УДК 004.93.2:159.9:331.44

А. І. Волянський
О. О. Войцеховська

РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ РИЗИКУ ПРОФЕСІЙНОГО ВИГОРАННЯ ПРАЦІВНИКІВ

Вінницький національний технічний університет

Анотація

В роботі проведено розвідувальний аналіз даних ризику професійного вигорання працівників для виявлення факторів, які найбільше впливають на рівень вигорання. Проведено аналіз структури датасету, перевірку пропущених значень, дослідження розподілу цільової змінної Burn Rate, а також аналіз зв'язків між рівнем вигорання та характеристиками працівників. Встановлено, що найбільший зв'язок із рівнем професійного вигорання мають показники Mental Fatigue Score, Resource Allocation та Designation.

Ключові слова: розвідувальний аналіз даних, машинне навчання, професійне вигорання, працівники.

Abstract

The paper presents an exploratory data analysis of employee professional burnout risk in order to identify the factors that have the greatest influence on burnout level. The structure of the dataset, missing values, distribution of the target variable Burn Rate, and relationships between burnout level and employee characteristics were analyzed. It was found that Mental Fatigue Score, Resource Allocation, and Designation have the strongest relationship with the level of professional burnout.

Keywords: exploratory data analysis, machine learning, professional burnout, employees.

Вступ

Професійне вигорання є актуальною проблемою сучасного ринку праці, оскільки воно безпосередньо пов'язане з продуктивністю працівників, якістю виконання професійних обов'язків, рівнем залученості персоналу та ризиком звільнення. За визначенням Всесвітньої організації охорони здоров'я, вигорання розглядається як професійне явище, що виникає внаслідок хронічного стресу на робочому місці, який не був успішно подоланий. В ICD-11 воно не класифікується як медичний стан, але описується як фактор, пов'язаний із професійною діяльністю та станом здоров'я людини [1] .

У сучасних організаціях аналіз ризику вигорання набуває особливого значення, оскільки дозволяє своєчасно виявляти групи працівників із підвищеним навантаженням, оцінювати вплив умов праці та приймати управлінські рішення щодо покращення робочого середовища. Використання методів аналізу даних дає змогу не лише описати поточний стан працівників, а й виявити приховані закономірності між організаційними, психологічними та робочими факторами.

Метою даного дослідження є проведення розвідувального аналізу даних ризику професійного вигорання працівників та визначення ознак, які мають найбільший зв'язок із цільовою змінною Burn Rate.

Результати дослідження

Для проведення аналізу було обрано датасет «Are Your Employees Burning Out?», який має відкритий доступ на платформі Kaggle. Набір даних призначений для аналізу та передбачення показника Burn Rate, що характеризує рівень професійного вигорання працівника. У наборі даних наведено інформацію про працівників організації, зокрема дату приєднання до компанії, стать, тип компанії, наявність умов для роботи з дому, рівень посади, розподіл ресурсів, показник ментальної втоми та рівень вигорання. Kaggle описує цей датасет як набір для розуміння того, яким буде Burn Rate працівника в організації [2] .

Приклад структури даних представлено на рисунку 1.

	Employee ID	Gender	Company Type	WFH Setup Available	Designation	Resource Allocation	Mental Fatigue Score	Burn Rate
0	fffe32003000360033003200	Female	Service	No	2.0	3.0	3.8	0.16
1	fffe3700360033003500	Male	Service	Yes	1.0	2.0	5.0	0.36
2	fffe31003300320037003900	Female	Product	Yes	2.0	NaN	5.8	0.49
3	fffe32003400380032003900	Male	Service	Yes	1.0	1.0	2.6	0.20
4	fffe31003900340031003600	Female	Service	No	3.0	7.0	6.9	0.52

Рис.1. Фрагмент набору даних ризику професійного вигорання працівників.

Датасет містить 22750 записів та 9 ознак. До основних змінних набору даних належать:

– Employee ID: унікальний ідентифікатор працівника;
– Gender: стать працівника;
– Company Type: тип компанії;
– WFH Setup Available: наявність можливості роботи з дому;
– Designation: рівень посади працівника;
– Resource Allocation: рівень залучення або навантаження ресурсів;
– Mental Fatigue Score: показник ментальної втоми;
– Burn Rate: цільова змінна, що відображає рівень професійного вигорання.

