

# РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ ДЛЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПЕРЕДБАЧЕННЯ ВПЛИВУ СПОСОБУ ЖИТТЯ НА ЯКІСТЬ СНУ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Вінницький національний технічний університет

## Анотація

*Робота присвячена підготовці та розвідувальному аналізу даних для подальшого використання в інформаційній технології передбачення впливу способу життя на якість сну методами машинного навчання. Проведено аналіз структури датасету та його ознак, а також досліджено можливі залежності між якістю сну та факторами способу життя.*

**Ключові слова:** інформаційні технології, якість сну, машинне навчання, аналіз даних, передбачення, спосіб життя, розвідувальний аналіз.

## Abstract

*The paper is devoted to the preparation and exploratory analysis of data for further use in the information technology of predicting the impact of lifestyle on sleep quality using machine learning methods. The structure of the dataset and its features were analyzed, and possible dependencies between sleep quality and lifestyle factors were investigated.*

**Keywords:** information technology, sleep quality, machine learning, data analysis, prediction, lifestyle, intelligence analysis.

## Вступ

Сучасні інформаційні технології відіграють важливу роль у сфері медицини та здоров'я, зокрема у вивченні факторів, що впливають на якість сну. Сон є критично важливим компонентом фізичного та психоемоційного стану людини, а його якість може залежати від численних аспектів способу життя, таких як рівень фізичної активності, стрес, режим харчування та інші фактори.

Методи машинного навчання відкривають нові можливості для аналізу цих факторів і побудови моделей прогнозування порушень сну. Одним із ключових етапів створення таких моделей є розвідувальний аналіз даних (EDA), який дозволяє виявити закономірності, перевірити кореляції між змінними та підготувати дані до подальшого моделювання.

## Розвідувальний аналіз

Для проведення аналізу було обрано набір даних, що має назву «Sleep Health and Lifestyle Dataset» та опублікований користувачем «Laksika Tharmalingam» у відкритому доступі для загального використання [1]. Даний датасет містить 400 записів і 13 атрибутів, що охоплюють широкий спектр змінних, пов'язаних зі сном і щоденними звичками.

Цільовою змінною у цьому дослідженні є якість сну (Quality of Sleep), яка оцінюється за шкалою від 1 до 10. Вона відображає суб'єктивне сприйняття якості відпочинку та може залежати від багатьох факторів, таких як тривалість сну (Sleep Duration), рівень стресу (Stress Level), фізична активність (Physical Activity Level) та наявність розладів сну (Sleep Disorder).

Серед інших атрибутів, що можуть впливати на якість сну, варто виділити категорію ІМТ (індекс маси тіла) (BMI Category), артеріальний тиск (Blood Pressure), частоту серцевих скорочень (Heart Rate), кількість щоденних кроків (Daily Steps), а також демографічні характеристики: стать (Gender), вік (Age) та професію (Occupation).

Цей набір даних дозволяє проаналізувати взаємозв'язки між способом життя, фізичним станом і якістю сну, що дає змогу визначити ключові фактори, які впливають на здоровий сон. На рисунку 1 представлено фрагмент датасету «Sleep Health and Lifestyle Dataset», який містить основні атрибути.

|   | Gender | Age | Occupation           | Sleep Duration | Quality of Sleep | Physical Activity Level | Stress Level | BMI Category |
|---|--------|-----|----------------------|----------------|------------------|-------------------------|--------------|--------------|
| 0 | Male   | 27  | Software Engineer    | 6.1            | 6                | 42                      | 6            | Overweight   |
| 1 | Male   | 28  | Doctor               | 6.2            | 6                | 60                      | 8            | Normal       |
| 2 | Male   | 28  | Doctor               | 6.2            | 6                | 60                      | 8            | Normal       |
| 3 | Male   | 28  | Sales Representative | 5.9            | 4                | 30                      | 8            | Obese        |
| 4 | Male   | 28  | Sales Representative | 5.9            | 4                | 30                      | 8            | Obese        |

Рис. 1. Приклад атрибутів, що містить набір даних

Попередньо проведено очищення даних та усунення пропусків. Для аналізу було побудовано графік розподілу рівнів якості сну, яка дозволяє оцінити характер варіації цільової змінної якості сну (Quality of Sleep) представлено на рисунку 2.

