що виявляється менш ефективним порівняно з очним навчанням. Це створює певні проблеми, оскільки під час очного навчання є можливість навчитися адаптуватися до нового середовища та розвивати міжособистісні стосунки з однолітками та викладачами [3].

Розглянемо детальніше основні фактори, що впливають на рівень стресу студентів закладів вищої освіти [4].

Академічний стрес – вважається одним із ключових чинників, що спричиняють напругу в студентів, проте цей аспект їхнього життя часто залишається поза увагою. Наслідки такого стресу проявляються не тільки на психологічному рівні, у вигляді депресії чи тривожності, але також можуть переходити у фізичні розлади, такі як: болі в животі, нудота та проблеми зі сном. Було встановлено, що академічний стрес часто спричиняється високими очікуваннями щодо результатів у навчанні, він може негативно впливати на загальну успішність. Таким чином, підвищений стрес призводить до зниження результатів успішності.

Соціально-економічний аспект стресу – окрім, тиску, пов'язаного з досягненням академічного успіху, стрес може формуватися під впливом зовнішніх чинників, таких як: сімейні обставини і культурні традиції. Наукові

дослідження свідчать, що студенти з родин із низьким рівнем доходу, переживають значно вищий рівень стресу. Зокрема, молодь, що переїжджає до великих міст для навчання, стикається з додатковими викликами, у вигляді підвищених витрат на проживання, харчування, транспорт тощо.

Психологічний стрес – це стан емоційного та розумового напруження, яка виникає тоді, коли людина сприймає вимоги життєвих ситуацій як, перевищення її адаптивної можливості. Це відчуття дисбалансу між тим, що від нас вимагається та наявними ресурсами, готуючи організм до боротьби або втеч. Цей внутрішній конфлікт може проявлятися на різних рівнях: від дратівливості до емоційних спалахів.

Особистісний стрес – це складний психоемоційний стан, який виникає на основі індивідуальних уявлень про світ, особистої системи цінностей, внутрішніх суперечностей та моральних дилем. Він може бути зумовлений конфліктом між бажаним і реальним, а також розривом між ідеалами особистості та її діями. Часто джерелом такого стресу виступають завищені очікування від себе, прагнення до досконалості, що створює тиск і веде до тривалого емоційного виснаження.

1.2 Аналіз наукових робіт дослідження стресу у студентів

У статті «Assessment and Analysis of Factors Impacting Stress in Higher Educational Institutes in Delhi-NCR», авторів Kajleen Kaur, Harpreet Kaur, Jasmin Kaur та Yogeita S.Mehra, досліджено чинники, що впливають на сприйняття стресу серед студентів вищих навчальних закладів. Дослідження ґрунтується на первинному опитуванні 633 студентів і використовує упорядковану логістичну регресію для аналізу впливу демографічних та університетських факторів [5].

На рисунку 1.1 представлено пункти, що використовувалися в дослідженні [4] для вимірювання сприйняття стресу студентами.

v1	Competition with my peers for grades is quite intense.
v2	My parents are critical of my academic performance.
v3	Over-expectations of my teachers stress me out.
v4	The biased/critical nature of my teachers stresses me out.
v5	My family has huge expectations of me.
v6	The size of the curriculum is excessive.
v7	I can manage extracurricular activities with my coursework.
v8	I cannot afford to miss classes.
v9	The examination questions are too lengthy.
v10	The examination questions are usually difficult.
v11	I am worried about failing in my final exams.
v12	I do not get sufficient personal/family time.
v13	I do not get sufficient recreational breaks.
v14	There are too many assignments to be submitted.
v15	Feel frustrated due to limited choices in the course curriculum.
v16	Feel stressed out due to social media addiction.
v17	The financial cost of studies is a cause of stress.
v18	I face gender discrimination at home/college.
v19	Feel stressed due to fear of not being selected for higher studies.
p1	I will be a successful student.
p2	I will be successful in my career.
p3	I fear not being able to get into a Higher Education programme of my choice.
p4	I am worried about getting a job that meets my expectations.
p5	I will have to do an additional course to increase my employability prospects.
p6	Experiencing Mental health issues (anxiety or depressive thoughts)

Рисунок 1.1 – Пункти, що використовуються для вимірювання сприйняття стресу студентами

Це дослідження виявило, що понад 65% студентів зустрічаються з середнім або високим рівнем стресу. Найбільш суттєвими чинниками, що впливають на сприйняття стресу, є: тиск, пов'язаним з іспитами, навчальною програмою та часом.

Також у роботі [5] вивчалася різниця у сприйнятті стресу залежно від щомісячного сімейного доходу студентів. Як показано на рисунку 1.2, сприйняття стресу в категорії з найменшим доходом є найвищим, тоді як у категорії з найвищим щомісячним доходом – найнижчим. Показник сприйняття стресу падає для категорій з вищими доходами.

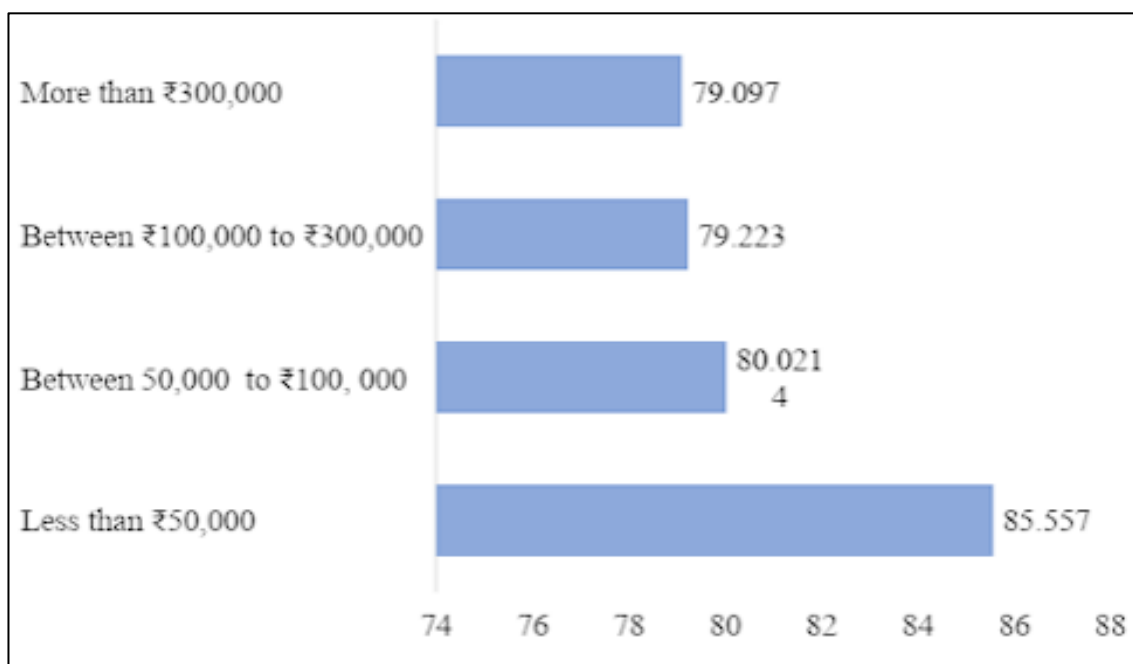


Рисунок 1.2 – Різниця у сприйнятті стресу в залежності від щомісячного сімейного доходу студентів

У статті «A study to assess the level of stress among college students in a selected Govt авторів Rais Ahmad, Priyanka Chaudhary та Rajwant Kaur Randhawa, досліджено рівень стресу серед студентів Державного університету медсестринства у Срінагарі, Джамму і Кашмір (Індія). Це дослідження, спрямоване на кількісну оцінку рівня стресу та його зв'язку з обраними демографічними змінними. На рисунку 1.3 представлено частотний розподіл рівня стресу, з якого видно, що більшість студентів (63,4%) повідомили про помірний рівень стресу, 3,3% – про високий рівень стресу і 33,3% – про низький рівень стресу [6].

To determine the level of stress, the study employed descriptive statistics (frequency distribution).
Table1. Frequency distribution of the level of stress. (N=30)

Levels of stress	Frequency (N)	Percentage (%)
Lower	10	33.3
Moderate	19	63.4
Higher	1	3.3
Total	30	100.0

Рисунок 1.3 – Частотний розподіл рівня стресу.

Дослідження базувалося на первинному опитуванні 30 студентів (N=30), відібраних випадковим чином. Для оцінки рівня стресу використовується Шкала сприйнятого стресу. Важливі результати цього дослідження показують, що рівень стресу студентів помірний (рис. 1.4).

Table 2. Summary of cross-tabulation for Gender and level of stress (N=30)

Gender		Level of overall stress			
		Lower	Moderate	Higher	Total
Male	Number	7	10	0	17
	%age	41.30%	58.70%	0.00%	100.00%
Female	Number	4	7	2	13
	%age	30.7%	54.0%	15.3%	100.00%
Total	Number	11	17	2	30
	%age	36.6%	56.6%	6.2%	100.00%

Рисунок 1.4 – Зв'язок між рівнем стресу та обраною демографічною змінною (стать)

Результати показали, що більшість учасників-чоловіків - 10 (58,7%) з 17 - мають помірний рівень стресу, решта учасників чоловічої статі - 7 (41,3%) - мали нижчий рівень стресу. Однак не було жодного студента-чоловіка, який би відчував високий рівень стресу. Що стосується учасниць, то більшість з них, 7 (54,0%) з 13, мали помірний рівень стресу, 4 (30,7%) - нижчий і 2 (6,2%) -

вищий рівень стресу. Розподіл частот чоловіків і жінок вказує на те, що вони мали більш-менш схожий рівень стресу, хоча студенти-жінки дещо більше схильні до стресу, ніж їхні колеги.

У роботі «Дослідження рівня стресу у студентів технічного закладу вищої освіти», авторів Кондратенка Є. М. та Церковної О. В. проведено комплексну оцінку рівня стресу серед студентів технічного закладу вищої освіти. Це дослідження є актуальним, оскільки враховує фактори, такі як: адаптація до університетського середовища, вплив пандемії Covid-19 та бойових дій на території України [7].

У своєму дослідженні автори провели анкетування 2334 студентів 1-3 курсів усіх факультетів ХНУРЕ, серед яких 1573 (67.4%) юнаків та 761 (32.6%) дівчат, середній вік яких склав 19 років. Для визначення рівня стресу і виявлення найбільш характерних ознак стресу було використано тест «Комплексна оцінка проявів стресу Ю.В. Щербатих». Респонденту пропонувалося співвідносити свої суб'єктивні переживання з різними ознаками стресу, поділеними на чотири групи: інтелектуальні, фізіологічні, емоційні та поведінкові (по 12 ознак у кожній групі).

Дослідження [7] показало, що в умовах воєнних дій на території України більшість населення, у тому числі й студенти, опинилися під впливом стресових факторів. Виявлено особливості стресової реакції студентів. У досліджуваних студентів найбільш виявлені емоційні симптоми стресу (12 з 12 показників набрали більше 40% студентів, інтелектуальні симптоми стресу були виявлені тим, що понад 40% студентів відзначили 5 із 12 симптомів. Згідно проведеного дослідження відносно менш вираженими виявились фізіологічні симптоми. Таким чином дослідження, показує, що на більшість студентів університету впливають ворожі фактори. Найбільш помітно, емоційні та інтелектуальні симптом стресу, які впливають на якість навчання та загальне самопочуття студентів.

У статті «College student's learning stress, psychological resilience and learning burnout: status quo and coping strategies» автори Zhen Gong, Huadi Wang, Mingxia Zhong та Yuling Shao досліджують взаємозв'язок між навчальним стресом, психологічною стійкістю та синдромом емоційного вигорання серед студентів. У дослідження взяли участь 16800 студентів, які пройшли стандартизовані опитувальники для оцінки рівня стресу, вигорання та психологічної стійкості. Результати показали, що психологічна стійкість значною мірою знижує негативний вплив навчального стресу на емоційне вигорання [8].

Для ілюстрації рівня психологічної стійкості студентів у різних аспектах автори наводять такі середні показники (рис. 1.5).

Items	Average scores
Target focus	17.34 ± 3.87
Emotion control	18.01 ± 4.05
Positive cognition	14.78 ± 4.33
Family support	20.68 ± 3.46
Interpersonal assistance	19.29 ± 5.01
Total score of psychological resilience	90.06 ± 11.85

Рисунок 1.5 – Середні бали психологічної стійкості студентів

На рисунку 1.5 видно, що найвищі середні значення спостерігаються у категоріях «Сімейна підтримка» та «Міжособистісна допомога», що підкреслює важливу роль соціального оточення у підтримці психологічної стійкості студентів.

Дослідження також показало, що студенти з високим рівнем психологічної стійкості значно рідше проявляли симптоми емоційного вигорання, навіть за умов інтенсивного навчального навантаження. Це підтверджує, що стійкість відіграє посередницьку роль у взаємозв'язку між стресом і психоемоційним станом. Автори, акцентують увагу на необхідності

впровадження профілактичних програм, спрямованих на розвиток адаптивного мислення, емоційного самоконтролю та навичок ефективного подолання стресових ситуацій.

У роботі «Family and Academic Stress and Their Impact on Student's Depression and Academic Performance» авторами Nguyen T., Pham H., Le V. проаналізовано взаємозв'язки між двома типами стресу – академічним і сімейним, та їх вплив на депресію і навчальні результати студентів. Для дослідження залучили 520 студентів із різних факультетів, які пройшли опитування, в якому оцінювали рівень стресу, симптоми депресії та середній бал успішності (рис. 1.6) [9].

Hypothesis	Constructs	Estimate	Critical ratio	p-value	Decision
H ₁	Acad. Strs → Std. Dev. Lev	0.201	2.021	0.039*	Accepted
H ₂	Fam. Strs → Std. Dev. Lev	0.358	3.997	0.001*	Accepted
H ₃	Std. Dev. Lev → Std. Acad. Perf	-0.319	-3.402	0.001*	Accepted

Рисунок 1.6 – Взаємозв'язок між стресом, депресією та академічною успішністю

Результати показали, що як академічний, так і сімейний стрес позитивно корелюють із рівнем депресії. У свою чергу, депресія виступає посередником і чинить негативний вплив на навчальні результати студентів. Прямий вплив стресу на академічну успішність виявився менш значущим, що підкреслює роль психічного стану в освітньому процесі.

У статті «Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups and COVID-19» авторами Smith J., Lee A. було проаналізовано, як різні чинники навчального навантаження впливають на психічний стан студентів. Дослідження охопило 750 студентів. Які оцінювали

інтенсивність академічного навантаження, фінансові труднощі, наявність соціальної підтримки та психоемоційні показники. Досліджено зв'язок між рівнем психічного благополуччя студентів та їхнім академічним стресом (рис 1.7) [10].

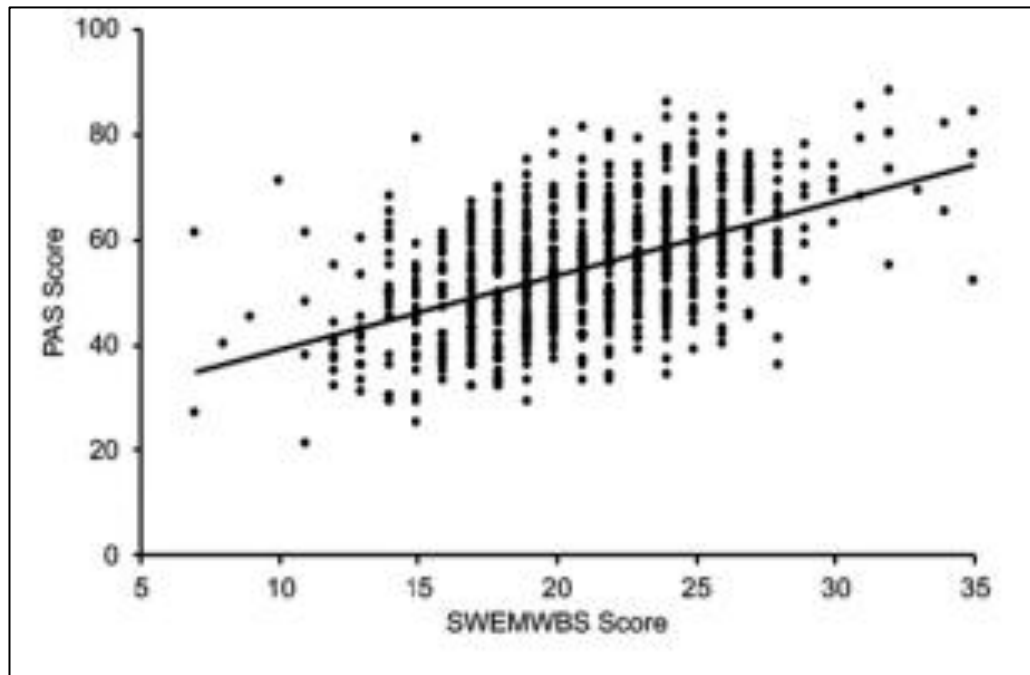


Рисунок 1.7 – Графік взаємозв'язків між SWEMWBS і PAS-показниками серед студентів

Графік ілюструє що, студенти з вищим загальним психоемоційним тонусом також демонструють більшу сприйнятливості до академічного навантаження.

Найсильнішим негативним фактором виявилось надмірне навчальне навантаження, яке значно знижувало рівень психічного здоров'я. COVID -19 також вплинув на академічну успішність студентів (рис. 1.8).

	No change/neutral response/ reduced stress		Increased stress	
	<i>n</i>	%	<i>n</i>	%
Gender				
Female	118	17.8	544	82.2
Male	47	33.3	94	66.7
Nonbinary	0	0	40	100
Race/ethnicity				
White or Caucasian	104	18.6	456	81.4
Black or African American	16	24.2	50	75.8
East Asian	20	25.6	58	74.4
Hispanic, Latino, or of Spanish origin	14	18.9	60	81.1
Middle Eastern, Native American, Alaskan Native, South Asian, other, or prefer not to answer	11	16.9	54	83.1
Year of study				
1	38	28.4	96	71.6
2	43	18.5	190	81.5
3	39	15.5	212	84.5
4	45	20	180	80

Рисунок 1. 8 – Таблиця впливу COVID-19 на рівень стресу студентів

Це дослідження продемонструвало, що студенти не однаково зазнають впливу академічного стресу чи стресу пов'язаного з пандемією. І що існують значні відмінності на рівні груп у психічному благополуччі. Зокрема, студенти другого курсу зазнали впливу академічного стресу та низьке психологічне благополуччя порівняно зі студентами інших років навчання.

В дослідженні «Stress, student burnout and study engagement – a cross sectional comparison of university students of different academic subjects» авторами Martiez L., Kim S. було проведено аналіз зв'язків між рівнем стресу, симптомами емоційного вигорання та ступенем залученості студентів до навчального процесу. У дослідженні взяли участь 1200 студентів, серед яких виявили, що 73,2% регулярно відчувають високий або помірний рівень стресу, тоді як близько третини мають симптоми вигорання, включаючи втрату мотивації та емоційне виснаження (рис 1.9) [11].

	Information Technology (I)	Engineering (E)	Medicine (M)	Sport and Health Sciences (S)	Economic Sciences (ES)	F	p	Eta ²	Pairwise Tukey comparison
	MW ± SD	MW ± SD	MW ± SD	MW ± SD	MW ± SD				
Exhaustion (Welch)	3.1 ± 1.61	2.8 ± 1.43	2.4 ± 1.56	2.3 ± 1.36	2.6 ± 1.57	9.117	<0.001	0.061	I > M; I > S; E > S
Cynicism (Welch)	1.3 ± 1.59	1.2 ± 1.48	0.6 ± 1.05	1.0 ± 1.27	1.1 ± 1.24	7.537	<0.001	0.052	M < I; M < E; M < S; M < ES
Reduced academic efficiency (Welch)	2.3 ± 1.71	2.3 ± 1.56	1.6 ± 1.41	1.6 ± 1.31	2.2 ± 1.54	11.827	<0.001	0.074	M < I; M < E; S < I; S < M; S < ES
Study Engagement	3.2 ± 1.09	3.2 ± 1.05	3.8 ± 1.01	3.3 ± 0.93	3.1 ± 0.87	14.426	<0.001	0.058	M > I; M > E; M > S; M > ES

Рисунок 1.9 – Рівень стресу, вигорання та залученості у навчання серед студентів

Студенти медицини найменш схильні до вигорання та зниженню академічної активності, і водночас демонструють найвищий рівень залученості до навчання. Студенти інформаційних технологій, інженерії, економічних наук мають вищий рівень вигорання та зниження академічної успішності, тоді як залученість до навчання є меншою. Студенти спеціальностей у сфері спорту та наук про здоров'я демонструють низькі показники вигорання та зниженої академічної ефективності. Автори пропонують запровадити програми емоційної підтримки та тренінги зі стресостійкості для попередження хронічного вигорання.

У роботі «Стресостійкість студентів закладів вищої освіти як психологічна проблема» авторка Світлана Валеріївна Пухно проаналізувала рівень стресу студентів першого курсу, акцентуючи увагу на етапі адаптації до нових умов навчання, зміни життєвого укладу та зростання вимог до самостійності. Опитано 67 студентів першого курсу гуманітарного та технічного профілю. Більшість студентів продемонстрували середній або дещо вищий за середній рівень стресостійкості, що свідчить про помірну адаптацію до освітнього навантаження. Низький рівень стресостійкості проявляється у чутливості до дії стресових факторів, тривалості емоційних переживань та ризику хронічного стресу. Натомість студенти з вищими показниками характеризуються більшою емоційною стабільністю та здатністю швидко відновлюватися після негативних впливів.

Ключову роль у формуванні адаптивних механізмів відіграє соціальне середовище. У СумДП ім. А.С. Макаренка акцент робиться на створенні комфортного психологічного клімату, активній взаємодії між курсами, участі студентів у волонтерських та громадських ініціативах. Це сприяє розвитку комунікативних навичок, зменшенню рівня стресу та підвищенню загальної стресостійкості [12].

1.3 Аналіз методів прогнозування стресу у студентів

Мова програмування Python пропонує широкий набір методів, для виконання прогнозування завдань різної складності, що адаптуються відповідно до поставлених цілей. Оскільки основним завданням роботи є прогнозування рівня стресу студентів закладів вищої освіти, потрібно обрати методи прогнозування, які знайдуть найбільш точний результат. Розглянемо найбільш поширені методи прогнозування в середовищі Python:

1. Метод k-найближчих сусідів (KNeighbors) – це метод класифікації об'єктів на основі найближчих навчальних прикладів у просторі ознак, відстань з найменшим значенням відповідає зразку і навчальний набір, найближчому до невідомого зразка (рис. 1.5) [13].

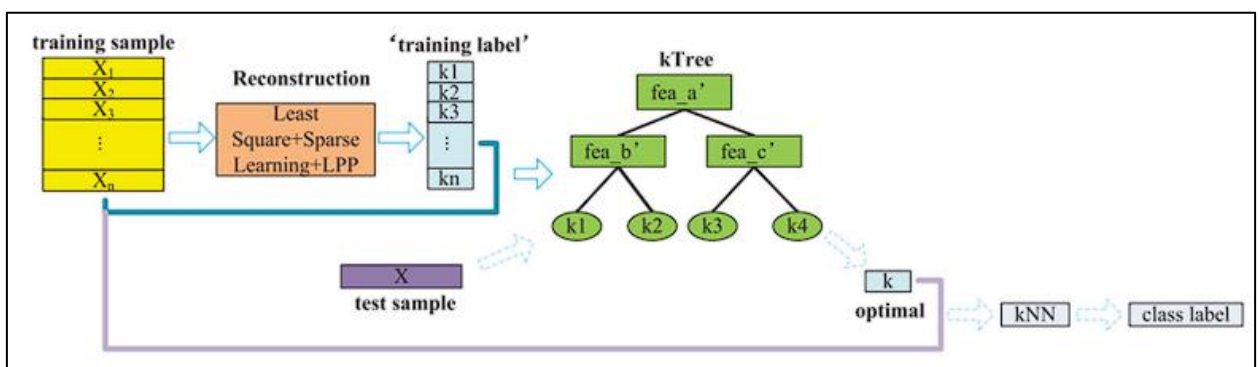


Рисунок 1.5 – Блок-схема методу KNN

2. Лінійна регресія (Linear Regression) – це один з найпростіших методів регресії, що відповідає на питання, чи впливає одне явище на інше і як, або як пов’язані між собою кілька змінних (рис. 1.6) [14].