На першому етапі було проведено перевірку пропущених значень (рис.2). Встановлено, що пропуски наявні у трьох змінних: Resource Allocation, Mental Fatigue Score та Burn Rate. Для ознаки Resource Allocation пропущено 1381 значення, що становить 6,07% від загальної кількості записів. Для Mental Fatigue Score пропущено 2117 значень, або 9,31%. Для цільової змінної Burn Rate пропущено 1124 значення, або 4,94%. Загалом хоча б один пропуск міститься у 4160 записах, що становить 18,29% від усього набору даних. Це свідчить про необхідність попередньої обробки даних перед побудовою моделей машинного навчання.

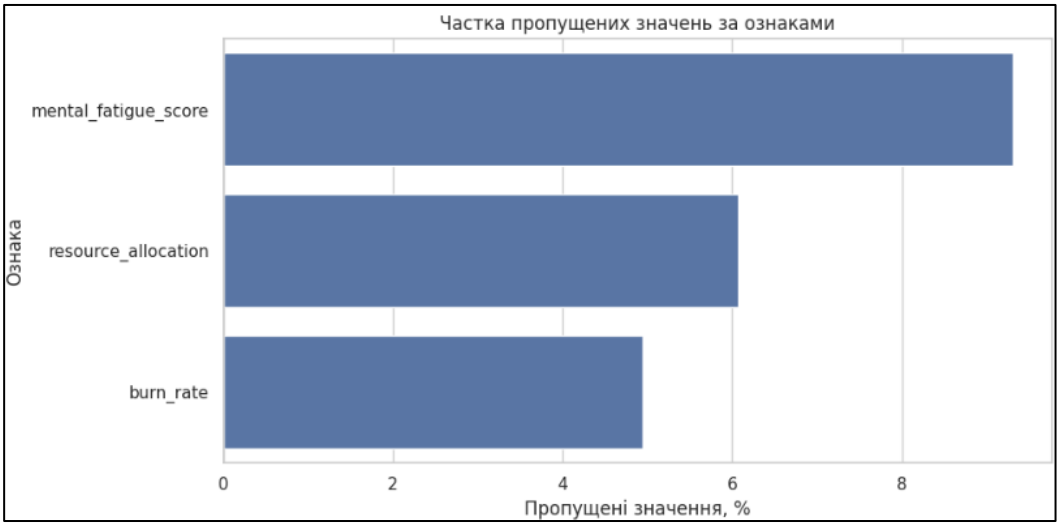

Рис. 2. Аналіз пропущених значень у наборі даних.

Далі було проаналізовано розподіл цільової змінної Burn Rate (рис. 3). Значення цього показника знаходяться в діапазоні від 0 до 1. Середнє значення Burn Rate становить 0,452, медіанне значення 0,45, а стандартне відхилення 0,198. Це свідчить про те, що більшість працівників мають середній рівень професійного вигорання, однак у вибірці також наявні працівники як з дуже низькими, так і з високими значеннями Burn Rate.

Для інтерпретації рівня ризику професійного вигорання цільову змінну можна поділити на три групи: низький, середній та високий рівень ризику. При поділі за фіксованими інтервалами було встановлено, що 28,1% працівників належать до групи з низьким рівнем вигорання, 57,6% до групи із середнім рівнем, а 14,3% до групи з високим рівнем професійного вигорання. Отже, найбільшу частку становлять працівники із середнім рівнем ризику, однак група з високим рівнем вигорання також є суттєвою та потребує окремої уваги.