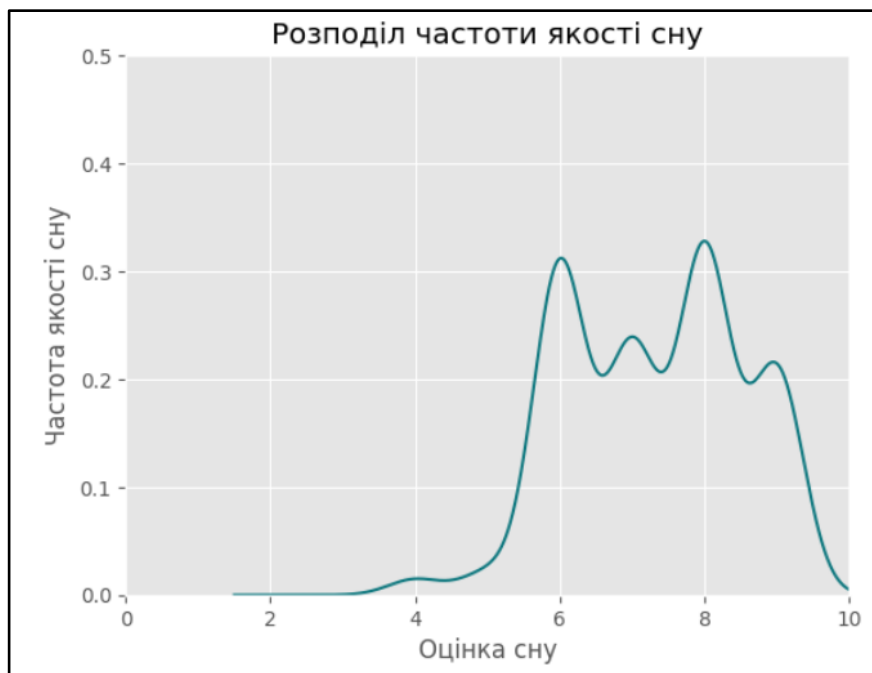


Рис. 2. Графік розподілу цільової величини якості сну (Quality of Sleep)

На графіку можна побачити, що розподіл є багатомодальним із найбільшим скупченням у діапазоні 6–9 балів. Низькі оцінки (0–4) зустрічаються рідко, що свідчить про переважно задовільну якість сну. Такий розподіл є корисним для подальшого аналізу, оскільки не має різких перекосів і дозволяє оцінювати фактори, що впливають на якість сну.

У рамках дослідження було проведено аналіз залежності між індексом маси тіла (ІМТ) та якістю сну. На рисунку 3 представлено графік, що ілюструє цей взаємозв'язок.



Рис. 3. Гістограма залежності між ІМТ та якістю сну

Первинний статистичний аналіз показав, що значення  $p$  (ймовірність отримати такі ж або ще більш екстремальні результати за умови, що нульова гіпотеза є істинною) є дуже малим, що вказує на статистично значущу залежність між ІМТ та якістю сну. Висунуто гіпотезу, що високий ІМТ може бути пов'язаний із погіршенням якості сну, тоді як нормальний або низький ІМТ сприяє її покращенню.

Дані результати є важливими для подальшого моделювання за допомогою методів машинного навчання, оскільки вони підтверджують можливий вплив фізіологічних параметрів на якість сну та можуть бути використані для створення прогнозних моделей.

На основі попереднього аналізу було побудовано кореляційну матрицю для аналізу взаємозв'язків між якістю сну та іншими змінними (рис. 3).