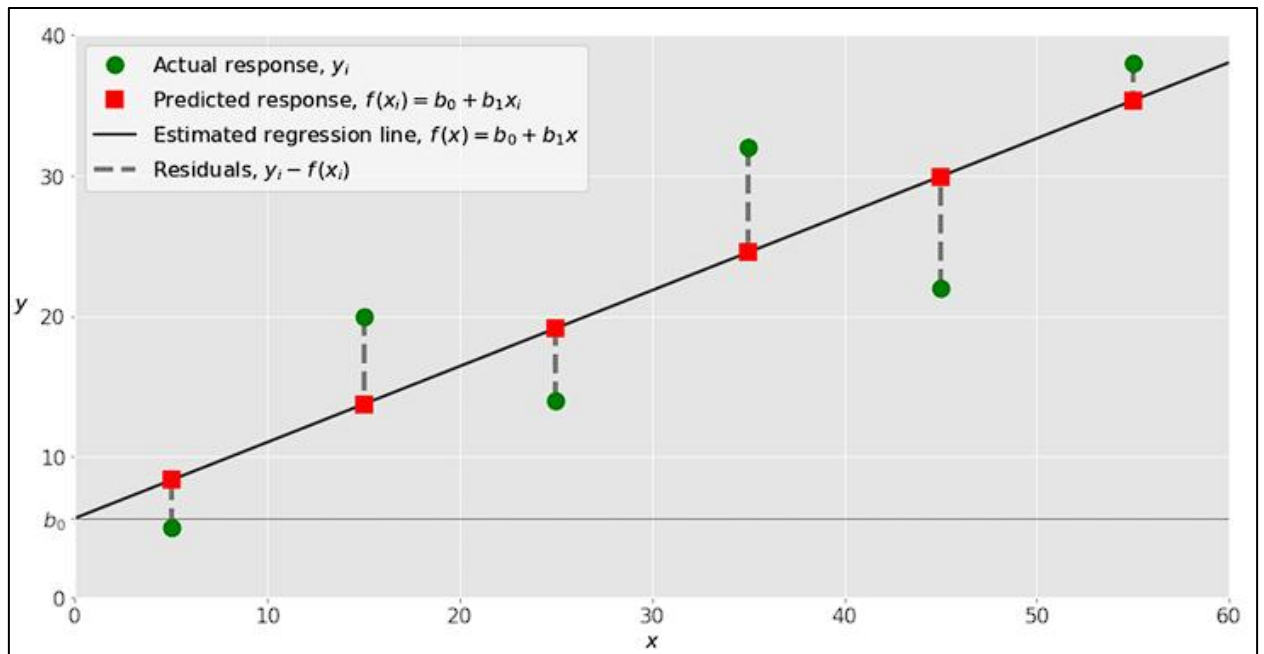


Рисунок 1.6 – Приклад лінійної регресії

3. Дерево рішень (Decision Tree) – представляє собою графічну модель, яка слугує інструментом для ухвалення рішень у різних сферах. Мета дерева прийняття рішень полягає у створенні моделі, що відтворює процес ухвалення рішень та дозволяє прогнозувати цільову зміну. Дерева рішень використовуються для прогнозування різних параметрів і характеристик даних (рис. 1.7) [15].

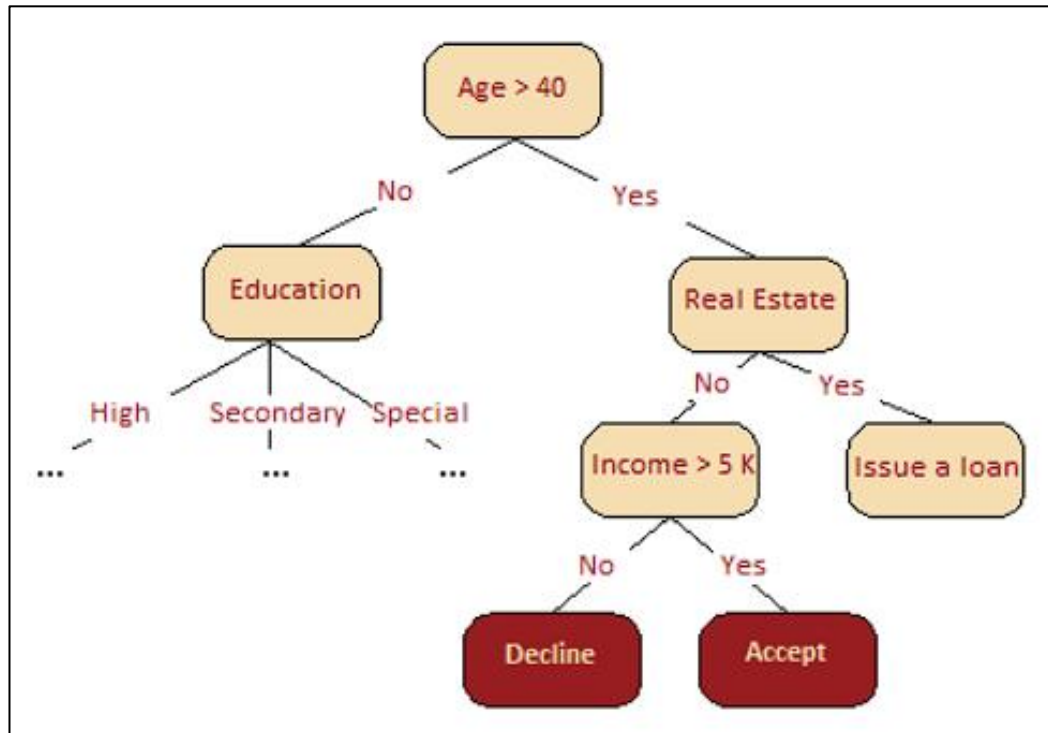


Рисунок 1. 7 – Приклад метода дерева рішень

4. Метод опорних векторів (Support Vector Machine) – це метод перетворення вхідних векторів у простір більшої розмірності, чим більша різниця або відстань паралельними гіперплощини, тим менша середньоквадратична похибка класифікації (рис. 1.8) [16].

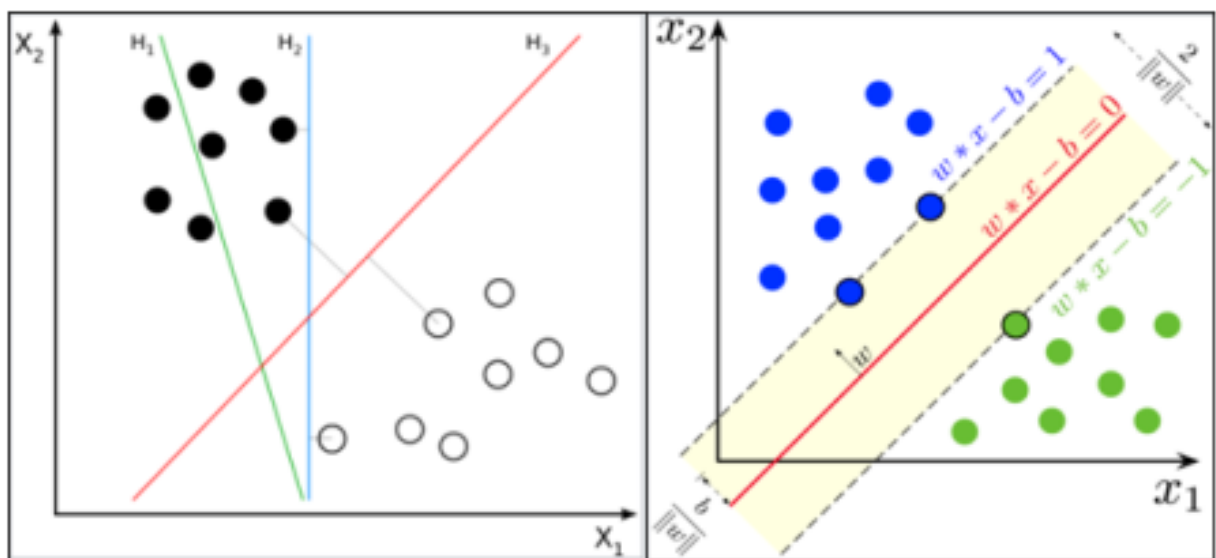


Рисунок 1.8 – Приклад метода опорних векторів

5. Випадковий ліс (Random Forest) – метод прийняття рішень, що представлений у вигляді дерева, функціонує на основі різних умов. RF складається з набору дерев рішень, утворюючи «ліс» класифікаторів, що обирають певний клас. Для навчання потрібні два параметри: кількість дерев та кількість випадкових ознак, а також базу даних з мітками класі. Метод є легким для розуміння і легкий в обробці відсутніх значень, без необхідності перетворення даних (рис. 1.9) [17].

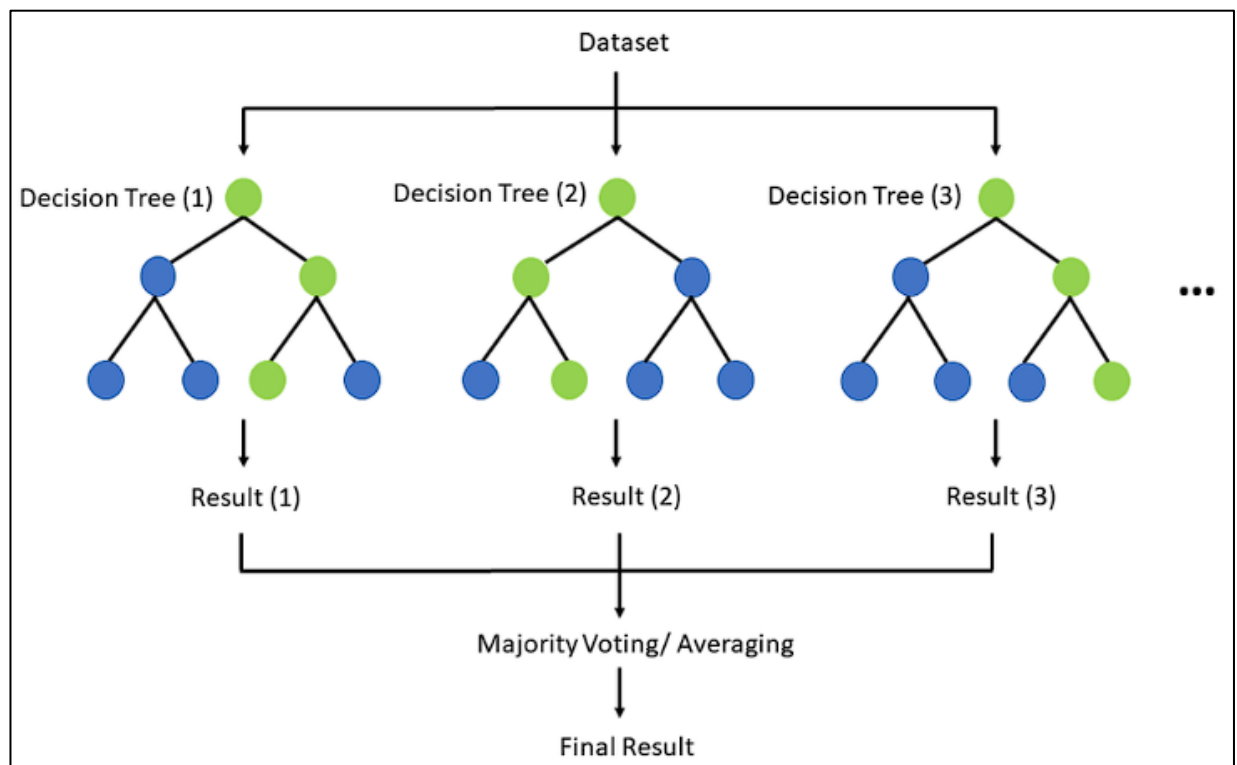


Рисунок 1. 9 – Приклад метода RandomForest

6. XGBoost – метод градієнтного бустингу, алгоритм має високу прогнозну здатність, крім того вона включає різноманітні методи регуляризації, що дозволяють знизити ризик перенавчання та підвищити загальну ефективність. Серед переваг цього підходу варто виділити його високу точність та можливість різноманітного використання регуляції (рис. 1.10) [18].

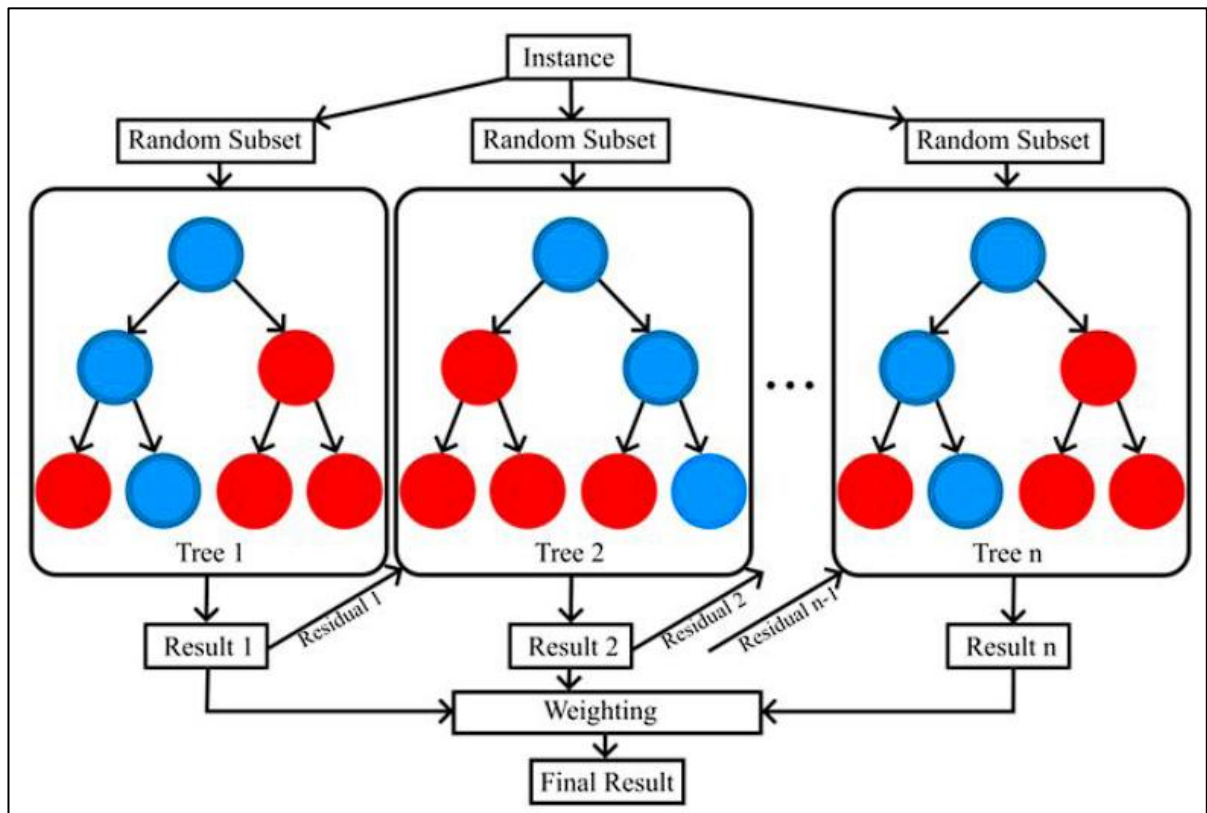


Рисунок 1.10 – Блок-схема методу XGBoost

Отже, здійснивши аналіз різних методів машинного навчання, для створення інформаційної технології прогнозування стресу у студентів обрано наступні: дерево рішень, випадковий ліс; k-найближчих сусідів.

1.4 Висновки

У цьому розділі було проведено аналіз предметної області прогнозування стресу студентів закладів вищої освіти та визначено актуальність теми. Розглянуто основні фактори, що впливають на рівень стресу студентів, а саме: академічний стрес, соціально-економічний, психологічний та особистісний стрес.

Проаналізовано близько 10 наукових публікацій на тематику прогнозування стресу у студентів. Аналіз вітчизняних та зарубіжних досліджень з проблеми стресу у студентів свідчить про високу актуальність цієї тематики в контексті сучасної освіти. Результати досліджень охоплюють

різні аспекти стресу – його причини, форми прояву, наслідки та зв'язок із навчальним середовищем. У більшості досліджень основними причинами стресу визнано академічне навантаження, тиск через оцінювання, відсутність балансу між навчанням і особистим життям, а також фінансові труднощі. Декілька авторів у своїх дослідженнях, а саме «College students' learning stress, psychological resilience and learning burnout: status quo and coping strategies» та «Family and Academic Stress and Their Impact on Student's Depression and Academic Performance» вказують на сильний зв'язок між стресом і симптомами депресії та емоційного вигорання, на їх думку, стрес виступає опосередкованим чинником, що впливає на зниження мотивації та академічну успішність. Дослідження українських науковців Кондратенко Є. М., Церковнонь О. В. та Пухно С. В. підкреслюють специфічний вплив воєнних дій, пандемії та адаптаційних труднощів на психоемоційний стан студентів. Дослідження авторів Rais Ahmad, Smith J., Kaur K. демонструють, що стать, курс навчання, спеціальність та рівень доходу можуть впливати на рівень стресу, а також на його сприйняття. Згідно дослідження Gong Z., психологічна стійкість, емоційний самоконтроль, підтримка родини та соціальних груп визнані факторами, що знижують негативний вплив стресу. Таким чином, аналіз показує, що стрес є типовим і системним явищем у студентському середовищі, особливо в умовах зростаючого академічного навантаження, соціально-економічної нестабільності та впливу глобальних подій (пандемія, війна).

Також проведено аналіз методів машинного навчання для створення інформаційної технології прогнозування стресу у студентів та обрано наступні: дерево рішень, випадковий ліс; k-найближчих сусідів.

2. ВИБІР ОПТИМАЛЬНОГО РІШЕННЯ ТА НАЛАШТУВАНЬ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

2.1 Аналіз середовища та мови програмування для розробки інформаційної технології прогнозування стресу у студентів

Головною метою цієї роботи є створення інформаційної технології для прогнозування рівня стресу студентів закладів вищої освіти. Для вирішення такого типу задач доцільно обрати мову програмування Python.

Python – це проста та мінімалістична мова програмування, програму можна розглядати як набір послідовних команд або алгоритмів, які об'єкт повинен виконати задля досягнення поставленої мети. Характеризується чітким логічним синтаксисом, що сприяє легкому читанню коду [19].

Ця мова підтримує різноманітні стилі програмування, такі як структурний та об'єктно-орієнтований. Також можливе використання інших підходів. Завдяки своїй гнучкості, Python дозволяє інтегрувати модульні компоненти, створені на інших мовах програмування. Наприклад, можна написати програму мовою C++ і імпортувати її в Python у вигляді модуля[20].

Python-скрипти можна запускати на різних операційних системах, включаючи: Windows, Linux, Amigo, Mac OS та інші. Окрім великої стандартної бібліотеки, вбудованої у Python, є широкий спектр безкоштовних бібліотек. Ці бібліотеки можуть бути з легкістю імпортовані та встановлені, що значно розширює функціональність середовища програмування.

Де використовують Python [20]:

1. Розробка веб-додатків;
2. Створення консольних програм;
3. Розробка мобільних додатків;
4. Машинне навчання;
5. Технології розпізнання мови;
6. Штучний інтелект.

Розглянемо також популярність машинного навчання на Python. За останні 20 років на різних мовах було створено та реалізовано багато різних алгоритмів машинного навчання. Завдяки своїй інтерактивній природі високого рівня та розвиненій екосистемі бібліотек, Python є привабливим вибором для розробки алгоритмів та аналізу даних. Бібліотека Scikit-learn використовує багатий інструментарій для навчання сучасних реалізацій алгоритмів машинного навчання. Вона одночасно зберігає інтуїтивно зрозумілий інтерфейс, тісно пов'язаний з Python, що робить її зручною та ефективною у використанні. Scikit-learn виділяється серед інших інструментів машинного навчання через забезпечення високої продуктивності в ньому використовується скомпільований код, він залежить лише від numpy та scipy, що полегшує його встановлення та використання [21].

Scikit-Learn пропонує широкий вибір алгоритмів для виконання задач класифікації, основні з них [21]:

- Метод k-найближчих сусідів (K-nearest Neighbors);
- Метод опорних векторів (Support Vector Machines);
- Класифікатор дерева рішень (Decision Tree Classifier)
- Випадковий ліс (RandomForest);
- Лінійний дискримінантний аналіз (Linear Discriminant Analysis);
- Логістична регресія (Logistic Regression).

Підсумовуючи, можна зазначити, що Python завдяки своїй простоті, гнучкості та великому набору бібліотек є відмінним вибором для виконання задачі прогнозування рівня стресу студентів закладів вищої освіти.

Для проведення дослідження було обрано онлайн-платформу Kaggle, яка спеціалізується на машинному навчанні та науці про дані. Kaggle – платформа для змагань з прогнозування, моделювання та аналітики. Ці навички, мають значну практичну користь для вирішення будь-яких реальних завдань, проведення розвідувального аналізу сприяє виявленню проблемних місць [22]. Jupyter Notebook – допомагає писати та виконувати код на платформі Kaggle,

це пришвидшує процес моделювання моделей та візуалізації даних, що нам необхідно для виконання завдання.

2.2 Розвідувальний аналіз даних

Для проведення аналізу було обрано набір даних у Kaggle, що має назву «Student Stress Factors: A Comprehensive Analysis» та опубліковано користувачем «Chhabii». Датасет має відкритий доступ для загального використання на платформі Kaggle [23]. Датасет містить дані про фактори стресу студентів, а саме: психологічні, фізіологічні, соціальні, фактори навколишнього середовища, академічні фактори, загально датасет складається з 21 ознаки, 1100 рядків (рис.2.1).

Інформація про датасет:						
Кількість рядків і стовпців: (1100, 21)						
Перші 5 рядків даних:						
	anxiety_level	self_esteem	mental_health_history	depression	headache	\
0	14	20	0	11	2	
1	15	8	1	15	5	
2	12	18	1	14	2	
3	16	12	1	15	4	
4	16	28	0	7	2	
Останні 5 рядків даних:						
	anxiety_level	self_esteem	mental_health_history	depression	headache	\
1095	11	17	0	14	3	
1096	9	12	0	8	0	
1097	4	26	0	3	1	
1098	21	0	1	19	5	
1099	18	6	1	15	3	

Рисунок 2.1 – Дані датасету «Student Stress Factors: A Comprehensive Analysis»

Розглянемо повний список ознак, які є у нашому наборі даних та виконаємо опис. Датасет включає в себе такі дані:

- Anxiety_level: рівень тривоги;
- Self_esteem: самооцінка;
- Mental_health_history: історія психічного здоров'я;
- Depression: депресія;
- Headache: головний біль;
- Blood_pressure: кров'яний тиск;
- Sleep_quality: якість сну;
- Breathing_problem: проблеми з диханням;
- Noise_level: рівень шуму;
- Living_conditions: умови життя;
- Safety: безпека;
- Basic_needs: базові потреби;
- Academic_performance: успішність;
- Study_load: навчальне навантаження;
- Teacher_student_relationship: стосунки між викладачем та студентом;
- Future_career_concerns: майбутня кар'єра;
- Social_support: соціальна підтримка;
- Peer_pressure: тиск з боку однолітків;
- Extracurricular_activities: позакласні заняття;
- Bullying: булінг;
- Stress_level: рівень стресу (цільова ознака).