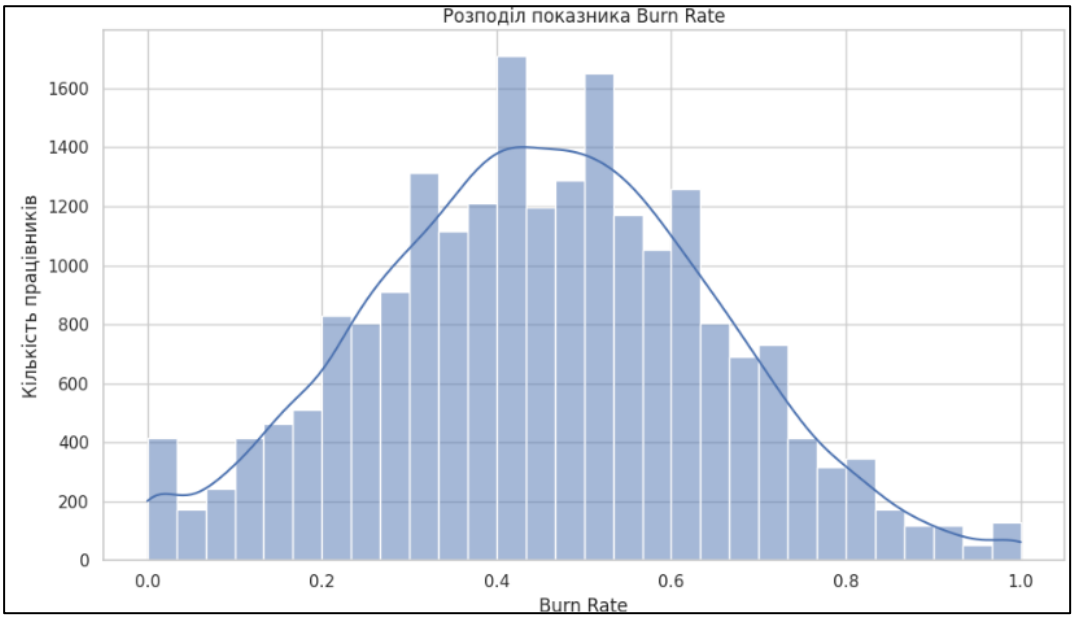

Рис. 3. Розподіл цільової змінної Burn Rate

Було також проаналізовано розподіл категоріальних ознак. За статтю вибірка є відносно збалансованою: 52,34% становлять жінки, 47,66% чоловіки. За типом компанії переважають працівники сервісних компаній 65,2%, тоді як працівники продуктових компаній становлять 34,8%. Можливість роботи з дому мають 54,02% працівників, тоді як 45,98% такої можливості не мають.

Окремий аналіз середнього рівня вигорання за категоріальними ознаками показав, що середнє значення Burn Rate для чоловіків становить 0,484, а для жінок 0,423. Це може свідчити про дещо вищий середній рівень вигорання серед чоловіків у межах даного набору даних. Водночас тип компанії майже не впливає на середній рівень вигорання: для продуктових компаній середнє значення Burn Rate становить 0,451, а для сервісних 0,453.

Більш помітна різниця спостерігається для ознаки WFH Setup Available (рис. 4). Працівники, які не мають можливості працювати з дому, мають середній Burn Rate на рівні 0,518, тоді як для працівників із доступною WFH-інфраструктурою середній показник становить 0,396. Це свідчить про потенційний зв'язок між гнучкістю робочого середовища та нижчим рівнем професійного вигорання.