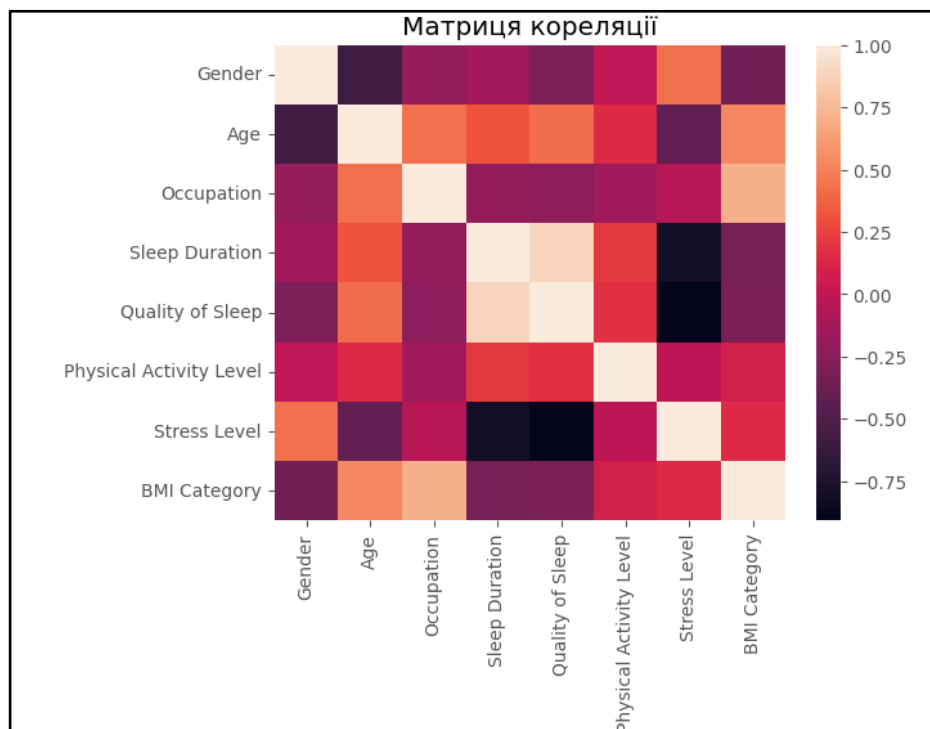


Рис. 4. Матриця кореляції

З аналізу кореляційної матриці (рис. 4) видно, що рівень стресу (*Stress Level*) має сильний негативний зв'язок із тривалістю (*Sleep Duration*) та якістю сну (*Quality of Sleep*), що підтверджує його негативний вплив на відпочинок. Водночас фізична активність (*Physical Activity Level*) позитивно корелює з якістю сну та допомагає знижувати рівень стресу. Категорія ІМТ (*BMI Category*) має помірний негативний зв'язок із якістю сну, що може свідчити про гірший сон у людей із вищим індексом маси тіла. Отримані результати підкреслюють важливість контролю стресу, активного способу життя та здорового режиму сну.

## Висновки

У ході розвідувального аналізу даних було досліджено взаємозв'язки між способом життя та якістю сну на основі набору даних «Sleep Health and Lifestyle Dataset». Аналіз показав, що рівень стресу має сильний негативний вплив на якість сну, тоді як фізична активність позитивно корелює з якістю відпочинку. Виявлено, що індекс маси тіла (BMI) має помірний негативний зв'язок із якістю сну, що свідчить про можливий негативний вплив надмірної ваги на сон.

Попередня обробка даних включала очищення та усунення пропусків, що дозволило покращити якість подальшого аналізу. Аналіз розподілу цільової змінної показав, що більшість спостережень знаходиться в діапазоні 6–9 балів за шкалою оцінки якості сну, що свідчить про переважно задовільний рівень відпочинку серед учасників вибірки.

Кореляційний аналіз підтвердив значущі залежності між основними характеристиками способу життя та якістю сну. Особливо помітним є негативний вплив підвищеного рівня стресу, що підкреслює важливість контролю психологічного стану для підтримання якісного відпочинку.

Отримані результати є важливими для подальшого застосування методів машинного навчання з метою прогнозування можливих порушень сну та розробки персоналізованих рекомендацій для його покращення.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Sleep Health and Lifestyle Dataset. Kaggle. 2024 [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>
2. Pandas Tutorial 2024 [Електронний ресурс] – Режим доступу: <https://www.w3schools.com/python/pandas/default.asp>
3. Matplotlib Pyplot Documentation. 2024 [Електронний ресурс]. – Режим доступу: [https://matplotlib.org/3.5.3/api/\\_as\\_gen/matplotlib.pyplot.html](https://matplotlib.org/3.5.3/api/_as_gen/matplotlib.pyplot.html)
4. A Complete Guide to Data Visualization in Python With Libraries & More. 2023 [Електронний ресурс]. – Режим доступу: <https://www.simplilearn.com/tutorials/python-tutorial/data-visualization-in-python>

**Білецька Мар'яна Володимирівна** – студентка групи СА-22б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця. e-mail: [marjnabilecka@gmail.com](mailto:marjnabilecka@gmail.com)

**Жуков Сергій Олександрович** – к.т.н., доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, e-mail: [sazhukov@gmail.com](mailto:sazhukov@gmail.com)

**Biletska Mariana** – student of Faculty of Intellectual Information Technologies and Automation, SA-22b, Vinnytsia National Technical University, Vinnytsia, e-mail [marjnabilecka@gmail.com](mailto:marjnabilecka@gmail.com)

**Zhukov Serhii O.** - Ph.D., Assistant Professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: [sazhukov@gmail.com](mailto:sazhukov@gmail.com)