Наступним кроком є виявлення аномалії у даних. Для цього використаємо функцію `describe()`, щоб візуально визначити можливі аномалії та визначити їхнє місцезнаходження у датасеті (рис.2.2-2.3).

```
| print("\nОписова статистика:")
| print(data.describe())
```

Рисунок 2.2 – Функція `describe()`

Описова статистика:					
	anxiety_level	self_esteem	mental_health_history	depression	\
count	1100.000000	1100.000000	1100.000000	1100.000000	
mean	11.063636	17.777273	0.492727	12.555455	
std	6.117558	8.944599	0.500175	7.727008	
min	0.000000	0.000000	0.000000	0.000000	
25%	6.000000	11.000000	0.000000	6.000000	
50%	11.000000	19.000000	0.000000	12.000000	
75%	16.000000	26.000000	1.000000	19.000000	
max	21.000000	30.000000	1.000000	27.000000	

	headache	blood_pressure	sleep_quality	breathing_problem	\
count	1100.000000	1100.000000	1100.000000	1100.000000	
mean	2.508182	2.181818	2.660000	2.753636	
std	1.409356	0.833575	1.548383	1.400713	
min	0.000000	1.000000	0.000000	0.000000	
25%	1.000000	1.000000	1.000000	2.000000	
50%	3.000000	2.000000	2.500000	3.000000	
75%	3.000000	3.000000	4.000000	4.000000	
max	5.000000	3.000000	5.000000	5.000000	

	noise_level	living_conditions	...	basic_needs	academic_performance	\
count	1100.000000	1100.000000	...	1100.000000	1100.000000	
mean	2.649091	2.518182	...	2.772727	2.772727	
std	1.328127	1.119208	...	1.433761	1.414594	
min	0.000000	0.000000	...	0.000000	0.000000	
25%	2.000000	2.000000	...	2.000000	2.000000	
50%	3.000000	2.000000	...	3.000000	2.000000	
75%	3.000000	3.000000	...	4.000000	4.000000	

Рисунок 2.3 – Результат виконання функції describe()

За результатами функції describe() візуально аномальних значень на цьому етапі не виявлено, мінімальне категоріальне значення є 1, тоді як максимальне значення не перевищує 5.

Змінна mental_health_history має більшість значень 0, змінна basic_needs та academic_performance мають значення 4 та 5.

Опис результатів:

1. Anxiety_level: середній рівень тривожності становить 11.06, що вказує на помірний рівень тривожності серед студентів
2. Self_esteem: середня самооцінка є відносно високою 17.77, проте значення різні що говорить, про різний рівень впевненості.
3. Mental_health_history: середнє значення 0.49, що вказує на наявність проблем з психічним здоров'ям.

4. Depression: середній рівень депресії 12.55, що може свідчити про наявність депресивних симптомів студентів.

5. Headache: середній рівень головного болю 2.50, студенти відчують іноді але не постійно відчувають головний біль.

6. Blood_pressure: середнє значення 2.18, що означає нормальний кров'яний тиск.

7. Sleep_quality: середнє значення 2.67, якість сну є ближче до задовільної.

8. Breathing_problem: середнє значення 2.75, доволі часті проблеми з диханням.

Далі потрібно перевірити дані на наявність пропущених значень, щоб врахувати яких значень забагато та не мають ніякої цінності для прогнозування (рис. 2.4).

Перевірка наявності пропущених значень:	
anxiety_level	0
self_esteem	0
mental_health_history	0
depression	0
headache	0
blood_pressure	0
sleep_quality	0
breathing_problem	0
noise_level	0
living_conditions	0
safety	0
basic_needs	0
academic_performance	0
study_load	0
teacher_student_relationship	0
future_career_concerns	0
social_support	0
peer_pressure	0
extracurricular_activities	0
bullying	0
stress_level	0
dtype: int64	

Рисунок 2.4 – Результат перевірки на пропущені значення

За результатами функції `info()` можна зробити висновки, що у наборі даних немає пустих значень, тобто усі параметри мають сто відсотків заповнення для кожного фактору стресу студентів.

Наступний крок, визначення кореляції між параметрами, та побудова кореляційної матриці. Для цього використовуємо функцію `corr()`.

На рисунку 2.8 представлено код функції `corr()`.

```
def correlation_analysis(data):
    correlation_matrix = data.corr()
```

Рисунок 2.8 – Функція `corr()`

На рисунку 2.9 представлено результат значень кореляції між параметрами.

```
Кореляція між рівнем стресу та іншими факторами:
stress_level          1.000000
bullying              0.751162
future_career_concerns 0.742619
anxiety_level         0.736795
depression            0.734379
headache              0.713484
extracurricular_activities 0.692977
peer_pressure         0.690684
noise_level           0.663371
mental_health_history  0.648644
study_load            0.634156
breathing_problem     0.573984
blood_pressure        0.394200
living_conditions     -0.581723
social_support        -0.632497
teacher_student_relationship -0.680163
basic_needs           -0.708968
safety               -0.709602
academic_performance -0.720922
sleep_quality         -0.749068
self_esteem           -0.756195
Name: stress_level, dtype: float64
```

Рисунок 2.9 – Результат значень кореляції між параметрами

Кореляційна матриця є ключовим статистичним методом для кількісного аналізу лінійних зв'язків між змінними. Вона представляє собою таблицю попарних кореляцій, де кожна клітинка містить коефіцієнт у діапазоні від -1 до +1. Додатні значення вказують на пряму залежність, від'ємні – на обернену, а нуль – на відсутність лінійної кореляції [24]. Отже, побудуємо кореляційну матрицю, щоб візуалізувати зв'язки між різними ознаками у наборі даних (рис 2.10).

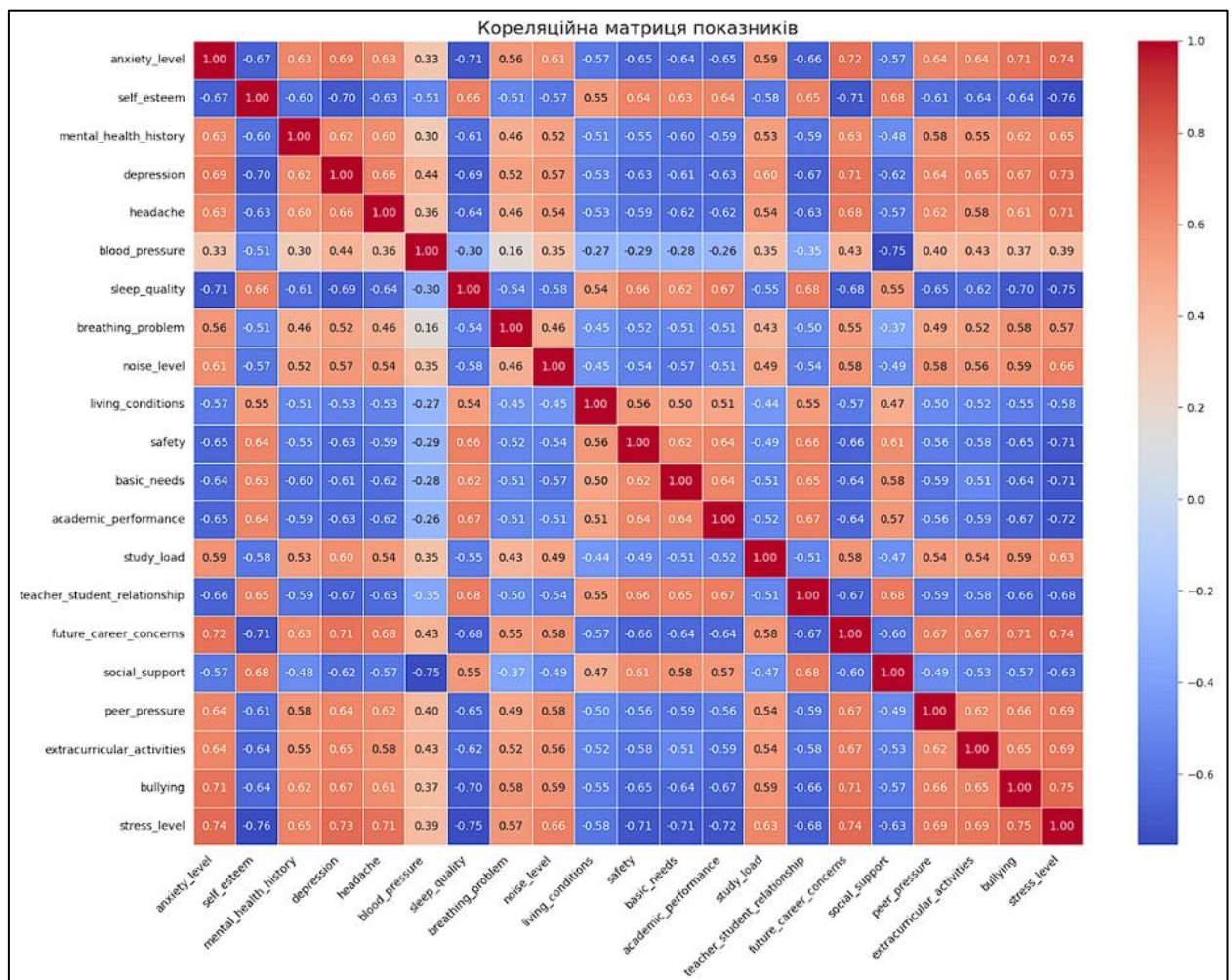


Рисунок 2.10 – Кореляційна матриця

Кожен елемент матриці містить значення коефіцієнта кореляції Пірсона, змінюючись від -1 до +1. Коефіцієнт кореляції Пірсона використовується для змінних на шкалі інтервалів які перебувають у лінійній залежності. Коефіцієнт кореляції позначається малою літерою r і його значення від 0 до -1 вказує на

негативну кореляцію, тобто зростання однієї змінної пропорційне зменшенню значення іншої змінної [25].

Отже, аналізуючи нашу кореляційну матрицю, виділяємо важливі взаємозв'язки між показниками:

1. `Anxiety_level`: рівень тривоги дорівнює значенню $+0,74$, що є сильним прямим зв'язком.
2. `Depression`: депресія дорівнює значенню $+0,71$, що свідчить про сильний прямий зв'язок.
3. `Bullying`: булінг дорівнює значенню $+0,69$, що вказує на те, що булінг є одним із факторів стресу студентів.
4. `Self_esteem`: самооцінка дорівнює значенню $-0,76$ що означає, чим вища самооцінка студента, тим нижчий рівень стресу і так само в протилежну сторону.
5. `Academic_performance`: успішність корелює на $-0,71$, що говорить про те, що високий рівень стресу негативно впливає на концентрацію та мотивацію, що призводить до зниження успішності.
6. `Mental_health_history`: історія психічного здоров'я корелює на $+0,63$, що підкреслює вразливість студентів з психічними проблемами до стресу.

Кореляційна матриця ілюструє складні взаємозв'язки між різними факторами життя студентів та рівнем стресу. Отже, найбільш впливовими факторами, що збільшують ймовірність виникнення стресу, є: тривога, депресія, булінг, проблеми зі сном, фінансові труднощі та занепокоєння щодо майбутньої кар'єри. У той час, такі аспекти, як висока самооцінка, хороша академічна успішність та позитивні взаємини з викладачами, понижують рівень стресу. Виявлені кореляції будуть використані для створення моделі машинного навчання, що допоможе точніше прогнозувати рівень стресу студентів.

Далі проведемо аналіз розподілу даних відносно факторів у датасеті. На рисунку 2.11 зображено графік розподілу `anxiety_level`.

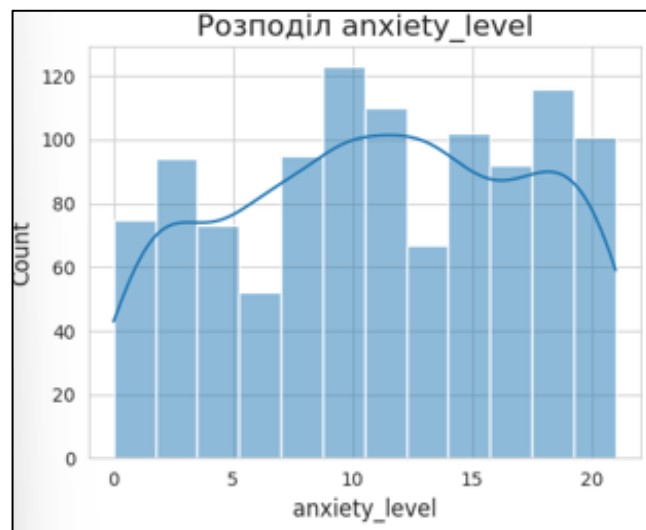


Рисунок 2.11 – Графік розподілу даних відповідно до рівня тривоги

Аналізуючи наведений графік, ми можемо побачити, що значна частина студентів демонструє середній рівень тривожності. Однак, варто звернути увагу на те, що існує також суттєва кількість студентів, які мають високий показник тривоги. Це підкреслює важливість та актуальність проблем психічного здоров'я студентів.

Побудуємо подібний графік до попереднього, але уже на основі фактору Self_esteem самооцінка (рис. 2.12).

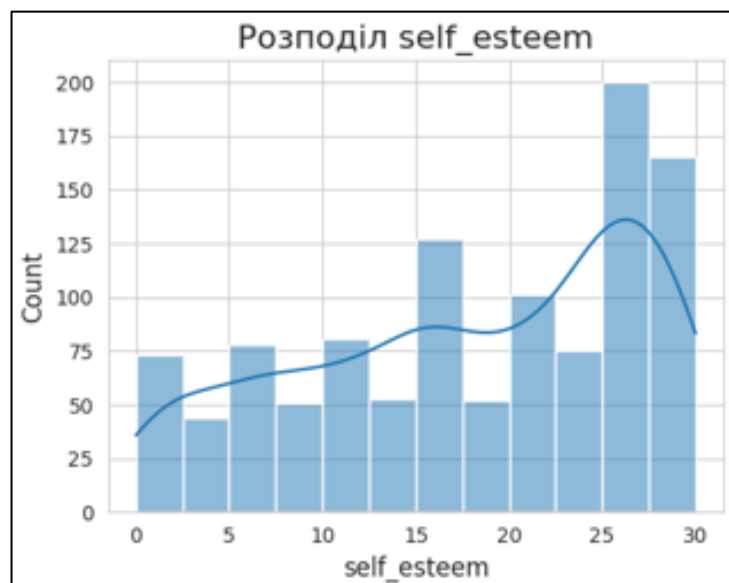


Рисунок 2.12 – Графік розподілу даних відповідно до рівня самооцінки

З графіку видно, що значна частина студентів має відносно високий рівень самооцінки. Але, наявність окремих груп з нижчою самооцінкою вказує на потребу в психологічній підтримці студентів.

Наступним кроком є побудова графіка, на основі фактору Depression (депресії) (рис. 2.13).

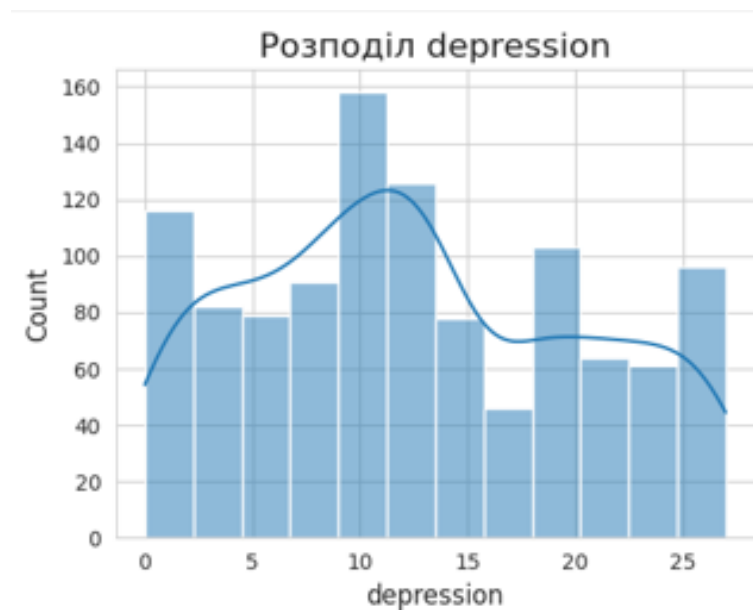


Рисунок 2.13 – Графік розподілу даних відповідно до рівня депресії

Аналізуючи результати візуалізації рівня депресії можемо бачити, що значна частина студентів, страждає від помірного рівня депресії, накладена крива, вказує на три основних максимумів, що говорить про різний рівень депресії студентів.

Далі побудуємо графік із рівнем розподілу даних якості сну (рис. 2.14).

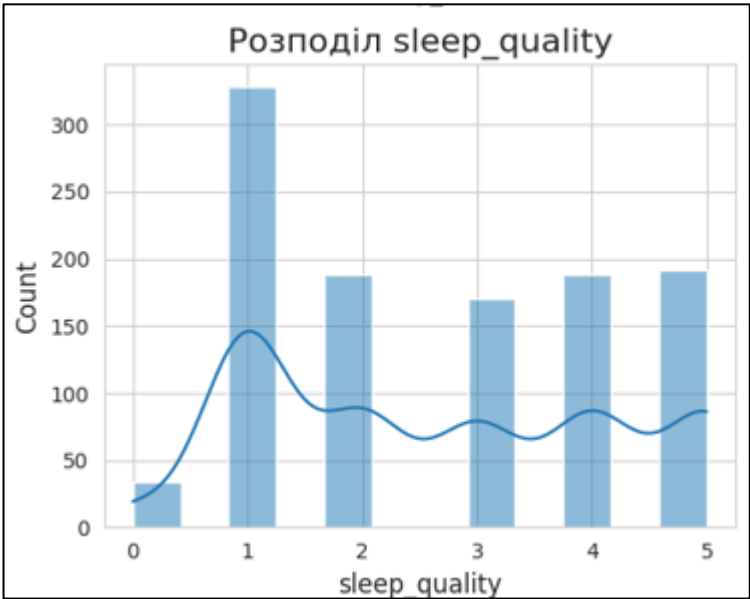


Рисунок 2.14 – Графік розподілу даних відповідно до рівня якості сну

На графіку зображено дані рівня якості сну студентів, понад 300 студентів оцінюють свою якість сну як «1». Це вказує на те, що велика кількість студентів має проблеми зі сном.

Наступний графік розподілу даних рівня академічної успішності (рис. 2.15).

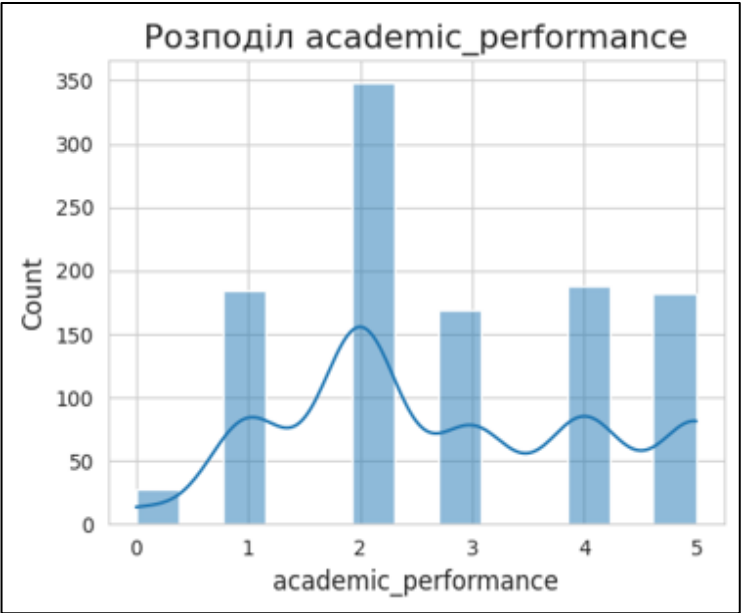


Рисунок 2.15 – Графік розподілу даних відповідно до рівня академічної успішності

Згідно даними графіку, близько 350 студентів мають оцінку успішності «2», що свідчить про середній рівень академічної успішності. Цей показник є важливим, він демонструє, що студенти потребують додаткової підтримки та уваги з боку навчального закладу.

Будуємо графік розподілу даних до рівня майбутньої кар'єри (рис. 2.16).

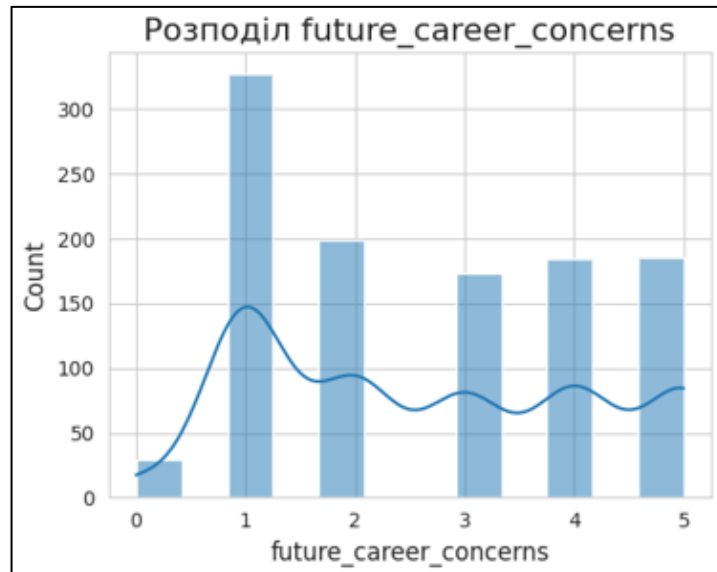


Рисунок 2.16 – Графік розподілу даних відповідно до рівня майбутньої кар'єри

У представленому графіку, близько 320 студентів оцінюють свій рівень занепокоєння майбутньою кар'єрою на рівні «2». Це свідчить про те, що певна кількість студентів демонструє мінімальний, але помітний ступінь тривоги щодо цього питання.

Візуалізуємо графік розподілу даних до рівня соціальної підтримки (рис. 2.17).

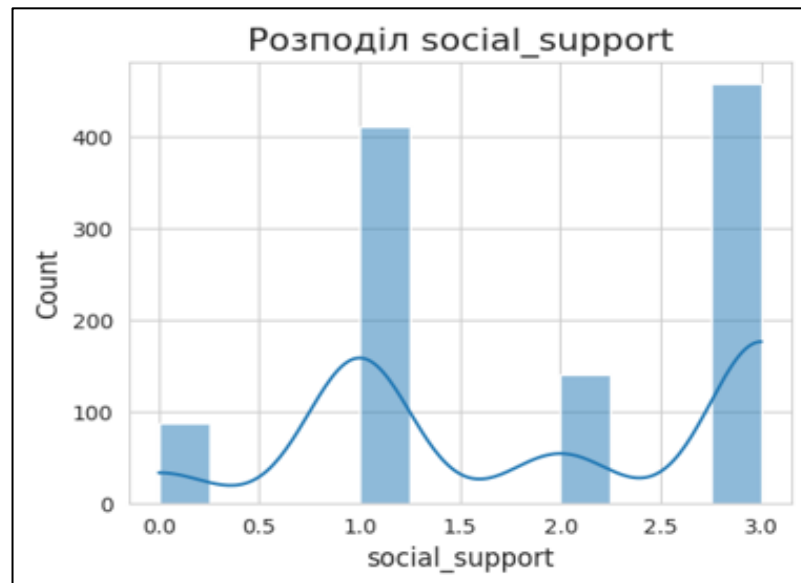


Рисунок 2.17 – Графік розподілу даних відповідно до рівня соціальної підтримки

На графіку відтворено дані, близько 400 студентів оцінюють свій рівень соціальної підтримки як «1» що означає помірний рівень підтримки. Але понад 450 студентів мають високий рівень соціальної підтримки «3», що означає про достатню кількість соціальної підтримки.

Проведемо детальніший аналіз графіків, що ілюструють вплив окремих факторів на рівень стресу студентів. Ці візуалізації дають змогу оцінити, які саме змінні мають найбільшу вагу при класифікації рівня стресу та виявити закономірності їх взаємозв'язку.

На рисунку 2.18 зображено важливість ознак академічного чинника

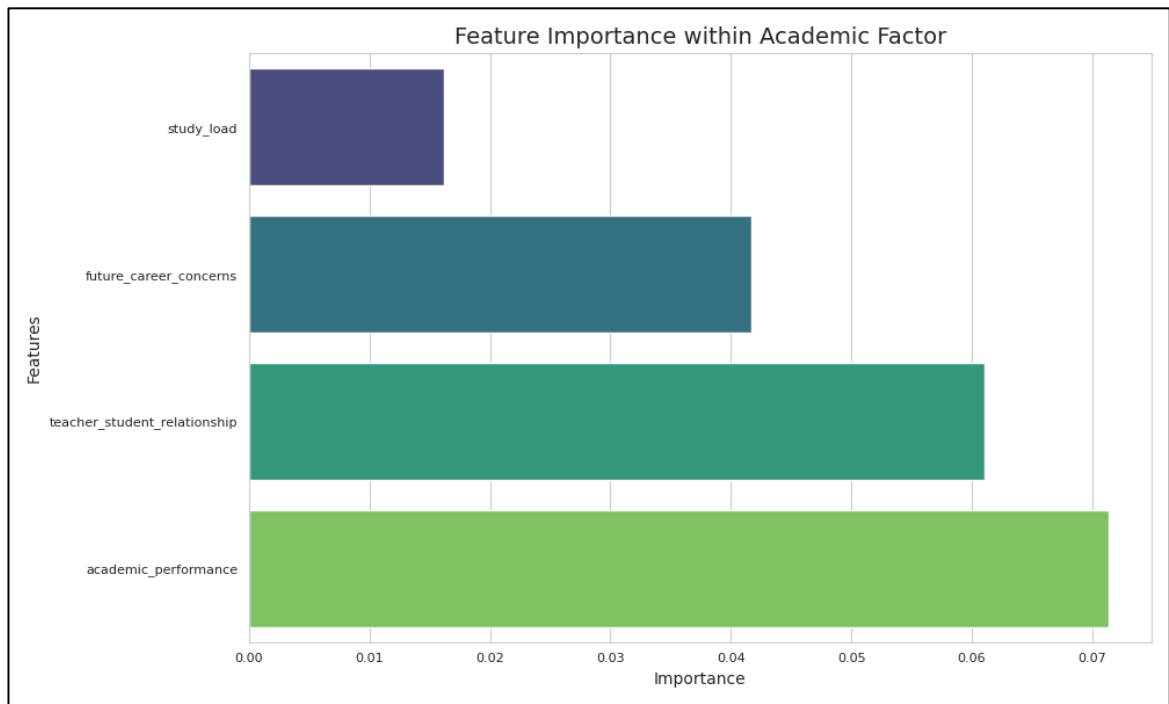


Рисунок 2.18 – Важливість ознак академічного чинника

Аналіз діаграми дозволяє зробити висновок, що серед академічних факторів найбільший вплив на рівень стресу студентів має саме успішність (academic performance). Це свідчить про тісний взаємозв'язок між оцінками студента та його емоційним станом: низькі результати навчання можуть провокувати відчуття тривоги, невпевненості, страху перед майбутнім.

Другу позицію за впливом посідає якість взаємодії між викладачем та студентом (teacher-student relationship). Позитивне або, навпаки, напружене спілкування з викладачами суттєво впливає на рівень психологічного комфорту в навчальному середовищі.

Ознака занепокоєння щодо майбутньої кар'єри також має значущий вплив, адже невизначеність у професійному напрямку є типовим джерелом стресу серед молоді. Найменший вплив у межах академічного чинника має навчальне навантаження, що, ймовірно, пояснюється адаптацією студентів до постійної роботи з великою кількістю інформації.

На рисунку 2.19 зображено важливість ознак навколишнього середовища.

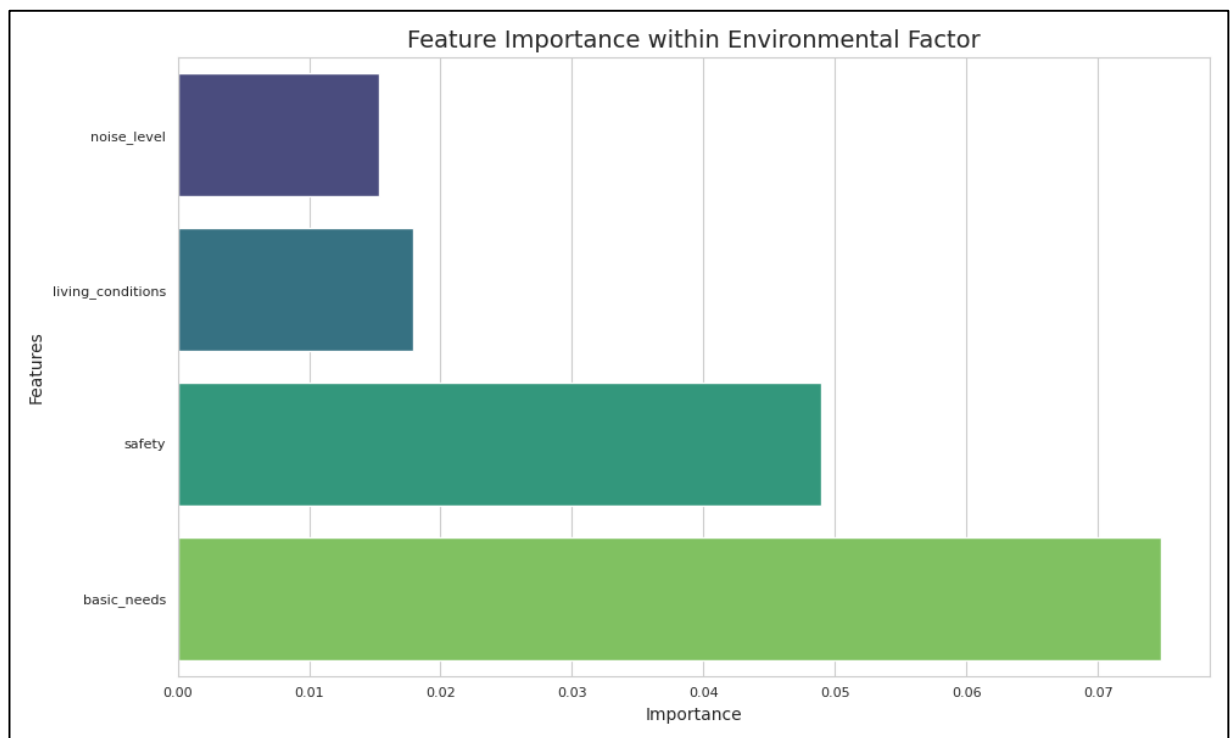


Рисунок 2.19 – Важливість ознак чинника навколишнього середовища

Найвищу вагу в цій групі чинників має означник базових потреб (basic needs), з показником важливості понад 0,7. Такий високий рівень свідчить про критичну роль матеріального благополуччя — доступу до житла, харчування, комфорту — у формуванні або посиленні стресових реакцій.

Другим за значущістю фактором є відчуття безпеки (safety) — його вплив пояснюється як загальним психоемоційним фоном, пов’язаним з умовами життя (зокрема у період воєнного стану), так і відчуттям вразливості перед зовнішніми обставинами.

Менший вплив мають такі чинники, як умови проживання та рівень шуму, проте й вони формують контекст, у якому перебуває студент, і здатні опосередковано впливати на стрес.

На рисунку 2.20 зображено важливість ознак фізіологічних чинників.

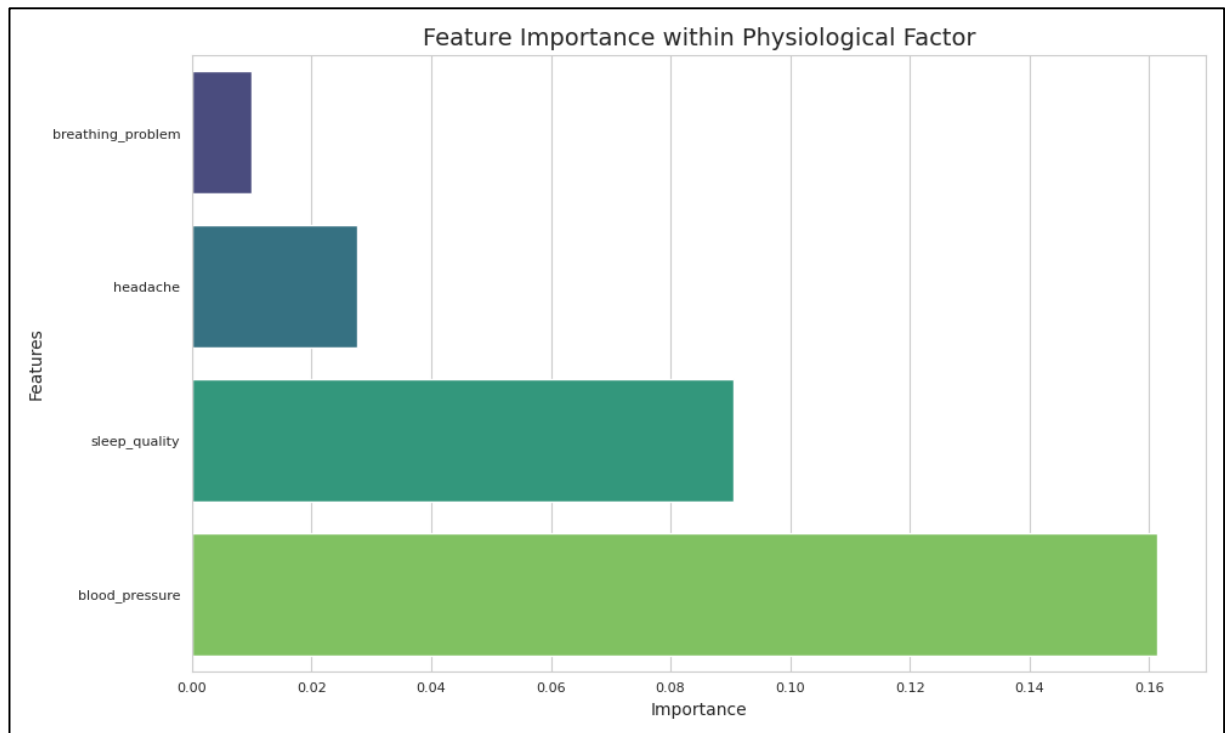


Рисунок 2.20 – Важливість ознак фізіологічних чинників

Найбільш значущим показником фізіологічного типу є кров'яний тиск (blood pressure), що підтверджує наявність прямого зв'язку між фізичним здоров'ям та психоемоційним станом.

Показник якості сну (sleep quality) посідає другу позицію: численні дослідження свідчать, що розлади сну не тільки є симптомами стресу, але й самі по собі підсилюють його прояви.

Головний біль (headache) і проблеми з диханням (breathing problems) мають меншу вагу, але також вказують на те, що навіть незначні фізичні недуги здатні посилювати загальне навантаження на нервову систему.

На рисунку 2.21 зображено важливість ознак психологічних чинників.

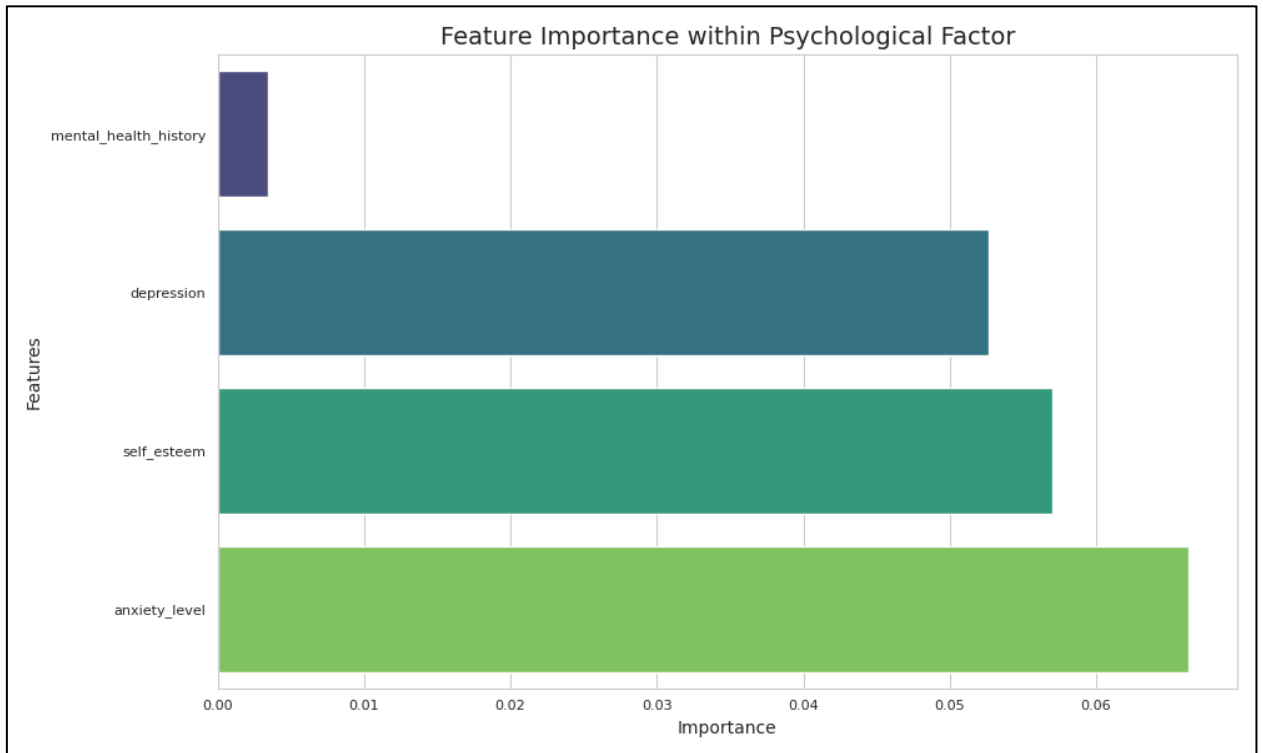


Рисунок 2.21 – Важливість ознак психологічних чинників

У групі психологічних змінних рівень тривожності (anxiety level) демонструє найбільший вплив на стрес. Це передбачувано, адже тривога — це основний маркер емоційного виснаження.

Самооцінка (self-esteem) виступає інверсно пропорційним фактором: чим вона вища, тим менший ризик виникнення стресу.

Депресивні прояви (depression) також відіграють суттєву роль — наявність навіть помірних симптомів значно підвищує ймовірність деструктивної реакції на зовнішні обставини.

На рисунку 2.22 зображено важливість ознак соціальних чинників.

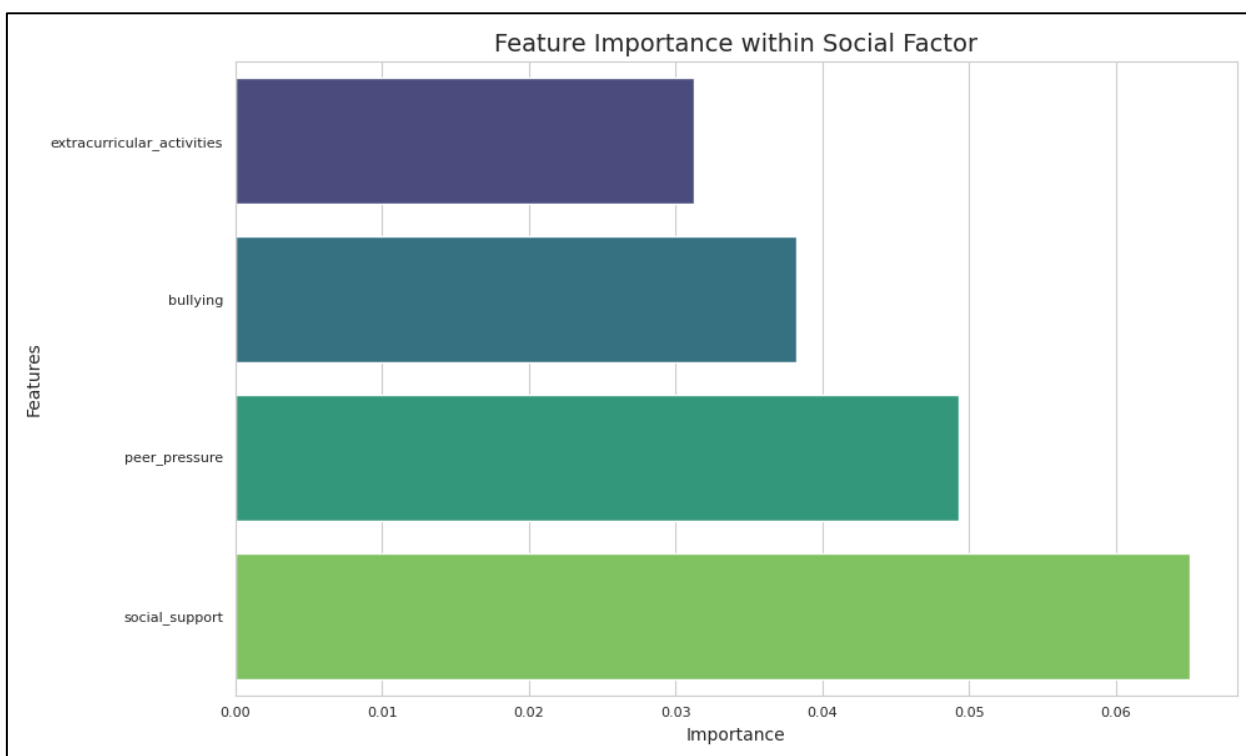


Рисунок 2.22 – Важливість ознак соціальних факторів

Соціальна сфера також має прямий вплив на рівень стресу. Найвпливовішим чинником виявлено соціальну підтримку (social support). Її відсутність — або відчуття ізоляції — здатні багаторазово підсилювати стресові реакції.

Тиск з боку однолітків (peer pressure) демонструє значний негативний вплив, особливо в умовах академічної конкуренції чи соціального порівняння.

Булінг (bullying) теж є важливим предиктором підвищеного рівня стресу, навіть якщо він має нижчий показник важливості.

Позанавчальна активність (extracurricular activities), навпаки, відіграє нейтральну або навіть компенсаторну роль, і тому демонструє найменший вплив у цій групі. Також, особливу увагу було приділено кореляції кожної ознаки з цільовою змінною stress_level.

На рисунку 2.23 візуалізовано взаємозв'язок між рівнем стресу та топ-5 факторами, що мають найвищу кореляцію з ним за абсолютним значенням, що допомагає визначити найбільш значущі предиктори.

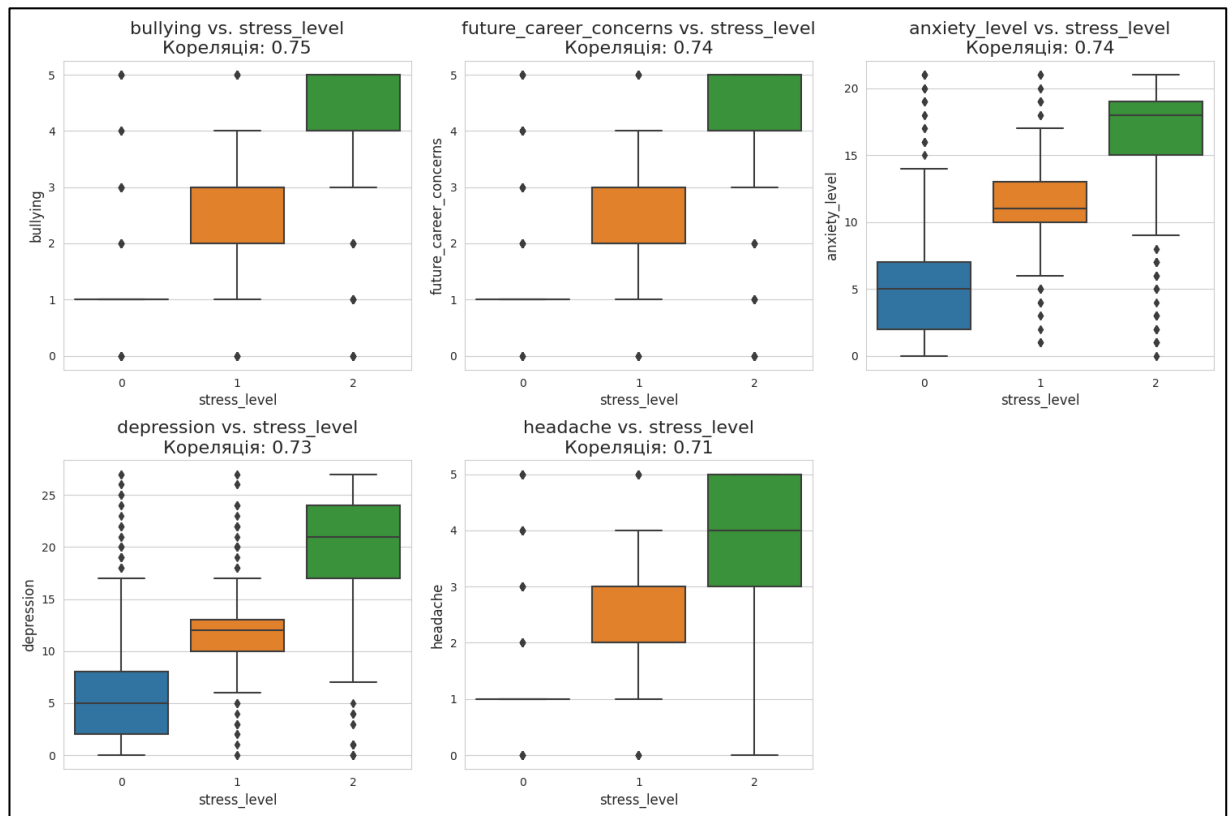


Рисунок 2.23 – Взаємозв'язок рівня стресу з найбільш корелюючими факторами

Для дослідження взаємозв'язку між рівнем стресу студентів та факторами було побудовано серію boxplot-графіків. Вони демонструють розподіл значень для різних груп респондентів залежно від рівня стресу (0 – низький, 1 – середній, 2 – високий). Кореляційні коефіцієнти вказують на ступінь зв'язку між цими змінними.

1) Bullying and stress level – кореляція: 0,75. Сильний позитивний зв'язок між рівнем булінгу та рівнем стресу. Студенти з високим рівнем стресу частіше зазнають булінг. Це свідчить про те, що соціальне середовище й міжособистісні конфлікти суттєво впливають на емоційне благополуччя.

2) Future career concerns and stress level – кореляція: 0,74. Показники свідчать про майже такий самий рівень залежності, як і при булінгу. Студенти, які мають високий рівень тривожності щодо майбутнього працевлаштування, демонструють підвищений рівень стресу.

3) Anxiety level and stress level – кореляція: 0,74. Високий рівень тривожності суттєво корелює з підвищеним рівнем стресу. У групі з високим стресом більшість студентів мають значення тривожності понад середній рівень. Це підтверджує теоретичне припущення про тісний зв'язок тривожних розладів зі стресовими станами.

4) Depression and stress level – кореляція: 0,73. Кореляційний зв'язок на рівні 0,73 демонструє значну асоціацію між рівнем депресивних симптомів та стресом. У групі з високим рівнем стресу переважають студенти з високими показниками депресії.

5) Headache and stress level – кореляція: 0,71. Фізичні прояви, такі як головні болі, також мають суттєвий позитивний зв'язок із рівнем стресу. У групі з високим стресом спостерігається збільшення середніх показників головного болю. Це вказує на психоматичну природу деяких симптомів, викликаних хронічним напруженням.

Усі п'ять факторів мають сильну позитивну кореляцію з рівнем стресу, що свідчить про їх суттєвий вплив. Найвищий зв'язок спостерігається між стресом і булінгом, а також тривожністю щодо майбутнього. Це дозволяє зробити висновок про необхідність комплексної підтримки студентів не лише на психологічному, а й на соціальному рівні.

Аналіз цільової змінної (stress_level) був наступним важливим кроком.

На рисунку 2.24 представлено код аналізу цільової змінної.

```

# Аналіз цільової змінної
def target_analysis(data):
    # Перевірка унікальних значень цільової змінної
    print("\nУнікальні значення рівня стресу:")
    print(data['stress_level'].value_counts().sort_index())

    # Гістограма розподілу стресу
    plt.figure(figsize=(10, 6))
    sns.countplot(data=data, x='stress_level')
    plt.title('Розподіл рівня стресу серед студентів (класи)')
    plt.xlabel('Рівень стресу (клас)')
    plt.ylabel('Кількість студентів')
    plt.savefig('stress_distribution.png')
    plt.close()

    # Кругова діаграма розподілу рівнів стресу
    plt.figure(figsize=(10, 8))
    stress_counts = data['stress_level'].value_counts()
    stress_percentages = (stress_counts / len(data) * 100).round(1)

    plt.pie(stress_percentages, labels=[f'Рівень {level}: {pct}%'
                                       for level, pct in zip(stress_counts.index, stress_percentages)],
            autopct='%1.1f%%', startangle=90)
    plt.title('Кількісний розподіл рівнів стресу (%)')
    plt.axis('equal')
    plt.savefig('stress_distribution_pie.png')
    plt.close()

```

Рисунок 2.24 – Аналіз цільової змінної

Цільовою змінною в даному дослідженні є `stress_level`, яка відображає рівень стресу студента і представлена у вигляді класів (наприклад, 0 – низький, 1 – середній, 2 – високий). Було детально проаналізовано розподіл студентів за цими рівнями стресу.

На рисунку 2.25 зображено результат унікального значення рівня стресу.

```

Унікальні значення рівня стресу:
stress_level
0      373
1      358
2      369
Name: count, dtype: int64

```

Рисунок 2.24 – Унікальні значення рівня стресу

На рисунку 2.25 представлено гістограму розподілу рівнів стресу, що наочно демонструє кількість студентів у кожному класі.

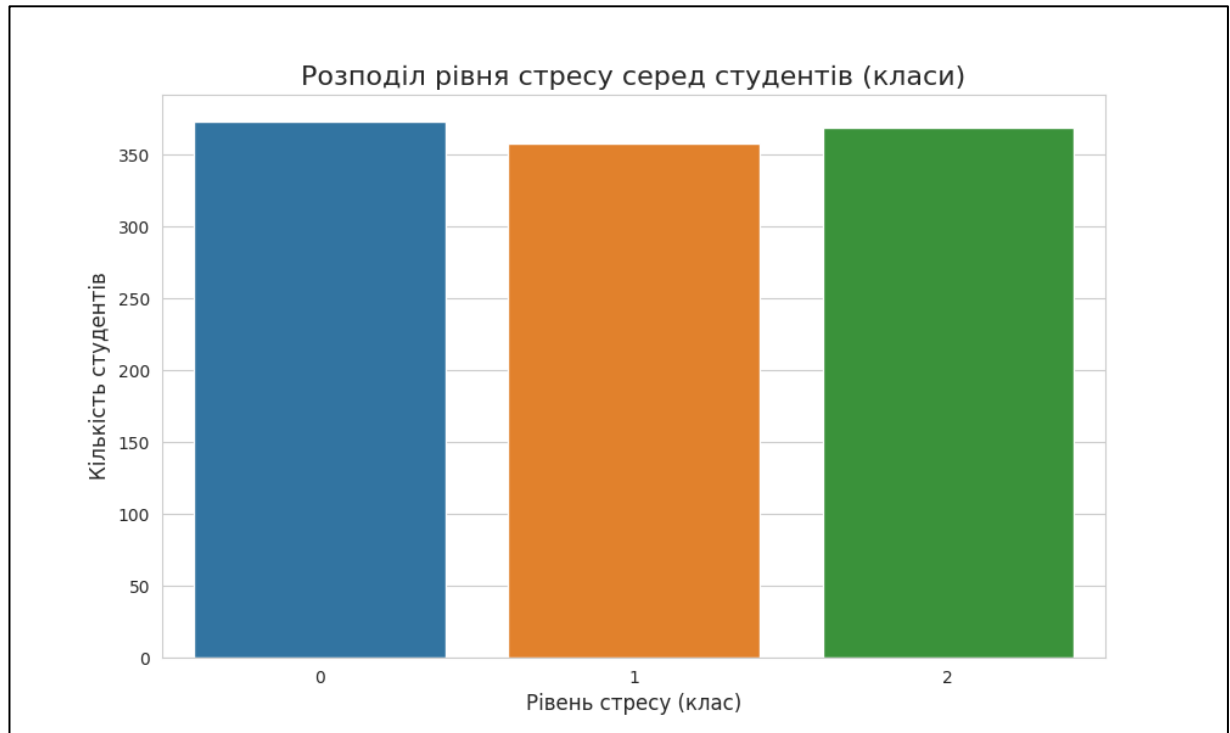


Рисунок 2.25 – Розподіл рівня стресу серед студентів

Для кращого розуміння кількісного співвідношення, на рисунках 2.26 показано код та на рисунку 2.27 зображено відсоткову кругову діаграму, яка відображає кількісний розподіл унікальних значеннях параметру `stress_level` у відсотках.

```
plt.figure(figsize=(10, 8))
stress_counts = data['stress_level'].value_counts()
stress_percentages = (stress_counts / len(data) * 100).round(1)

plt.pie(stress_percentages, labels=[f'Рівень {level}: {pct}%'
                                   for level, pct in zip(stress_counts.index, stress_percentages)]
        autopct='%1.1f%%', startangle=90)
plt.title('Кількісний розподіл рівнів стресу (%)')
plt.axis('equal')
plt.savefig('stress_distribution_pie.png')
plt.close()
```

Рисунок 2. 26 – Код розподілу рівнів стресу у відсотках за допомогою кругової діаграми

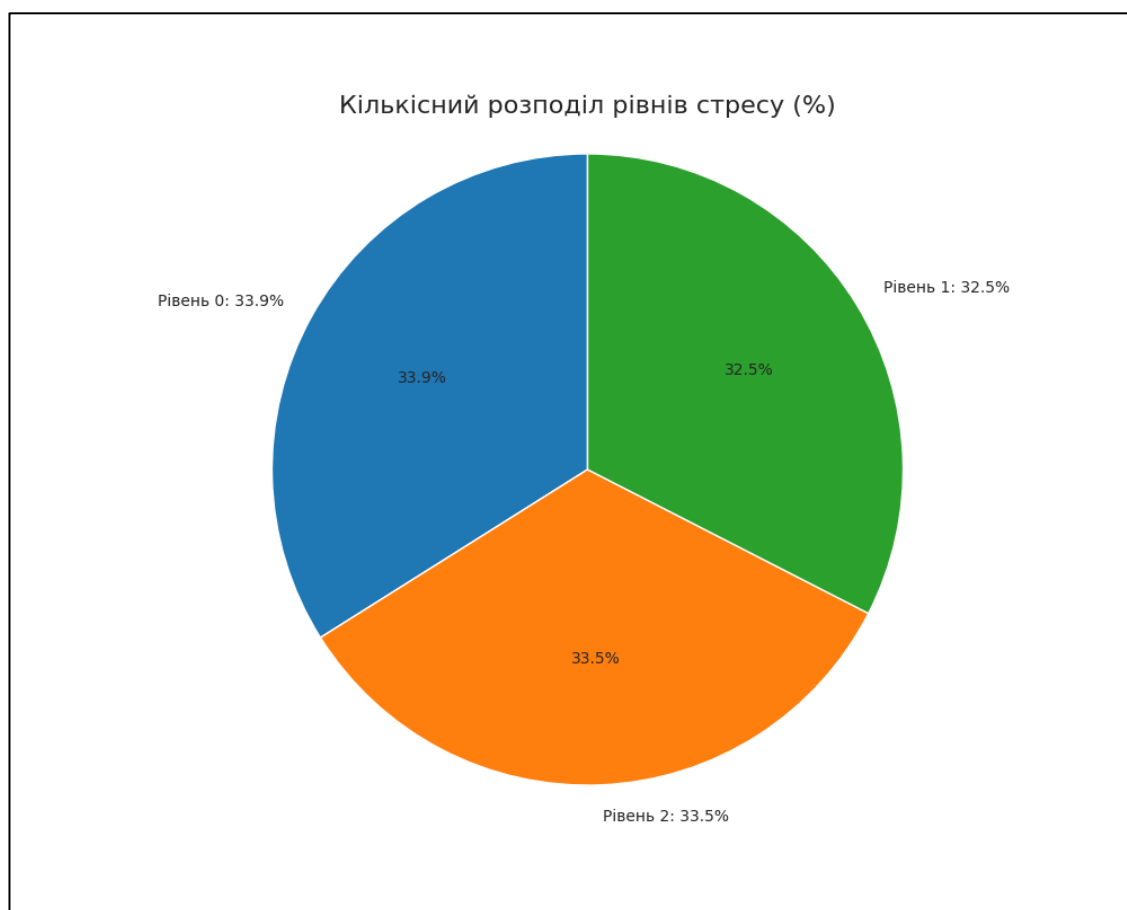


Рисунок 2.27 – Кількісний розподіл рівнів стресу у відсотках

Отже, згідно з результатами аналізу кількісних значень параметра рівня стресу, можна стверджувати, що розподіл між унікальними значеннями рівня стресу студентів закладів вищої освіти, представлений у цьому датасеті, є рівномірним. Для перевірки результатів дослідження при побудові моделей машинного навчання будемо використовувати метрику `F1_Score`, адже маємо задачу класифікації. Остаточна оцінка моделей буде залежати саме від цієї метрики.

2.3 Висновки

У цьому підрозділі було здійснено розвідувальний аналіз даних: перевірено набір на пропущені значення, проведено статистичний опис ознак, побудовано кореляційну матрицю та визначено найбільш впливові фактори стресу. Аналіз показав, що найбільш впливовими факторами, що збільшують

ймовірність виникнення стресу, є: тривога, депресія, булінг, проблеми зі сном, фінансові труднощі та занепокоєння щодо майбутньої кар'єри. У той час, такі аспекти, як висока самооцінка, хороша академічна успішність та позитивні взаємини з викладачами, понижують рівень стресу.

Для проведення дослідження було обрано мову програмування Python, яка завдяки своїй простоті, гнучкості та великому набору бібліотек є відмінним вибором для виконання задачі прогнозування рівня стресу студентів закладів вищої освіти. В якості середовища розробки було обрано онлайн-платформу Kaggle, яка спеціалізується на машинному навчанні та науці про дані. Також виконано візуалізацію розподілу ключових змінних та проаналізовано взаємозв'язки між ними і рівнем стресу. За результатами аналізу кількісних значень параметра рівня стресу, було виявлено, що розподіл між унікальними значеннями рівня стресу студентів закладів вищої освіти, представлений у цьому датасеті, є рівномірним. Отриманні результати стали основною для подальшого навчання моделей машинного навчання.

3. РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ СТРЕСУ У СТУДЕНТІВ

3.1 Розроблення інформаційної технології

Для розробки інформаційної технології, спрямованої на аналіз та прогнозування рівня стресу серед студентів закладів вищої освіти, необхідно створити комплекс методів машинного навчання. Ці методи повинні працювати з попередньо підготовленими даними, на яких вони навчатимуться та проходити тестування. Для цієї задачі було вирішено обрати наступні методи: Random Forest, Decision Tree, K-nearest Neighbors.

Щоб ці моделі працювали ефективно, для кожної з них потрібно визначити набір параметрів, за якими буде відбуватися налаштування та оптимізація. Для підбору найбільш результативної комбінації параметрів доцільно застосувати метод GridSearchCV, який поєднує простоту реалізації та ефективність. Якщо отриманні результати не відповідатимуть очікуваній точності, налаштування параметрів буде змінюватися до досягнення задовільних показників.

Після отриманих результатів, необхідно здійснити їх порівняльний аналіз і вибрати найкращу з них для подальшого використання. Обрану модель слід додатково проаналізувати з точки зору важливості окремих ознак, тобто оцінити, наскільки кожна характеристика вхідних даних вплинула на прийняття рішень у процесі класифікації рівня стресу.

На рисунку 3.1 представлено блок-схему алгоритму роботи інформаційної технології прогнозування стресу студентів.

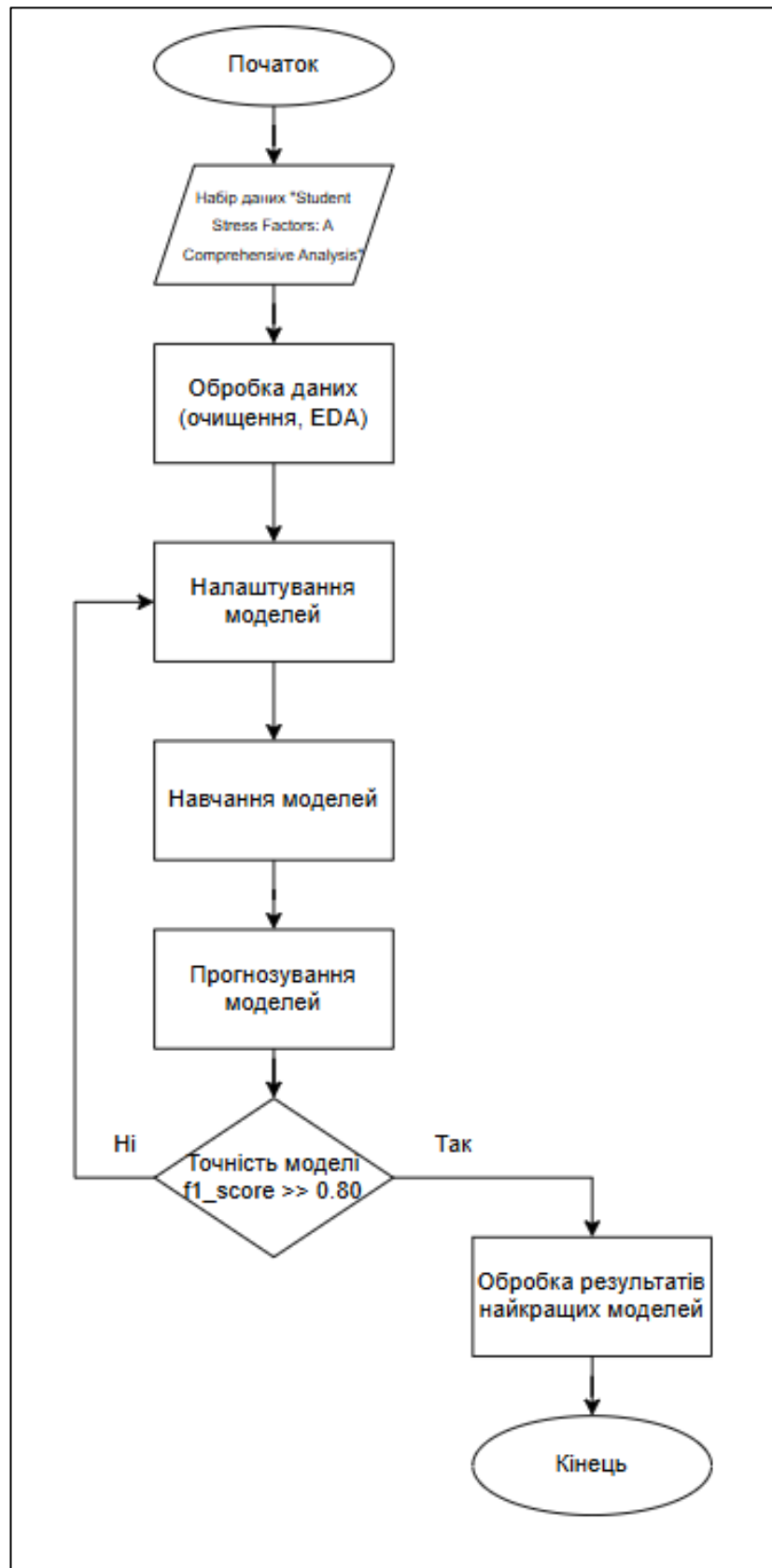


Рисунок 3.1 – Блок-схема алгоритму роботи інформаційної технології прогнозування стресу у студентів

Наведена блок-схема ілюструє алгоритм функціонування інформаційної технології прогнозування стресу у студентів закладів вищої освіти. Вона описує етап обробки даних, які надалі використовуються для навчання моделей із попереднім налаштуванням параметрів. Після завершення навчання моделі застосовуються для прогнозування, під час якого оцінюється їх точність. Якщо точність моделі не перевищує 80% модель повторно направляється на навчання, доки не буде досягнуто прийнятного результату. У результаті визначається найефективніша модель, після чого здійснюється аналіз важливості ознак, що впливають на її роботу.

Для опису функціональності розробленої інформаційної технології прогнозування рівня стресу студентів було використано нотацію UML (Unified Modeling Language), зокрема Use Case діаграму, що дозволяє візуалізувати сценарії взаємодії користувачів із системою.

На рисунку 3.2 представлено UML Use Case діаграма функціонування інформаційної технології прогнозування рівня стресу, на якій в якості ключових учасників (акторів) - студента, дослідника та систему машинного навчання, а також перелік основних функціональних можливостей.

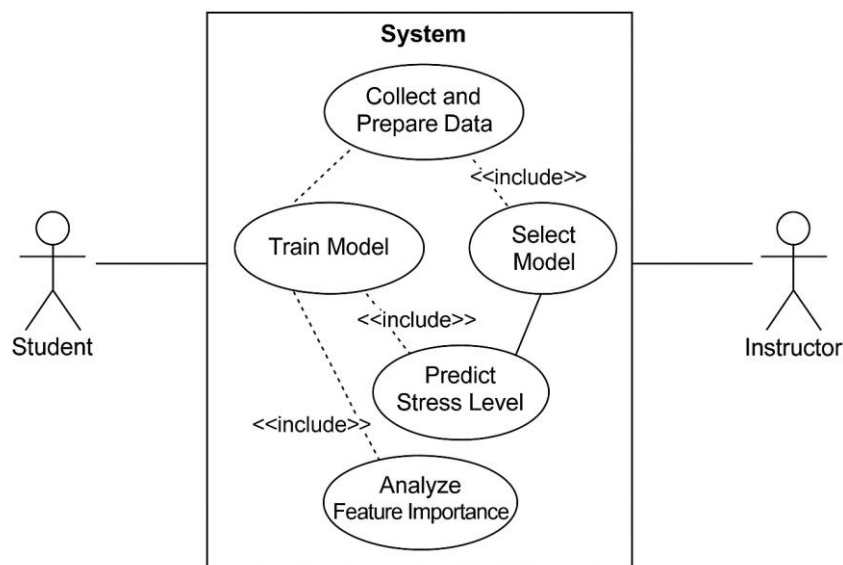


Рисунок 3.2 – UML Use Case діаграма функціонування інформаційної технології прогнозування рівня стресу

До основних варіантів використання належить:

- Надання вхідних даних – дослідник надає інформацію про студентів, необхідну для подальшої обробки;
- Попередня обробка даних – відбувається перевірка, очищення та нормалізація даних;
- Тренування моделі – навчання моделі на основі зібраного датасету;
- Прогнозування рівня стресу – система, використовує алгоритми машинного навчання, генерує прогноз рівня стресу;
- Аналіз важливості ознак – система аналізує, які фактори найбільше вплинули на результат прогнозування.

Для візуалізації архітектури програмного забезпечення, що реалізує аналіз факторів стресу студентів із використанням моделей машинного навчання, було побудовано UML-діаграму компонентів.

На рисунку 3.3 зображено UML-діаграму компонентів.

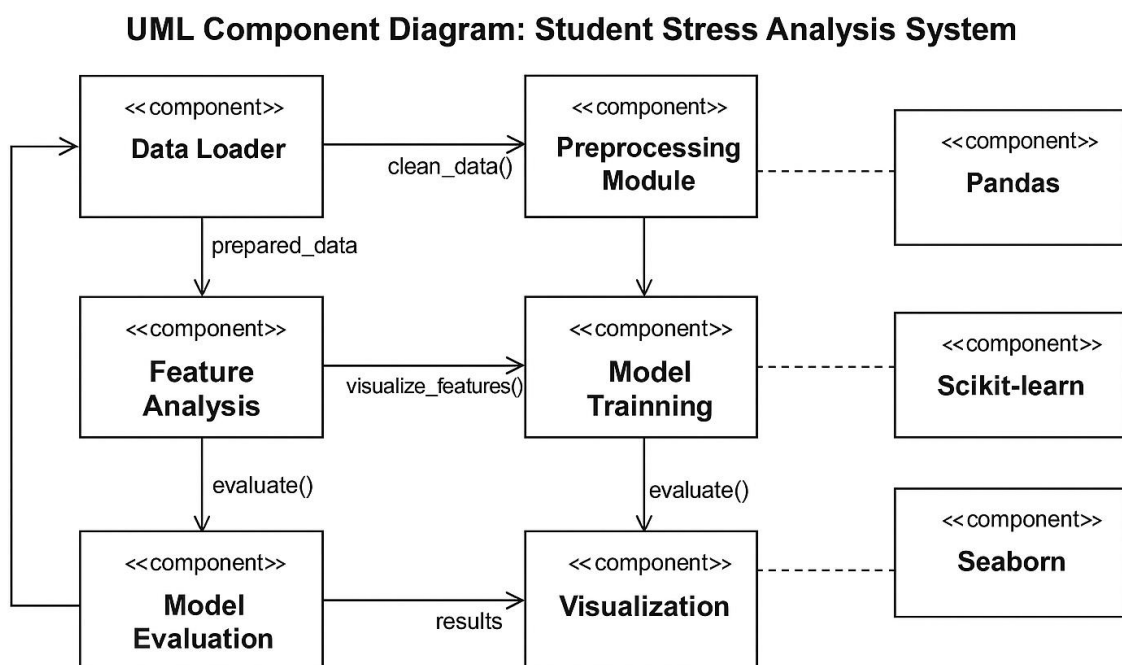


Рисунок 3.3 – UML-діаграма компонентів

На діаграмі зображено ключові логічні компоненти системи та взаємозв'язки між ними, а саме: Data Loader відповідає за імпорт вхідних даних про студентів. Preprocessing Module здійснює попередню обробку даних, для реалізації використовує бібліотеку Pandas, що зображено за допомогою залежності. Feature Analysis визначає та аналізує ключові ознаки, що включає побудову кореляційну матрицю, перевірку важливості ознак. Передає результати візуалізації до модуля Model Training, що реалізує навчання моделей Decision Tree, Random Fores, K-nearest Neighbors, використовуючи бібліотеку Scikit-learn. Model Evaluation здійснює оцінку продуктивності моделей за допомогою метрик якості: accuracy, precision, recall. F1-score. Visualization візуалізує результати аналізу та моделювання, залежить від бібліотеки Seaborn.

Отже UML-діаграма компонентів дозволяє представити модульну архітектуру програмної системи, демонструючи як структуровано відбувається аналіз факторів стресу серед студентів.

3.2 Налаштування моделей

Наступним етапом стала підготовка даних для моделювання.

На рисунку 3.14 зображено функцію для підготовки даних для моделювання.

```
def prepare_data(data):
    # Виділення цільової змінної та предикторів
    X = data.drop('stress_level', axis=1)
    y = data['stress_level']

    # Розділення даних на тренувальну та тестову вибірки (стратифіковане розбиття)
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42, stratify=y
    )

    # Збереження оригінальних колонок для інтерпретації результатів
    feature_names = X.columns

    # Стандартизація даних
    scaler = StandardScaler()
    X_train_scaled = scaler.fit_transform(X_train)
    X_test_scaled = scaler.transform(X_test)

    print(f"\nРозмір тренувальної вибірки: {X_train.shape}")
    print(f"Розмір тестової вибірки: {X_test.shape}")
    print(f"Розподіл класів у тренувальній вибірці:")
    print(y_train.value_counts().sort_index())
    print(f"Розподіл класів у тестовій вибірці:")
    print(y_test.value_counts().sort_index())

    return X, X_train, X_test, X_train_scaled, X_test_scaled, y_train, y_test, feature_names, scaler
```

Рисунок 3.4 – Функція для підготовки даних для моделювання

Цей процес включав декілька ключових кроків. По-перше, відбулося виділення цільової змінної та предикторів: датасет було розділено на матрицю ознак (X) та вектор цільової змінної (y – stress_level). По-друге, дані було розділено на тренувальну (80%) та тестову (20%) вибірки за допомогою функції train_test_split. Для забезпечення репрезентативності класів у обох вибірках використовувалося стратифіковане розбиття (stratify=y), а для відтворюваності результатів було зафіксовано random_state=42. Розмір тренувальної вибірки склав 880 рядків, а тестової – 220 рядків. Розподіл класів у тренувальній вибірці зображено на рисунку 3.5.

```
Розподіл класів у тренувальній вибірці:
stress_level
0      299
1      286
2      295
Name: count, dtype: int64
```

Рисунок 3.5 – Розподіл класів у тренувальній вибірці

Розподіл класів у тестовій вибірці зображено на рисунку 3.6.

```
Розподіл класів у тестовій вибірці:  
stress_level  
0      74  
1      72  
2      74  
Name: count, dtype: int64
```

Рисунок 3.6 – Розподіл класів у тестовій вибірці

По-третє, було виконано масштабування ознак: числові ознаки в тренувальній та тестовій вибірках було масштабовано за допомогою StandardScaler. Ця процедура є необхідною для коректної роботи деяких алгоритмів машинного навчання, які чутливі до масштабу вхідних даних.

Для підвищення ефективності моделей класифікації було проведено оптимізацію їх гіперпараметрів.

На рисунку 3.8 зображено функцію пошуку оптимальних параметрів моделей.

```

def optimize_models(X_train_scaled, y_train):
    print("\nОптимізація гіперпараметрів моделей:")

    # Параметри для пошуку (для КЛАСИФІКАЦІЙНИХ моделей)
    param_grid = {
        'Випадковий ліс': {
            'model': RandomForestClassifier(random_state=42),
            'params': {
                'n_estimators': [50, 100, 200],
                'max_depth': [5, 10, 15],
                'min_samples_split': [2, 5],
                'min_samples_leaf': [1, 2]
            }
        },
        'Дерево рішень': {
            'model': DecisionTreeClassifier(
                random_state=42,
                class_weight='balanced'
            ),
            'params': {
                'max_depth': [3, 4, 5],
                'min_samples_split': [5, 10],
                'min_samples_leaf': [2, 4],
                'max_features': ['sqrt', 'log2'],
                'ccp_alpha': [0.01, 0.02, 0.03]
            }
        },
        'К-найближчих сусідів': {
            'model': KNeighborsClassifier(),
            'params': {
                'n_neighbors': [3, 5, 7, 9],
                'weights': ['uniform', 'distance'],
                'p': [1, 2]
            }
        }
    }

    best_params = {}
    best_models = {}

```

Рисунок 3.7 – Функція пошуку оптимальних параметрів моделей

Цей процес здійснювався за допомогою методу GridSearchCV, який дозволяє автоматично підібрати найкращу комбінацію гіперпараметрів із заданого набору. Метод працює шляхом перебору всіх можливих варіантів та оцінки їх якості за допомогою крос-валідації (з параметром cv=5) за метрикою `f1_weighted`.

Оптимізація проводилася для таких моделей класифікації, як Випадковий ліс (Random Forest Classifier), Дерево рішень (Decision Tree Classifier) та К-найближчих сусідів (K-Nearest Neighbors Classifier). Для кожної моделі було визначено сітку гіперпараметрів для пошуку. Наприклад, для Випадкового лісу перебиралися такі параметри, як кількість дерев

(`n_estimators`), максимальна глибина (`max_depth`), мінімальна кількість зразків для розділення вузла (`min_samples_split`) та мінімальна кількість зразків у листковому вузлі (`min_samples_leaf`). Аналогічні набори параметрів були визначені для Дерева рішень (зокрема, `max_depth`, `min_samples_split`, `min_samples_leaf`) та К-найближчих сусідів (наприклад, `n_neighbors`, `weights`).

На рисунку 3.8 зображено оптимізацію гіперпараметрів моделей

```

Оптимізація гіперпараметрів моделей:

Оптимізація моделі: Випадковий ліс
Найкращі параметри: {'max_depth': 15, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 50}
Найкращий F1-score: 0.8866

Оптимізація моделі: Дерево рішень
Найкращі параметри: {'ccp_alpha': 0.03, 'max_depth': 3, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 5}
Найкращий F1-score: 0.8855

Оптимізація моделі: К-найближчих сусідів
Найкращі параметри: {'n_neighbors': 3, 'p': 1, 'weights': 'uniform'}
Найкращий F1-score: 0.8798

```

Рисунок 3.8 – Оптимізація гіперпараметрів моделей

За результатами роботи `GridSearchCV` для кожної моделі було визначено оптимальні значення гіперпараметрів, які забезпечують найкращу якість класифікації на тренувальних даних при крос-валідації. Ці оптимальні параметри в подальшому були використані для навчання фінальних моделей. Так, найкращі параметри для моделі "Випадковий ліс" виявилися '`max_depth`': 15, '`min_samples_leaf`': 1, '`min_samples_split`': 5, '`n_estimators`': 50 з F1-score 0.8866. Для моделі "Дерево рішень" оптимальними стали параметри '`max_depth`': 7, '`min_samples_leaf`': 2, '`min_samples_split`': 10 з F1-score 0.885. А для моделі "К-найближчих сусідів" найкращими параметрами були '`n_neighbors`': 3, '`p`': 1, '`weights`': 'uniform' з F1-score 0.8798.

3.3 Використання моделей машинного навчання

З метою виявлення найважливіших факторів, що впливають на рівень стресу студентів, було застосовано підхід машинного навчання з

використанням класифікаційних моделей. У процесі дослідження попередньо здійснено підготовку даних, побудову моделей, та оцінку їх ефективності.

Датасет було поділено на тренувальну та тестову вибірки із збереженням пропорційне класів.

На рисунку 3.9 зображено поділ датасету на тренувальну та тестову вибірку.

```
def prepare_data(data):

    X = data.drop('stress_level', axis=1)
    y = data['stress_level']

    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42, stratify=y
    )

    feature_names = X.columns

    scaler = StandardScaler()
    X_train_scaled = scaler.fit_transform(X_train)
    X_test_scaled = scaler.transform(X_test)

    print(f"\nРозмір тренувальної вибірки: {X_train.shape}")
    print(f"Розмір тестової вибірки: {X_test.shape}")
    print(f"Розподіл класів у тренувальній вибірці:")
    print(y_train.value_counts().sort_index())
    print(f"Розподіл класів у тестовій вибірці:")
    print(y_test.value_counts().sort_index())

    return X, X_train, X_test, X_train_scaled, X_test_scaled, y_train, y_test, feature_names, scaler
```

Рисунок 3.9 – Поділ датасету на тренувальну та тестову вибірку

Для класифікації рівня стресу було обрано три моделі:

- 1) Дерево рішень (Decision Tree);
- 2) Метод К-найближчих сусідів (K-nearest Neighbors);
- 3) Випадковий ліс (Random Forest).

Моделі ініціалізуються із заданими гіперпараметрами, які підібрані емпірично з метою досягнення оптимального балансу між точність та узагальненістю.

На рисунку 3.10 зображено ініціалізацію моделей класифікації з використанням гіперпараметрів.

```
models = {
    'Випадковий ліс': RandomForestClassifier(
        random_state=42,
        n_estimators=100,
        max_depth=5,
        min_samples_split=10,
        min_samples_leaf=4,
        max_features='sqrt'
    ),
    'Дерево рішень': DecisionTreeClassifier(
        random_state=42,
        max_depth=5,
        min_samples_split=5,
        min_samples_leaf=2,
        max_features='sqrt'
    ),
    'К-найближчих сусідів': KNeighborsClassifier(n_neighbors=3)
}
```

Рисунок 3.10 – Ініціалізація моделей класифікації з використанням гіперпараметрів

Кожна з моделей навчається на тренувальній вибірці, після чого виконується прогнозування як на тренувальній, так і на тестовій частинах даних.

На рисунку 3.11 зображено навчання моделей на тренувальній вибірці та отримання прогнозів.

```
model.fit(X_train_scaled, y_train)

y_train_pred = model.predict(X_train_scaled)
train_predictions[name] = y_train_pred
```

Рисунок 3.11 – Навчання моделей на тренувальній вибірці та отримання прогнозів

Для кожної моделі обчислюються основні метрики якості класифікації. Метрики розраховуються окремо для тренувальної та тестової вибірок.

На рисунку 3.12 зображено обчислення тренувальної метрики якості класифікації.

```
train_accuracy = accuracy_score(y_train, y_train_pred)
train_precision = precision_score(y_train, y_train_pred, average='weighted')
train_recall = recall_score(y_train, y_train_pred, average='weighted')
train_f1 = f1_score(y_train, y_train_pred, average='weighted')
```

Рисунок 3.12 – Обчислення тренувальної метрики якості класифікації

На рисунку 3.13 зображено обчислення тестової метрики якості класифікації

```
test_accuracy = accuracy_score(y_test, y_test_pred)
test_precision = precision_score(y_test, y_test_pred, average='weighted')
test_recall = recall_score(y_test, y_test_pred, average='weighted')
test_f1 = f1_score(y_test, y_test_pred, average='weighted')
```

Рисунок 3.13 – Обчислення тестової метрики якості класифікації

Додатково проводиться крос-валідація для оцінки стабільності моделей.

На рисунку 3.14 зображено реалізацію валідації для оцінювання стабільності моделей.

```
cv_scores = cross_val_score(model, X_train_scaled, y_train, cv=5, scoring='f1_weighted')
cv_f1_mean = cv_scores.mean()
cv_f1_std = cv_scores.std()
```

Рисунок 3.14 – Реалізація валідації для оцінювання стабільності моделей

На основі значення F1-метрики на тестовій вибірці визначається найефективніша модель.

На рисунку 3.15 визначено найкращої моделі за метрикою F1 на тестовій вибірці.

```
t_model_name = max(results, key=lambda x: results[x]['Test_F1'])  
nt(f"\nНайкраща модель за метрикою F1-score: {best_model_name}")
```

Рисунок 3.15 – Визначення найкращої моделі за метрикою F1 на тестовій вибірці

На рисунку 3.16 виведено результат найкращої моделі за метрикою F1 на тестовій вибірці.

```
Найкраща модель за метрикою F1-score: Випадковий ліс
```

Рисунок 3.16 – Результат найкращої моделі за метрикою F1 на тестовій вибірці

Для моделі Random Forest додатково проводиться аналіз важливості ознак, що дозволяє визначити, які саме фактори найбільше впливають на рівень стресу.

На рисунку 3.17 зображено аналіз важливості ознак за допомогою моделі Random Forest.

Важливість ознак згідно з моделлю випадкового лісу:		
	Feature	Importance
5	blood_pressure	0.183790
6	sleep_quality	0.091024
14	teacher_student_relationship	0.069895
12	academic_performance	0.068264
16	social_support	0.067137
11	basic_needs	0.065226
3	depression	0.064263
0	anxiety_level	0.052537
15	future_career_concerns	0.052222
10	safety	0.050893

Рисунок 3.17 – Аналіз важливості ознак за допомогою моделі Random Forest

Для визначення найбільш ефективної моделі для класифікації рівня стресу було проведено порівняльний аналіз продуктивності трьох моделей, а саме: Випадковий ліс, Дерево рішень, К-найближчих сусідів.

На рисунку 3.18 виведено порівняльну таблицю результатів моделей.

Модель	Train Accuracy	Train F1	Test Accuracy	Test F1	CV F1-score
Випадковий ліс	0.9795	0.9796	0.8818	0.8817	0.8707 ± 0.0218
Дерево рішень	0.9330	0.9330	0.8636	0.8633	0.8855 ± 0.0182
К-найближчих сусідів	0.9364	0.9364	0.8500	0.8498	0.8740 ± 0.0197

Рисунок 3.18 – Порівняльна таблиця результатів моделей

Модель «випадковий ліс» демонструє найвищу продуктивність на тестовій вибірці з показником 0,8817.

Модель «Дерево рішень» посідає друге місце за продуктивністю на тестових даних, а саме 0,8633. Що є досить добрий результатом, враховуючи її простоту.

Модель «К-найближчих сусідів» показала дещо нижчі результати на тестовій вибірці – 0.8498.

Таким чином, на основі проведеного аналізу, модель «Випадковий ліс» визначена як найбільш ефективна для класифікації рівня стресу студентів.

На рисунку 3.19 представлені матриці плутанини для кожної з розроблених моделей, розраховані як на тренувальних, так і на тестових даних.

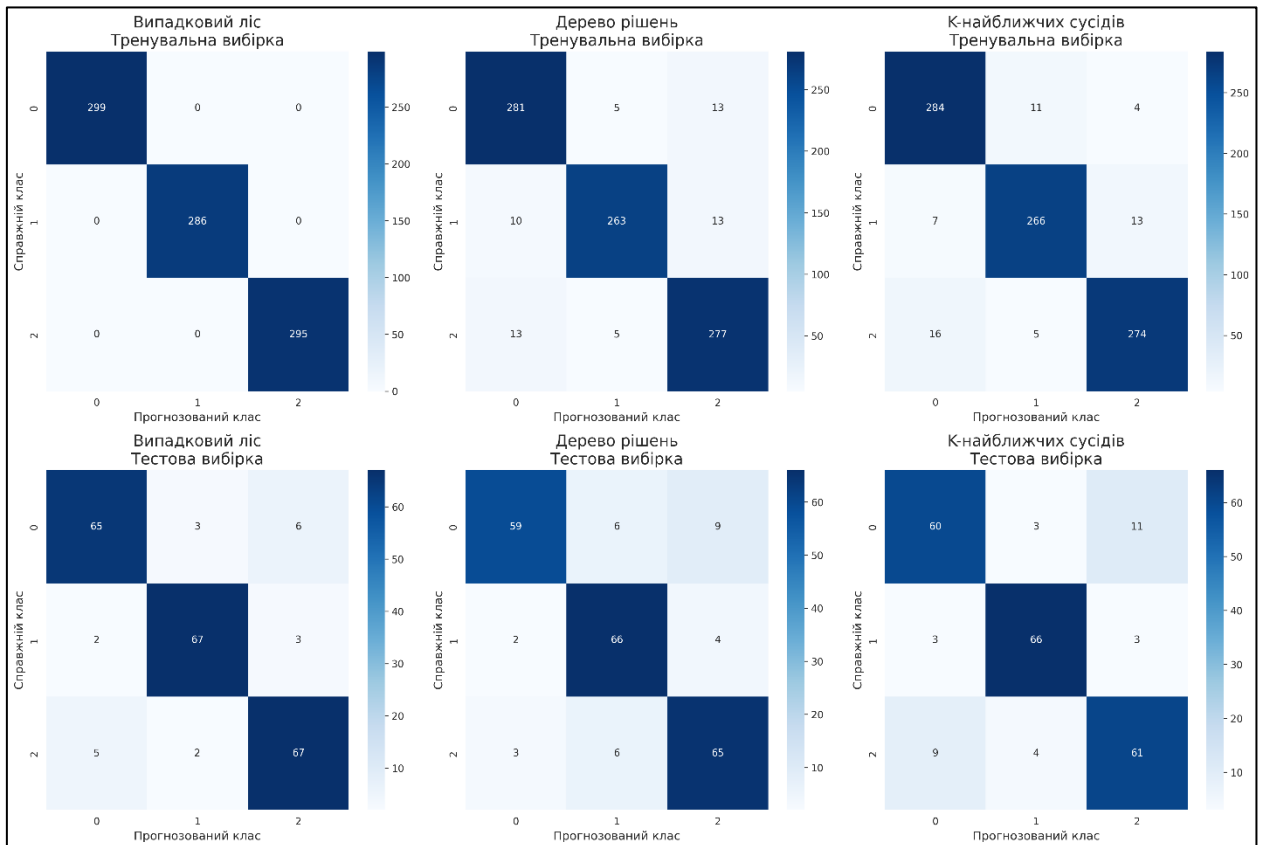


Рисунок 3.19 – Матриці плутанини для моделей класифікації

Аналіз цих матриць показує, наприклад, що модель "Випадковий ліс" демонструє найкращу узагальнюючу здатність на тестовій вибірці серед представлених моделей, маючи найменшу кількість помилок. Хоча на тренувальних даних він показав ідеальну точність, його результати на тестових даних є надійними.

На рисунку 3.20 візуалізовано топ-10 найважливіших ознак.

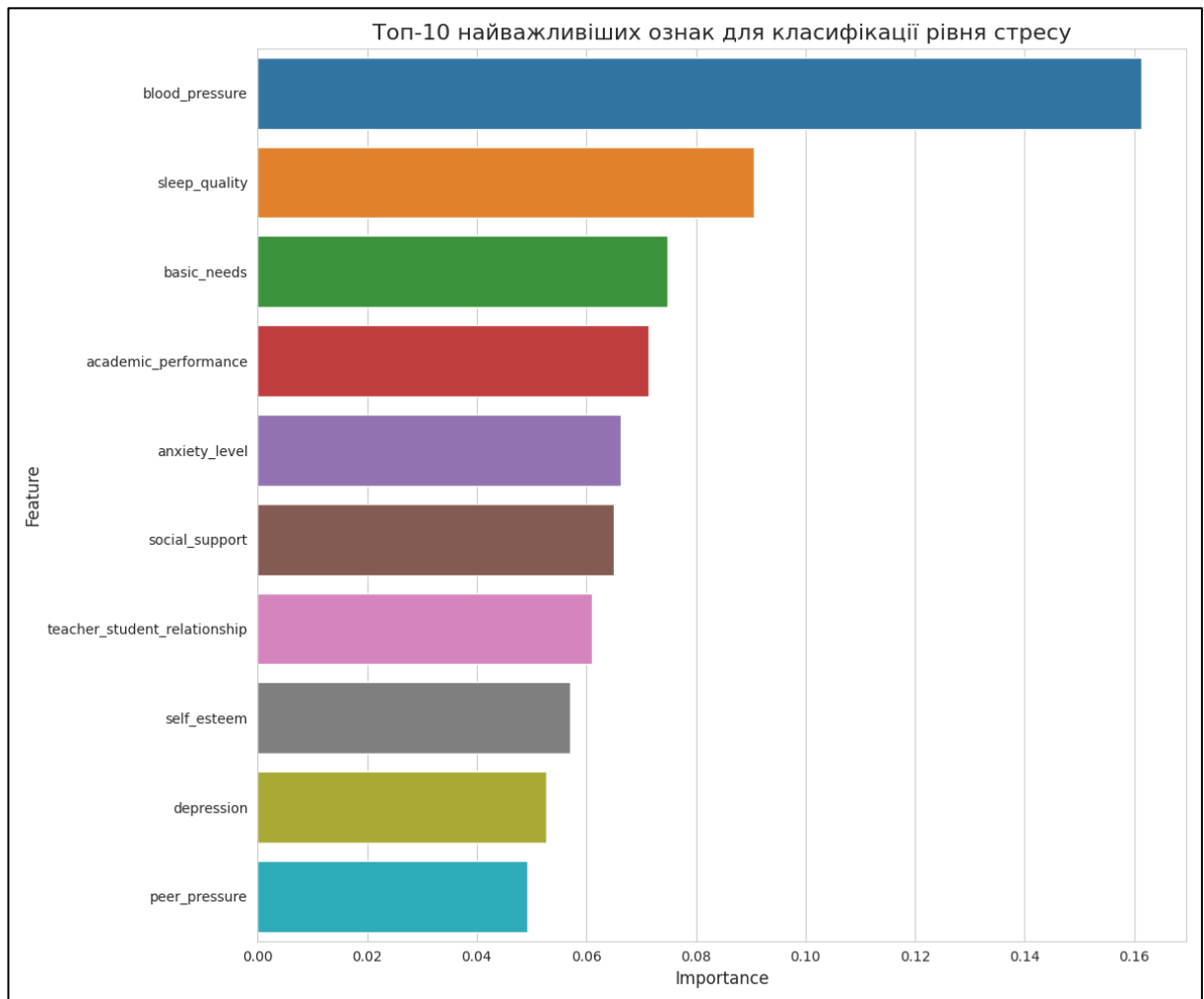


Рисунок 3.20 – Важливість ознак за моделлю Випадковий ліс

Згідно з результатів найбільш значуща ознака – blood pressure, що свідчить про сильний фізіологічний зв'язок між рівнем стресу та показниками здоров'я.

Sleep quality – висока важливість цієї змінної свідчить про те, що поганий або нестабільний сон є частим наслідком чи провокатором стресу.

Basic needs – недостатнє задоволення базових потреб має помітний вплив на емоційний стан.

Academic performance – низька успішність може бути як наслідком, так і причиною стресових станів.

Anxiety level - високий рівень тривожності часто супроводжує високий рівень стресу.

Social support – чим менше підтримки студент відчуває з боку близьких, тим вища ймовірність розвитку стресу.

Teacher student relationship – якість взаємодії з викладачами також впливає на рівень емоційного навантаження.

Depression – наявність депресивних симптомів істотно збільшує ризик підвищеного рівня стресу.

Peer pressure – наявність цього чинника, сприяє формуванню стресу через постійне соціальне напруження та страх бути неприйнятим.

Для кращого розуміння логіки роботи моделі Дерева рішень було побудовано її графічне представлення (Decision Tree Visualization), яке наведено на рисунку 3.21.

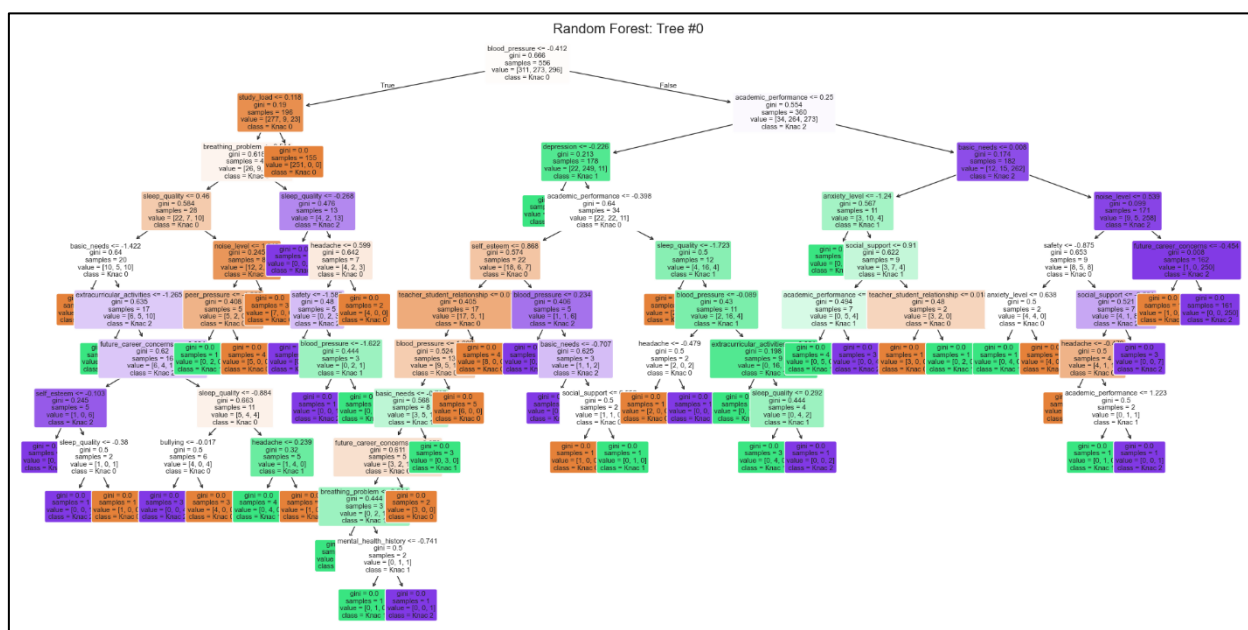


Рисунок 3.21 – Візуалізація Random Forest

Також було здійснено графічне порівняння метрик моделей. На рисунку 3.22 представлено стовпчасту діаграму, що порівнює ключові метрики (Accuracy, Precision, Recall, F1-score) для всіх розроблених моделей на тестовій вибірці, що дозволяє наочно оцінити їх відносну ефективність.

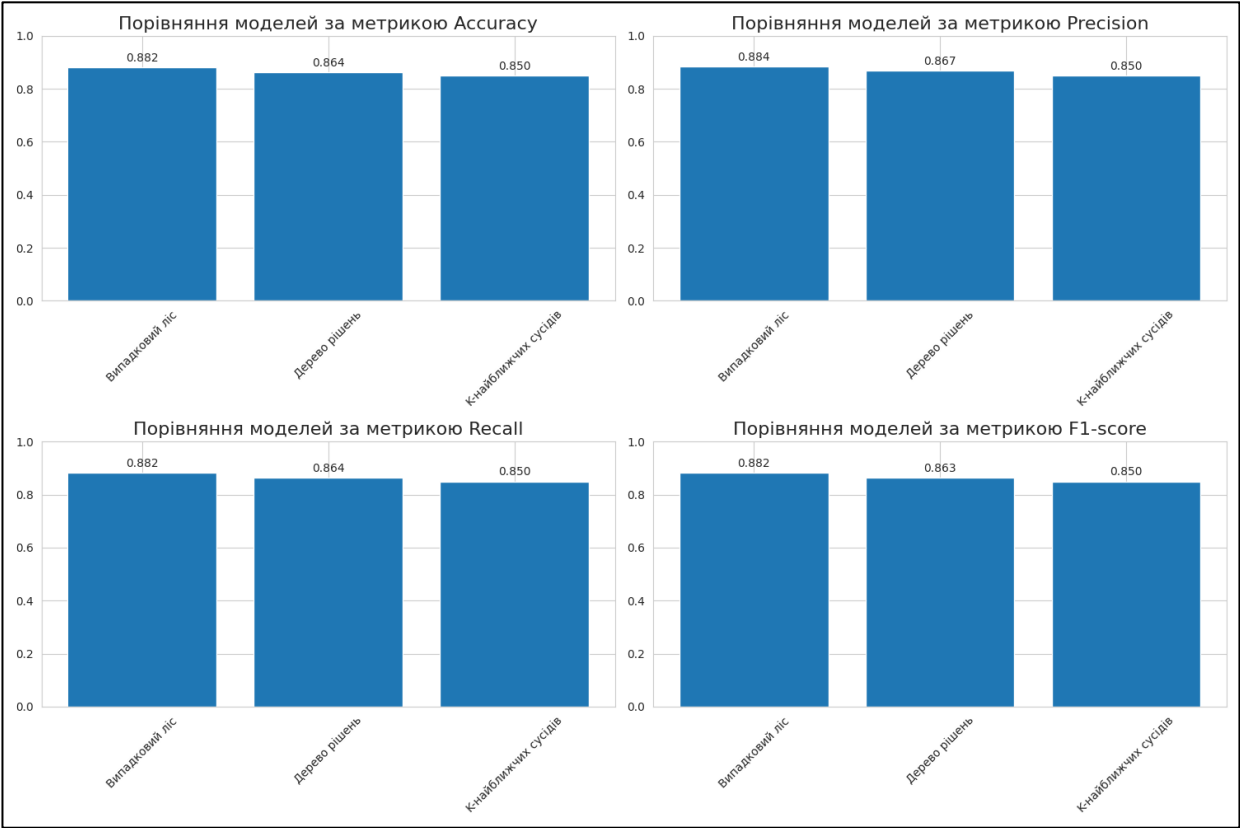


Рисунок 3.22 – Порівняння метрик моделей на тестовій вибірці

Крім того, для кожної моделі було сформовано детальні звіти класифікації (Detailed Classification Reports).

На рисунку 3.23 зображено детальний звіт класифікації.

ВИПАДКОВИЙ ЛІС:				
	precision	recall	f1-score	support
Клас 0	0.91	0.84	0.87	74
Клас 1	0.89	0.89	0.89	72
Клас 2	0.85	0.92	0.88	74
accuracy			0.88	220
macro avg	0.88	0.88	0.88	220
weighted avg	0.88	0.88	0.88	220
ДЕРЕВО РІШЕНЬ:				
	precision	recall	f1-score	support
Клас 0	0.92	0.80	0.86	74
Клас 1	0.85	0.92	0.88	72
Клас 2	0.83	0.88	0.86	74
accuracy			0.86	220
macro avg	0.87	0.86	0.86	220
weighted avg	0.87	0.86	0.86	220
K-НАЙБЛИЖЧИХ СУСІДІВ:				
	precision	recall	f1-score	support
Клас 0	0.83	0.81	0.82	74
Клас 1	0.90	0.92	0.91	72
Клас 2	0.81	0.82	0.82	74
accuracy			0.85	220
macro avg	0.85	0.85	0.85	220
weighted avg	0.85	0.85	0.85	220

Рисунок 3.23 – Детальний звіт класифікації

Звіт, представлений на рисунку 3.23, включає значення Precision, Recall, F1-score та support для кожного класу рівня стресу, надаючи повну картину ефективності моделей щодо розпізнавання кожного окремого рівня стресу. Дані звіти виводяться у консоль під час виконання скрипту.

Отже, серед розроблених моделей класифікації, а саме "Випадкового лісу", "Дерева рішень" та "K-найближчих сусідів", найкращі результати на тестовій вибірці продемонструвала модель "Випадковий ліс". Її показники показали найкращі результати серед усіх варіантів – точність (accuracy) становила 0.88, F1-score– 0.88, recall -0,88.

3.4 Висновки

В даному розділі було розроблено інформаційну технологію для прогнозування рівня стресу студентів закладів вищої освіти. Були застосовані наступні моделі машинного навчання: Random Forest, Decision Tree, KNeighbors, виконано їх налаштування. Зроблено порівняння оцінок моделей за різними метриками, а саме: Accuracy, Precision, Recall, F1-score. За результатами оцінки даних метрик усі моделі показали хороші результати, а саме точності вище 85%. Також було обрано найкращу модель за метрикою F1_Score, нею виявилась Random Forest із точністю 88%. Дослідження показали, що для моделі Random Forest найвпливовішою ознакою є «Blood pressure».

ВИСНОВКИ

В ході виконання бакалаврської кваліфікаційної роботи проведено аналіз предметної області прогнозування стресу студентів закладів вищої освіти та визначено актуальність теми.

Проаналізовано близько 10 наукових публікацій, присвячених прогнозуванню стресу у студентів. Аналіз вітчизняних та зарубіжних досліджень підтверджує високу актуальність теми в контексті сучасної освіти. Основними причинами стресу виступають: академічне навантаження, оцінювання, дисбаланс між навчанням і особистим життям, а також фінансові труднощі. Частина авторів досліджень вказує на зв'язок стресу з депресією та емоційним вигоранням, що негативно впливає на мотивацію й успішність. Українські науковці підкреслюють додатковий вплив пандемії та воєнних дій на психоемоційний стан студентів. Разом з тим, окремі дослідження звертають увагу на відмінності у сприйнятті стресу залежно від статі, спеціальності та соціального статусу. До факторів, що пом'якшують вплив стресу відносять: психологічну стійкість, емоційний самоконтроль і соціальну підтримку. Проаналізувавши статті, стає чітко зрозуміло, що стрес розглядається як системне явище, що має значний вплив на студентське середовище в умовах зростаючих освітніх вимог і зовнішніх викликів.

Проведено аналіз сучасних методів машинного навчання для створення інформаційної технології прогнозування стресу у студентів, що підходять для вирішення задачі класифікації та обрано наступні: дерево рішень (Decision Tree), метод k-найближчих сусідів (K-Nearest Neighbors) та випадковий ліс (Random Forest).

Для проведення дослідження було обрано мову програмування Python, яка завдяки своїй простоті, гнучкості та великому набору бібліотек є відмінним вибором для виконання задачі прогнозування рівня стресу студентів закладів вищої освіти. В якості середовища розробки було обрано онлайн-платформу Kaggle, яка спеціалізується на машинному навчанні та науці про дані.

Проведено попередню обробку, очищення та нормалізацію даних. Виконано розвідувальний аналіз, що дозволив виявити значущі закономірності у розподілі рівня стресу та сформувані інформативні ознаки для навчання моделей. Аналіз показав, що найбільш впливовими факторами, що збільшують ймовірність виникнення стресу, є: тривога, депресія, булінг, проблеми зі сном, фінансові труднощі та занепокоєння щодо майбутньої кар'єри. У той час, такі аспекти, як висока самооцінка, хороша академічна успішність та позитивні взаємини з викладачами, понижують рівень стресу.

Розроблено інформаційну технологію для прогнозування стресу у студентів закладів вищої освіти, яка базується на алгоритмах машинного навчання, а саме: Випадковий ліс (Random Forest), Дерево рішень (Decision Tree), К-найближчих сусідів (K-Neighbors). Проведено навчання моделей та їх тестування з використанням метрик точності, повноти, F1- score, а також кросвалідації. За результатами оцінювання, найкращу якість прогнозування продемонструвала модель Випадковий ліс (Random Forest) із точністю прогнозування 0,88. Дослідження показали, що для моделі Random Forest найвпливовішою ознакою є «Blood pressure» [26].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Верчинська М.О., Войцеховська О.О. Розвідувальний аналіз даних для інформаційної технології аналізу та прогнозування стресу у студентів закладів вищої освіти методами машинного навчання. *Міжнародна науково-практична інтернет-конференція студентів, аспірантів та молодих науковців «Молодь в науці: дослідження, проблеми, перспективи»*: збірник матеріалів конференції. Вінниця, 2025. URL : <https://conferences.vntu.edu.ua/index.php/mn/mn2025/paper/viewFile/25749/21297>
2. Böke B. N., Mills D. J., Mettler J., Heath N. L. Stress and Coping Patterns of University Students, *Journal of College Student Development*. 2019. 60(1). P. 85-103. URL: <https://dx.doi.org/10.1353/csd.2019.0005> (дата звернення: 15.05.2025).
3. Tamlina Larysa, Hohol Oleksandra, Ihnatenko Yaroslava. Exploring the Impact of Military Confluctis on Mental Health of Students: The Case of Ukraine. *Munich Personal RePEc Archive (MPRA)*. 2023. Paper 119435. 23 p. URL: <https://mpra.ub.uni-muenchen.de/119435/> (дата звернення 15.05.2025).
4. Kajleen Kaur, Harpreet Kaur, Jasmit Kaur, Yogieta S. Mehra. Assessment and Analysis of Factors Impacting Stress in Students in Higher Educational Institutes in Delhi, *International Journal of Experimental*. 2024. Vol. 41. P. 305-321. URL: [85dc93a1f284c2698568ea253d07cf0e302d.pdf](https://doi.org/10.1353/csd.2019.0005) (дата звернення 18.05.2025).
5. Ahmad R., Chaudhary, P., Randhawa, R. K. A study to assess the level of stress among college students in a selected Govt. College of Nursing, Srinagar, J&K, India. *International Journal of Creative Research Thoughts*. 2021. Vol.9. Issue 5. P. 295-307. URL: <https://ijert.org/papers/IJCRT2105883.pdf> (дата звернення 17.05.2025).
6. Кондратенко Є. М., Церковна О. В. Дослідження рівня стресу у студентів технічного закладу вищої освіти : *матеріали наук.-практ. конф. «Молодіжна наукова ліга. Науковий простір: аналіз, сучасний стан, тренди та перспективи»*. м. Івано-Франківськ, 15 грудня 2023р. Івано-Франківськ, 2023.

- C. 682–683. URL: <https://archive.liga.science/index.php/conference-proceedings/article/view/656/664> (дата звернення 17.05.2025).
7. Gong Z, Wang H, Zhong M, Shao Y. College students' learning stress, psychological resilience and learning burnout: status quo and coping strategies. *BMC Psychiatry*. 2023. Article 389. DOI: [10.1186/s12888-023-04783-z](https://doi.org/10.1186/s12888-023-04783-z) (дата звернення 17.05.2025).
8. Nguyen T., Pham, H., Le, V. Family and Academic Stress and Their Impact on Student's Depression and Academic Performance. *Journal of Education and Health Promotion*. 2022. Article 16:13:869337. DOI: [10.3389/fpsyh.2022.869337](https://doi.org/10.3389/fpsyh.2022.869337) (дата звернення 18.05.2025).
9. Smith J., Lee A. Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups and COVID-19. *Frontiers in Psychology*. 2022. Volume 13. Article 13:886344. URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.886344/full> (дата звернення 18.05.2025).
10. Martinez L., Kim S. Stress, student burnout and study engagement – a cross sectional comparison of university students of different academic subjects. *BMC Psychology*. 2025. Article 293. URL: <https://bmcp psychology.biomedcentral.com/articles/10.1186/s40359-025-02602-6> (дата звернення 18.05.2025).
11. Пухно С. В. Стресостійкість студентів закладів вищої освіти як психологічна проблема. *Слобожанський науковий вісник. Серія: Психологія*. 2023. №. 1 С. 44-48. URL: <https://journals.spu.sumy.ua/index.php/psy/article/view/58> (дата звернення 20.05.2025).
12. Cai Qingyu. Case Study of Model Selection on Customer Information Task Based on Machine Learning Algorithms. *BCP Business & Management*. 2023. vol. 38. P. 2827-2835 URL: <https://bcp publication.org/index.php/BM/article/view/4197> (дата звернення 19.05.2025).

13. Liner Regression in Python. Real Python. 2024. URL: <https://realpython.com/linear-regression-in-python/> (дата звернення 19.05.2025).
14. Омельченко І. Г., Чуйко Г. П. Метод дерев рішень у комп'ютерній інженерії: матеріали XXVI Всеукраїнської науково-практичної конференції «Могилянські читання – 2023: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти, м. Миколаїв, 6-10 листопада 2023р. Миколаїв, 2023. С. 431-432 URL: <https://du.chmnu.edu.ua/wp-content/uploads/2024/01/%D0%9C%D0%BE%D0%B3%D0%B8%D0%BB%D1%8F%D0%BD%D1%81%D0%BA-%D1%87%D0%B8%D1%82%D0%B0%D0%BD%D0%BD%D1%8F-2023-%D0%A2%D0%95%D0%97%D0%98.pdf#page=431> (дата звернення 19.05.2025).
15. Груша В. М. Застосування методу опорних векторів для визначення необхідності штучного поливу рослин на основі індукції флуоресценції хлорофілу. *Cybernetics and Computer Technologies*. 2024. № 2. С. 67-73. URL: http://cctech.org.ua/images/docs/Articles/2024/paper_24_2_7.pdf (дата звернення 20.05.2025).
16. Dragutin Petkovic, Russ Altman, Mike Wong, Arthur Vigil. A. Improving the explainability of Random Forest classifier – user centered approach. *PMC Disclaimer*. 2023. Vol. 23. P. 204-215. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5728671/> (дата звернення 20.05.2025).
17. Колесник А. М., Ткач, М. В. Пошук аномалій в системах автентифікації за допомогою методів машинного навчання: матеріали Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених. Навчально-науковий Фізико-технічний інститут. м. Київ, 11-12 травня 2023р. Київ, 2023. С. 387-391. URL: <https://ela.kpi.ua/server/api/core/bitstreams/3f5a2107-12d5-43cf-b167-41373cf3664f/content> (дата звернення 20.05.2025).
18. Замуруєва О. В., Кримусь, А. С., Ольхова, Н. В. Об'єктно-орієнтоване програмування в Python : курс лекцій. Луцьк. Вежа-Друк. 2018. 64 с. URL:

- <https://evnuir.vnu.edu.ua/bitstream/123456789/14344/1/Python.pdf> (дата звернення 20.05.2025).
19. Васильєв О. М. Програмування мовою Python: Bohdan Books. 2022. 504 с. URL: https://books.google.com/books/about/%D0%9F%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F_%D0%BC%D0%BE%D0%B2%D0%BE%D1%8E_Py.html?id=11hcEAAAQBAJ (дата звернення 21.05.2025).
20. Що таке Python? URL: <https://acode.com.ua/intro-python/#toc-7> (дата звернення 22.05.2025).
21. Кравченко С. М., Гришкун, Є. О., Власенко, О. В. Методи класифікації машинного навчання з використанням бібліотеки scikit-learn. *Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки*. 2020. С. 124-125. URL: https://www.tech.vernadskyjournals.in.ua/journals/2020/3_2020/part_1/21.pdf (дата звернення 23.05.2025).
22. Мокін В. Б., Дратований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних. Навчальний посібник. м. Вінниця. 2024. 263 с. URL: <https://iq.vntu.edu.ua/repository/getfile.php/8163.pdf> (дата звернення 23.05.2025).
23. Chhabii. Student Stress Factors: A Comprehensive Analysis. 2024. URL: <https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis> (дата звернення 2.05.2025).
24. Demystifying the Correlation Matrix in Data Science. Towards Data Science. 2024. – URL: <https://towardsdatascience.com/demystifying-the-correlation-matrix-in-data-science-6b8a4482b6e2/> (дата звернення 25.05.2025).
25. Martina Udovičić, Ksenija Baždarić, Lidija Bilić-Zulle1, Mladen Petrovečki1. What we need to know when calculating the coefficient of correlation. *Biochemia Medica* 2007. Vol. 17(1). P. 10-15. URL: https://www.biochemia-medica.com/assets/images/upload/xml_tif/Udovicic_M_et_al_-

[What we need to know when calculating the coefficient of correlation.pdf](#)
(дата звернення 25.05.2025).

26. Stress of students of higher education university by Maryna Verchynska.
2025. URL: <https://www.kaggle.com/code/marynaverchynska2/stress-of-students-of-higher-education-university> (дата звернення 26.05.2025).

Додаток А
(обов'язковий)

Міністерство освіти і науки України
Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

_____ д.т.н., проф. Віталій МОКІН

« ____ » _____ 2025 р.

ТЕХНІЧНЕ ЗАВДАННЯ
на бакалаврську кваліфікаційну роботу
ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ
СТРЕСУ У СТУДЕНТІВ ЗАКЛАДІВ ВИЩОЇ ОСВІТИ
08-34.БДР.002.00.000 ТЗ

Керівник: Ph.D., ст. викл. каф. САІТ

_____ Ольга ВОЙЦЕХОВСЬКА

«__» _____ 2025 р.

Розробила студентка гр. 2ІСТ-216

_____ Марина ВЕРЧИНСЬКА

«__» _____ 2025 р.

Вінниця 2025

1. Підстава для проведення робіт.

Підставою для виконання роботи є наказ №__ по ВНТУ від «__» _____ 2025р., та індивідуальне завдання на БКР, затверджене протоколом №__ засідання кафедри САІТ від «__» _____ 2025р.

2. Джерела розробки:

1) Tamilina Larysa, Hohol Oleksandra, Ihnatenko Yaroslava. Exploring the Impact of Military Confluctis on Mental Health of Students: The Case of Ukraine. *Munich Personal RePEc Archive (MPRA)*. 2023. 5с. – URL: <https://mpra.ub.uni-muenchen.de/119435/>.

2) Nguyen T., Pham, H., Le, V. Family and Academic Stress and Their Impact on Student's Depression and Academic Performance. *Journal of Education and Health Promotion*. 2023. – URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9243415/>.

3. Мета і призначення роботи. Метою дослідження є підвищення точності прогнозування стресу у студентів закладів вищої освіти шляхом створення інформаційної технології.

4. Вихідні дані для проведення робіт: Kaggle Dataset «Student Stress Factors: A Comprehensive Analysis» URL: - <https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis/suggestions>

5. Методи дослідження: розвідувальний аналіз даних, інженерія ознак, методи машинного навчання (Random Forest, Decision Tree, KNN), GridSearch.

6. Етапи роботи і терміни їх виконання:

- | | | | |
|---|-------|---|-------|
| а) Аналіз предметної області | _____ | – | _____ |
| б) Аналіз наукових робіт дослідження стресу у студентів | _____ | – | _____ |
| в) Вибір оптимального рішення та налаштувань для розв'язання поставленої задачі | _____ | – | _____ |
| г) Розроблення інформаційної технології аналізу та прогнозування стресу у студентів | _____ | – | _____ |
| д) Оформлення матеріалів до захисту БКР | _____ | – | _____ |

7.Очікувані результати та порядок реалізації

Отримання інформаційної технології для прогнозування стресу у студентів закладів вищої освіти.

8.Вимоги до розробленої документації

Текстова та ілюстративна частини роботи оформлені у відповідності до вимог «Методичних вказівок до виконання бакалаврських кваліфікаційних робіт для студентів спеціальностей: 124 «Системний аналіз», 126 «Інформаційні системи та технології» (освітня програма «Прикладні інформаційні технології»)).

7. Порядок приймання роботи

Публічний захист	«__» _____	2025 р.
Початок розробки	«__» _____	2025 р.
Граничні терміни виконання БКР	«__» _____	2025 р.

Розробила студентка групи 2ІСТ-216 _____ Марина ВЕРЧИНСЬКА

Додаток Б

(обов'язковий)

ПРОТОКОЛ ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Назва роботи: «Інформаційна технологія аналізу та прогнозування стресу у студентів закладів вищої освіти»

Тип роботи: бакалаврська кваліфікаційна робота

Підрозділ: кафедра САІТ, ФІТА, гр. 2ІСТ-216

Коефіцієнт подібності текстових запозичень, виявлених у роботі системою StrikePlagiarism 2,14 %

Висновок щодо перевірки кваліфікаційної роботи (відмітити потрібно):

- Запозичення, виявлені у роботі, є законними і не містять ознак плагіату, фабрикації, фальсифікації. Роботу прийняти до захисту
- У роботі не виявлено ознак плагіату, фабрикації, фальсифікації, але надмірна кількість текстових запозичень та/або наявність типових розрахунків не дозволяють прийняти рішення про оригінальність та самостійність її виконання. Роботу направити на доопрацювання.
- У роботі виявлено ознаки плагіату та/або текстових маніпуляцій як спроб укриття плагіату, фабрикації, фальсифікації, що суперечить вимогам законодавства та нормам академічної доброчесності. Робота до захисту не приймається.

Експертна комісія:

Віталій МОКІН, зав. каф. САІТ

_____ (підпис)

Євгеній КРИЖАНОВСЬКИЙ, доц. каф. САІТ

_____ (підпис)

Особа, відповідальна за перевірку _____

(підпис)

Сергій ЖУКОВ

З висновком експертної комісії ознайомлений(-на)

Керівник _____

(підпис)

Ольга ВОЙЦЕХОВСЬКА

Здобувач _____

(підпис)

Верчинська Марина

Додаток В

(довідниковий)

Фрагмент лістингу програми

Код ноутбуку «Stress of students of higher education university»

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.metrics import (accuracy_score, precision_score, recall_score, f1_score,
                             classification_report, confusion_matrix)

from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.pipeline import Pipeline

import warnings

warnings.filterwarnings('ignore')

plt.rcParams['font.family'] = 'Deja Vu Sans'
sns.set_style('whitegrid')
plt.rcParams['figure.figsize'] = (12, 8)
plt.rcParams['axes.titlesize'] = 16
plt.rcParams['axes.labelsize'] = 12

# Завантаження датасету
def load_data():
    try:
        data = pd.read_csv('/kaggle/input/stressleveldataset/StressLevelDataset.csv')
    return data

```

```

except Exception as e:
    print(f"Помилка завантаження датасету: {e}")
    return None

# Функція для розвідувального аналізу даних
def exploratory_data_analysis(data):
    print("Інформація про датасет:")
    print(f"Кількість рядків і стовпців: {data.shape}")
    print("\nПриклад рядків даних:")
    print(data.head())
    print("\n...")
    print(data.tail())

    print("\nОписова статистика:")
    print(data.describe())

    print("\nПеревірка наявності пропущених значень:")
    missing_values = data.isnull().sum()
    print(missing_values)
    if missing_values.sum() > 0:
        print("Заповнення пропущених значень...")
        data = data.fillna(data.median())

    print("\nАналіз пропущених значень у даних:")
    missing_info = data.isnull().sum().to_frame(name='Кількість NaN')
    missing_info['% від загального'] = (missing_info['Кількість NaN'] / len(data) *
100).round(2)
    missing_info = missing_info[missing_info['Кількість NaN'] > 0].sort_values(by='Кількість
NaN', ascending=False)

    if not missing_info.empty:
        print("Є пропущені значення у таких колонках:")
        print(missing_info)
        print("\nЗагальна кількість відсутніх значень:", data.isnull().sum().sum())

```

```

    print("Заповнення пропущених значень медіанами для числових колонок...")
    data = data.fillna(data.median(numeric_only=True))
else:
    print("Усі колонки повністю заповнені. Пропущених значень немає.")

# Дослідження характеристик цільової змінної
target_analysis(data)

# Кореляційна матриця та візуалізація
correlation_analysis(data)

# Розподіл особливостей даних
feature_distribution_analysis(data)

return data

# Аналіз цільової змінної
def target_analysis(data):
    # Перевірка унікальних значень цільової змінної
    print("\nУнікальні значення рівня стресу:")
    print(data['stress_level'].value_counts().sort_index())

    # Гістограма розподілу стресу
    plt.figure(figsize=(10, 6))
    sns.countplot(data=data, x='stress_level')
    plt.title('Розподіл рівня стресу серед студентів (класи)')
    plt.xlabel('Рівень стресу (клас)')
    plt.ylabel('Кількість студентів')
    plt.savefig('stress_distribution.png')
    plt.close()

    # Кругова діаграма розподілу рівнів стресу
    plt.figure(figsize=(10, 8))

```

```

stress_counts = data['stress_level'].value_counts()
stress_percentages = (stress_counts / len(data) * 100).round(1)

plt.pie(stress_percentages, labels=[f'Рівень {level}: {pct}%'
                                   for level, pct in zip(stress_counts.index, stress_percentages)],
        autopct='%1.1f%%', startangle=90)
plt.title('Кількісний розподіл рівнів стресу (%)')
plt.axis('equal')
plt.savefig('stress_distribution_pie.png')
plt.close()

print("\nСтатистика цільової змінної 'stress_level':")
print(data['stress_level'].describe())

# Кореляційний аналіз
def correlation_analysis(data):
    # Створюємо копію даних для безпечної модифікації
    data_copy = data.copy()

    # Приведення до числового типу (на випадок, якщо тип був object)
    cols_to_convert = ['anxiety_level', 'stress_level']
    for col in cols_to_convert:
        if col in data_copy.columns:
            # Виправляємо помилку: передаємо значення колонки, а не її назву
            data_copy[col] = pd.to_numeric(data_copy[col], errors='coerce')

    # Заповнення пропущених значень для числових колонок
    data_copy = data_copy.fillna(data_copy.median(numeric_only=True))

    print("\nТипи даних у колонках після перетворення:")
    print(data_copy.dtypes)

    # Побудова кореляційної матриці

```

```

correlation_matrix = data_copy.corr()

# Кореляція з цільовою змінною
if 'stress_level' in correlation_matrix.columns:
    corr_with_target = correlation_matrix['stress_level'].sort_values(ascending=False)
    print("\nКореляція між рівнем стресу та іншими факторами:")
    print(corr_with_target)
else:
    print("\nПопередження: колонка 'stress_level' відсутня у кореляційній матриці.")
    corr_with_target = pd.Series()

# Візуалізація кореляційної матриці
plt.figure(figsize=(16, 12))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)
plt.title('Кореляційна матриця показників')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.savefig('correlation_matrix.png')
plt.close()

# Візуалізація топ-5 найбільш корельованих з цільовою змінною факторів
if not corr_with_target.empty:
    top_correlated = corr_with_target[1:6] # Виключити 'stress_level' сам з себе

    plt.figure(figsize=(15, 10))
    for i, (feature, corr_value) in enumerate(top_correlated.items(), 1):
        plt.subplot(2, 3, i)
        sns.boxplot(x='stress_level', y=feature, data=data_copy)
        plt.title(f'{feature} vs. stress_level\nКореляція: {corr_value:.2f}')

    plt.tight_layout()
    plt.savefig('top_correlations.png')
    plt.close()

return correlation_matrix

```

Додаток Г
(обов'язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ
СТРЕСУ У СТУДЕНТІВ ЗАКЛАДІВ ВИЩОЇ ОСВІТИ**

Нормоконтроль: к.т.н., доцент

_____ Сергій ЖУКОВ

«___» _____ 2025 р.

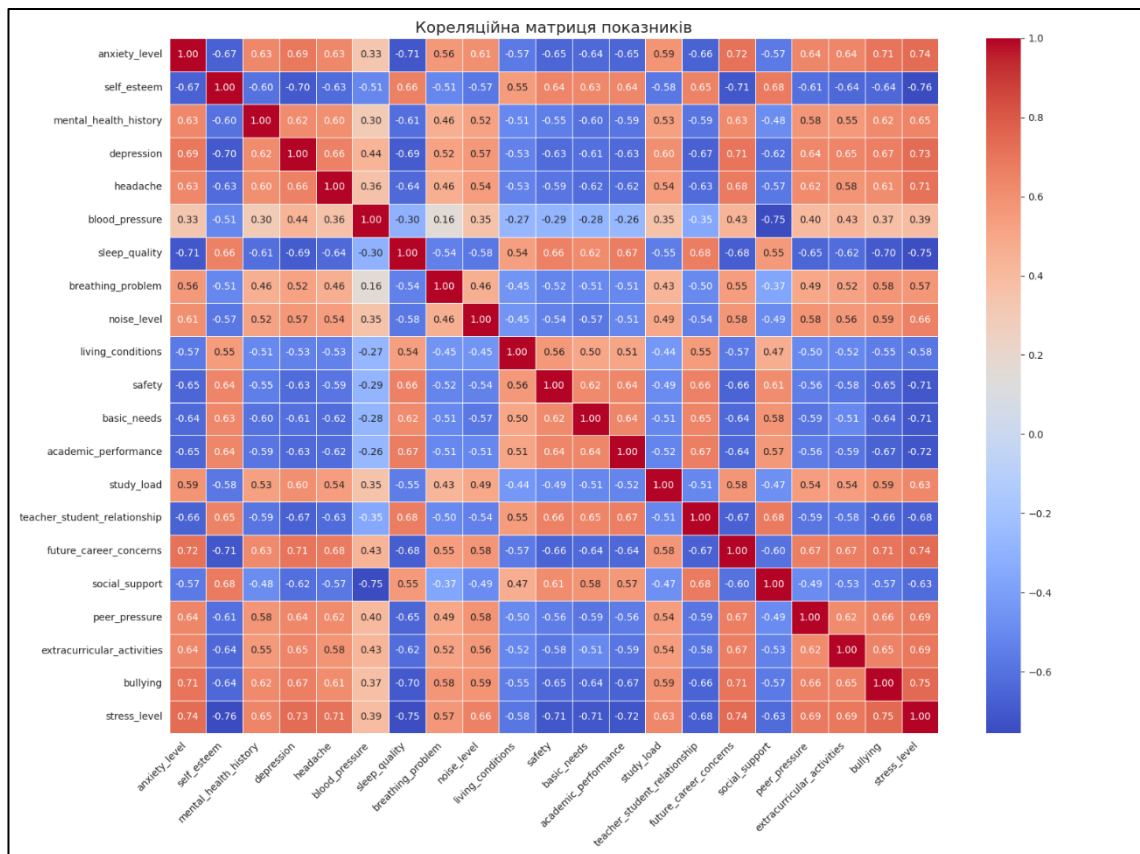


Рисунок Г.1 – Кореляційна матриця даних факторів рівня стресу студентів

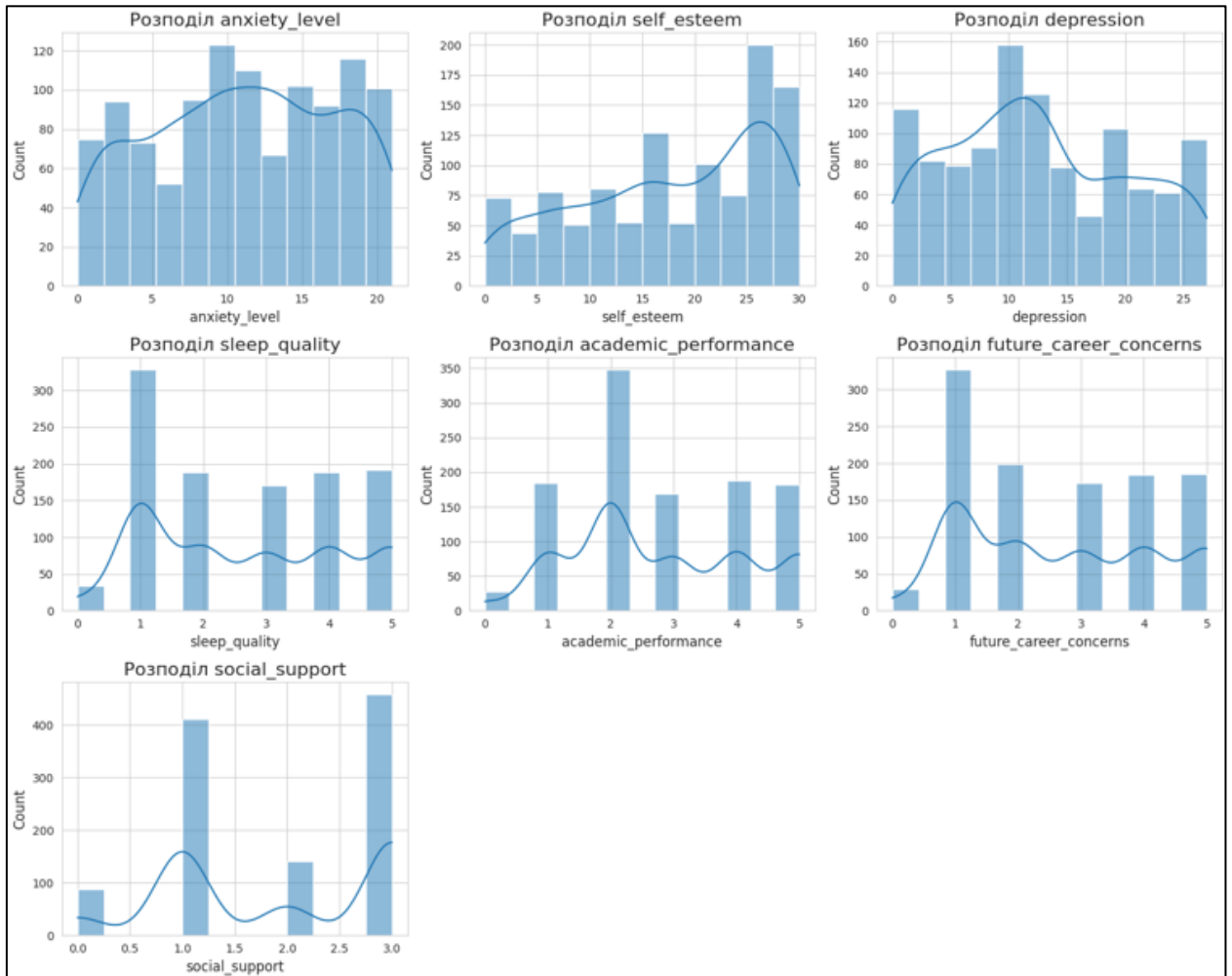


Рисунок Г.2 – Гістограми розподілу даних

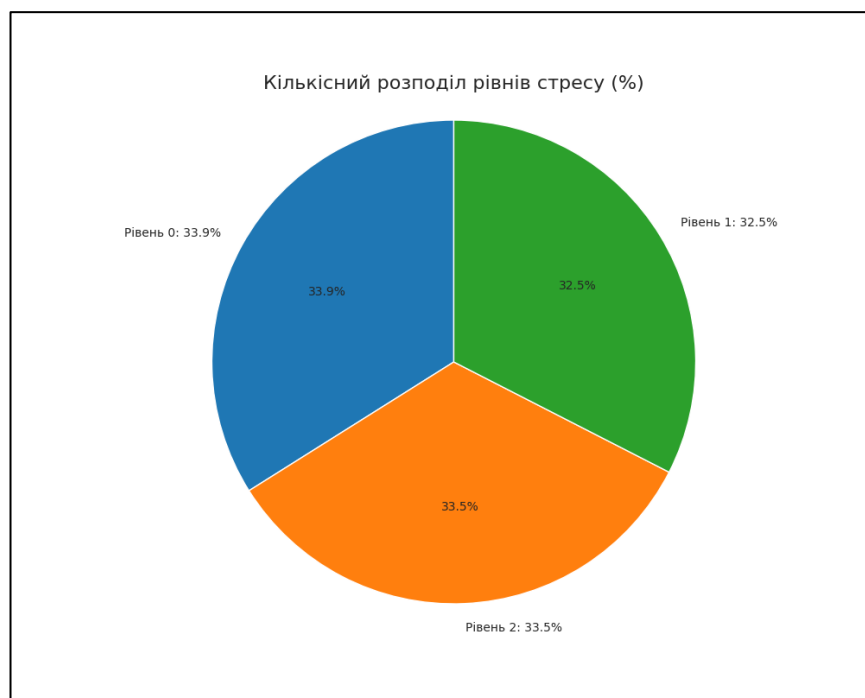


Рисунок Г.3 – Кількісний розподіл рівнів стресу у відсотках

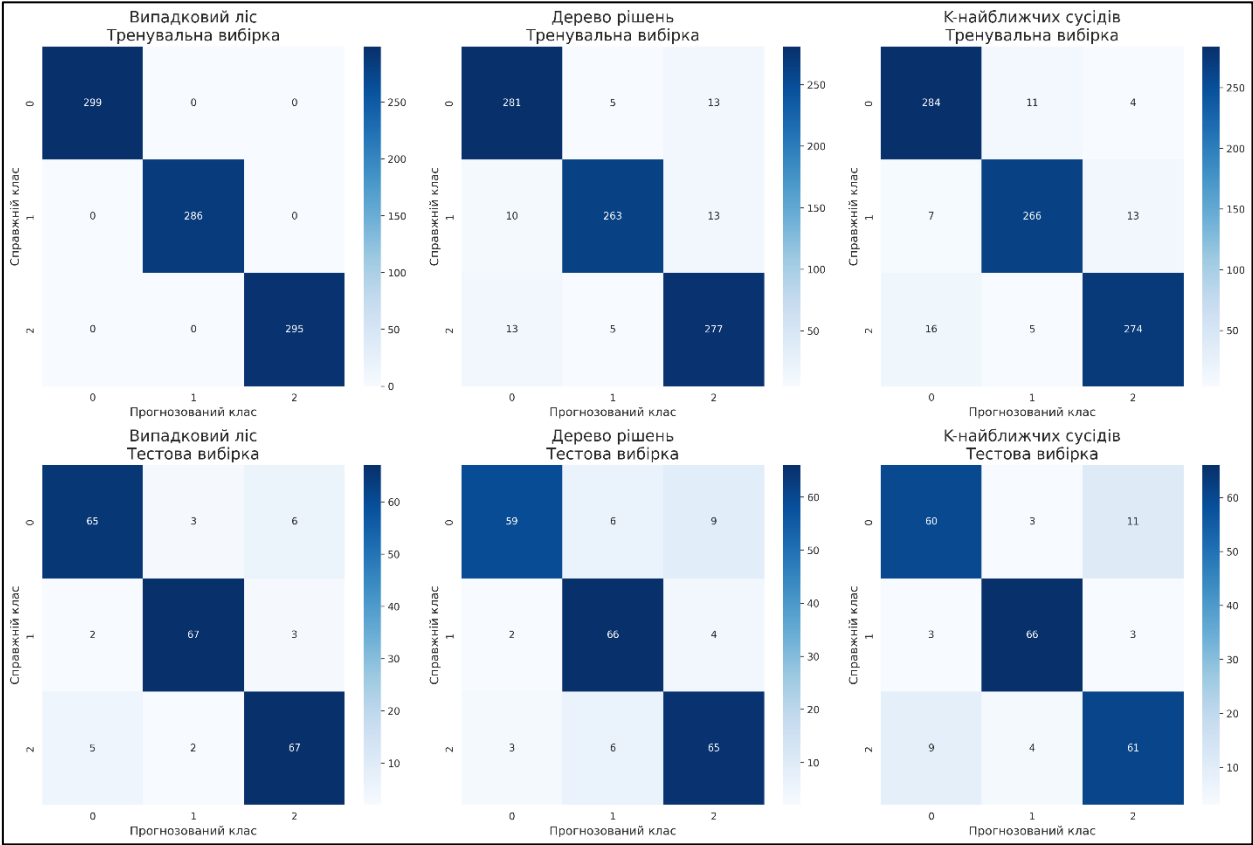


Рисунок Г.4 – Матриці плутанини для моделей класифікації

Модель	Train Accuracy	Train F1	Test Accuracy	Test F1	CV F1-score
Випадковий ліс	0.9795	0.9796	0.8818	0.8817	0.8707 ± 0.0218
Дерево рішень	0.9330	0.9330	0.8636	0.8633	0.8855 ± 0.0182
K-найближчих сусідів	0.9364	0.9364	0.8500	0.8498	0.8740 ± 0.0197

Рисунок Г.5 – Оцінка точності роботи моделей

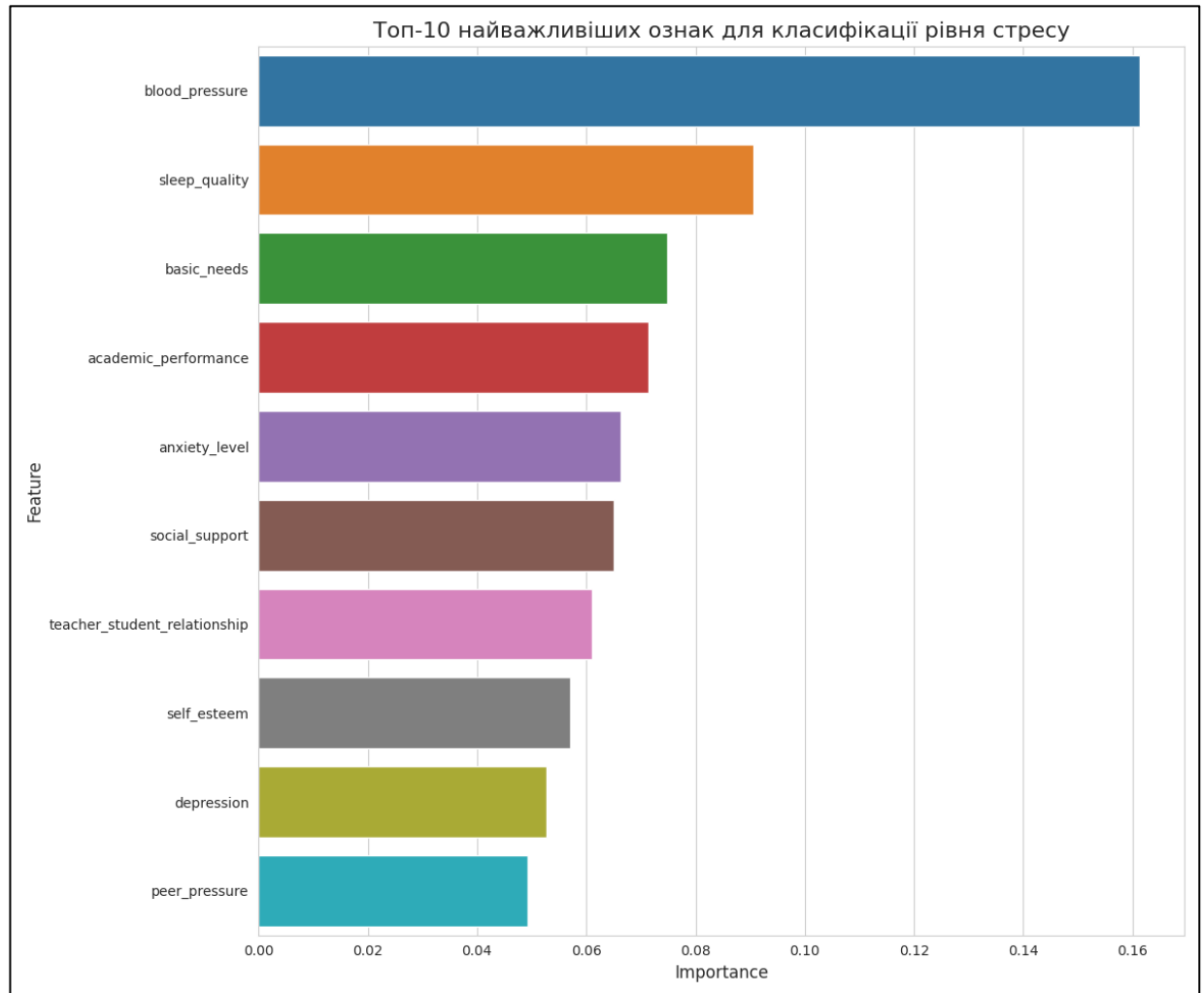


Рисунок Г.6 – Графік важливості ознак класифікації рівня стресу за моделлю Випадковий ліс