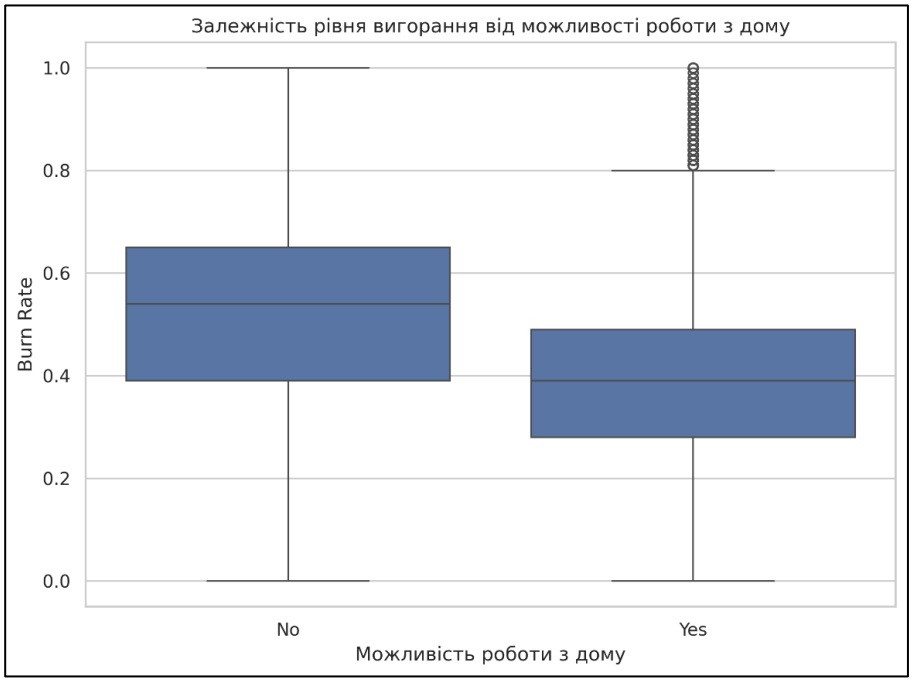

Рис. 4. Залежність рівня вигорання від наявності можливості роботи з дому

Для визначення взаємозв'язку між числовими ознаками й цільовою змінною було побудовано матрицю кореляції (рис. 5). Кореляційний аналіз є типовим інструментом розвідувального аналізу, що дозволяє оцінити парні зв'язки між числовими змінними; у pandas для цього використовується метод DataFrame.corr() [5] .

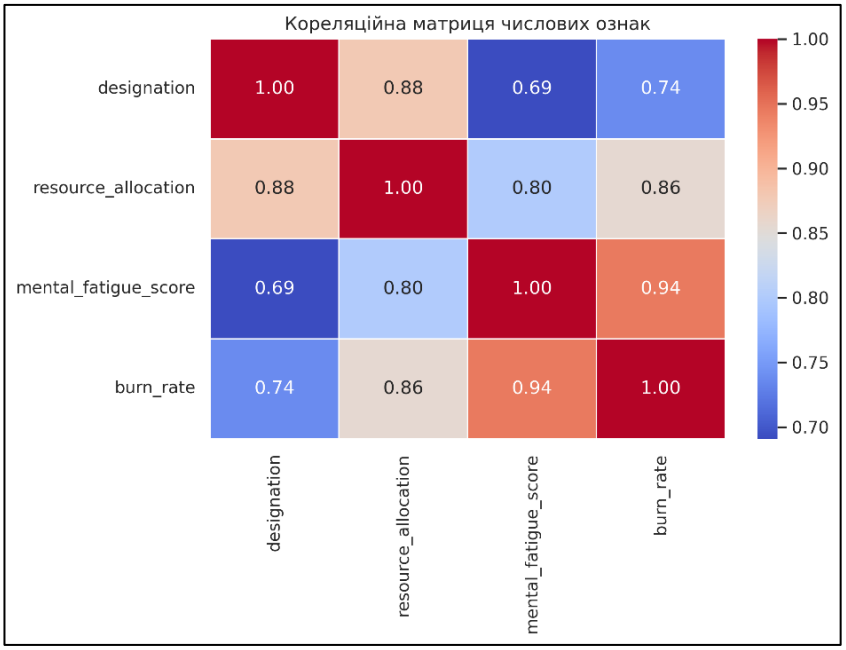

Рис. 5. Кореляційна матриця числових ознак

Найвищий коефіцієнт кореляції з рівнем вигорання має ознака Mental Fatigue Score 0,945. Це означає, що рівень ментальної втоми є найсильніше пов'язаним із професійним вигоранням. Другим за впливовістю показником є Resource Allocation з коефіцієнтом кореляції 0,856. Це свідчить про те, що зростання рівня навантаження працівника супроводжується зростанням ризику професійного вигорання. Ознака Designation також має достатньо високий зв'язок із Burn Rate, її коефіцієнт кореляції становить 0,738.

Додатково було досліджено залежність Burn Rate від рівня посади (рис. 6). Для працівників з найнижчим рівнем Designation середнє значення Burn Rate становить 0,151, тоді як для працівників з найвищим рівнем Designation воно зростає до 0,857. Це може свідчити про те, що вищий рівень посади пов'язаний із більшою відповідальністю, інтенсивністю роботи та, відповідно, підвищеним ризиком професійного вигорання.

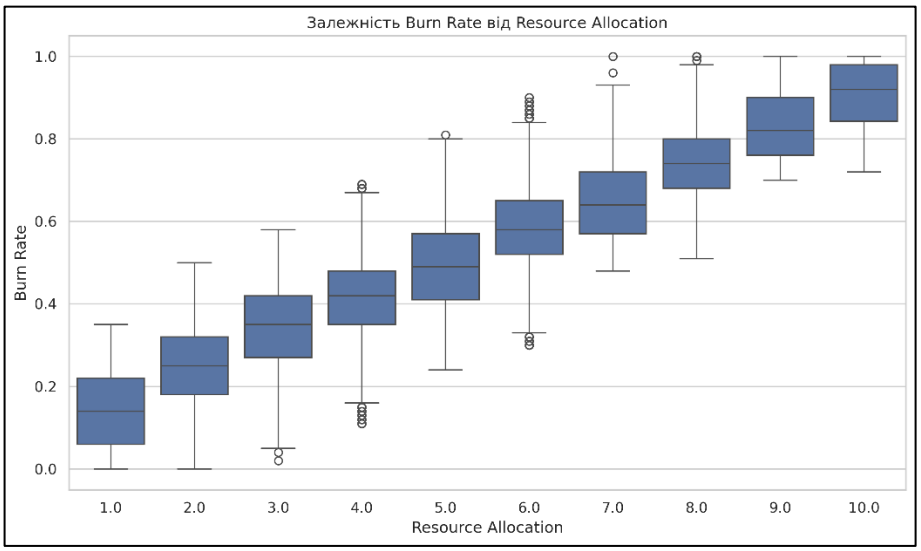

Рис. 6. Залежність Burn Rate від Resource Allocation та Mental Fatigue Score

Аналогічна залежність спостерігається для Resource Allocation. При мінімальному значенні цієї ознаки середній Burn Rate становить всього 0,138, а при максимальному 0,904. Це підтверджує, що розподіл ресурсів та рівень навантаження є одним із ключових факторів, пов'язаних із ризиком професійного вигорання.

Висновки

У роботі проведено розвідувальний аналіз даних ризику професійного вигорання працівників. Було проаналізовано структуру датасету, типи ознак, пропущені значення, розподіл цільової змінної Burn Rate та зв'язки між характеристиками працівників і рівнем вигорання.

Встановлено, що найбільший зв'язок із рівнем професійного вигорання мають ознаки Mental Fatigue Score, Resource Allocation та Designation. Найсильніша кореляція спостерігається між Burn Rate і Mental Fatigue Score, що підтверджує важливість психологічного стану працівника при оцінюванні ризику професійного вигорання. Також виявлено, що працівники без можливості роботи з дому мають вищий середній рівень Burn Rate, ніж працівники, для яких така можливість доступна.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. World Health Organization. Burn-out an occupational phenomenon. 2019. [Електронний ресурс]. – Режим доступу: https://www.who.int/standards/classifications/frequently-asked-questions/burn-out-an-occupational-phenomenon
2. Are Your Employees Burning Out? Kaggle Dataset. [Електронний ресурс]. – Режим доступу: https://www.kaggle.com/datasets/blurredmachine/are-your-employees-burning-out
3. Maslach C., Leiter M. P. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry. 2016. Vol. 15, No. 2. P. 103–111. doi: 10.1002/wps.20311
4. Мокін В. Б., Дратований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник. Вінниця : ВНТУ, 2024. URL: https://docs.vntu.edu.ua/card.php?id=8163
5. Pandas Documentation. DataFrame.corr. [Електронний ресурс]. – Режим доступу: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html
6. Scikit-learn Documentation. Imputation of missing values. [Електронний ресурс]. – Режим доступу: https://scikit-learn.org/stable/modules/impute.html

Волянський Андрій Ігорович – студент групи СА-22б, Факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: v0lyndr@gmail.com

Войцеховська Ольга Олександрівна – PhD, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, e-mail: olgav1085@gmail.com

Volianskyi Andii I. – student of group SA-22b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: v0lyndr@gmail.com

Voitsekhovska Olha O. – PhD, Associate Professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia,