ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІСЬКОГО ГОСПОДАРСТВА ІМЕНІ О. М. БЕКЕТОВА

Пояснювальна записка
до кваліфікаційної роботи бакалавра

на тему: Дослідження плинності кадрів в ІТ-компаніях методами машинного навчання

Виконав: студент _4_ курсу,
групи КН 2021-1
спеціальності
_122 Комп'ютерні науки_____
(шифр і назва спеціальності)

__Михайло ГАВРИЛОВ_____
(прізвище та ініціали)

Керівник: __Юрій ПАХОМОВ_____
(прізвище та ініціали)

Рецензент: Наталія Сізова_____
(прізвище та ініціали)

м. Харків – 2025 рік

Харківський національний університет міського господарства імені О. М. Бекетова
(повне найменування закладу вищої освіти)
Навчально-науковий Інститут енергетичної, інформаційної та транспортної інфраструктури
Кафедра комп'ютерних наук та інформаційних технологій
Рівень вищої освіти перший (бакалаврський)
Спеціальність 122 Комп'ютерні науки
(шифр і назва)

ЗАТВЕРДЖУЮ
Завідувач кафедри КНтаІТ

Марина НОВОЖИЛОВА
«27» 06 2025 року

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Гаврилову Михайлу Дмитровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження плинності кадрів в ІТ-компаніях методами машинного навчання

керівник роботи Пахомов Ю.В., к.т.н., доцент кафедри КНтаІТ
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом закладу вищої освіти від «09» травня 2025 р. № 341-03.

2. Термін подання студентом роботи 16.06.2025р

3. Вихідні дані до роботи датасет про плинність кадрів в компанії

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) проаналізувати причини та наслідки плинності кадрів в ІТ-секторі; дослідити можливості HR-аналітики та застосування методів машинного навчання для передбачення плинності; здійснити попередній аналіз даних з наявного датасету; реалізувати моделі машинного навчання; провести порівняльний аналіз ефективності моделей; розробити інтерфейс користувача для демонстрації роботи системи прогнозування; надати практичні рекомендації для HR-менеджменту на основі результатів прогнозування

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)
Презентація на 19 слайдах

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Юрій ПАХОМОВ, к.т.н., доцент каф. КНтаІТ	26.05.25	02.06.25
2	Юрій ПАХОМОВ, к.т.н., доцент каф. КНтаІТ	01.06.25	09.06.25
3	Юрій ПАХОМОВ, к.т.н., доцент каф. КНтаІТ	05.06.25	13.06.25
4	Вікторія МАЛИШЕВА, к.т.н., доцент кафедри безпеки життєдіяльності	06. 06.25	16. 06.25

7. Дата видачі завдання 26.05.2025р

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк викон. етапів роботи	Примітка
1	Вибір теми кваліфікаційної роботи	26.05.2025	Викон.
2	Затвердження тем, наукових керівників, завдань та календарного плану підготовки кваліфікаційної роботи	26.05.2025	Викон.
3	Написання І розділу	02.06.2025	Викон.
4	Написання ІІ розділу	09.06.2025	Викон.
5	Написання ІІІ розділу	13.06.2025	Викон.
6	Написання IV розділу Охорона праці	16.06.2025	Викон.
8	Подання кваліфікаційної роботи керівнику	17.06.2025	Викон.
9	Робота по усуненню зауважень керівника, уточнення і доповнення практичного матеріалу, оформлення додатків до роботи	18.06.2025	Викон.
11	Захист матеріалів кваліфікаційної роботи на засіданні кафедри	20.06.2025	Викон.
12	Офіційний захист матеріалів кваліфікаційної роботи на засіданні екзаменаційної комісії	28.06.2025	Викон.

Студент _____ (підпис) Гаврилов М.Д. (прізвище та ініціали)

Керівник роботи _____ (підпис) Пахомов Ю.В. (прізвище та ініціали)

АНОТАЦІЯ

Структура та обсяг роботи. Пояснювальна записка кваліфікаційної роботи бакалавра студента групи КН 2021-1 спеціальності 122 Комп'ютерні науки Гаврилова Михайла Дмитровича за темою «Дослідження плинності кадрів в IT-компаніях методами машинного навчання» складається з 4 розділів, містить 39 рисунків, 4 таблиць, 15 джерел.

Кваліфікаційну роботу бакалавра присвячено розробці та дослідженню моделей машинного навчання для прогнозування плинності кадрів в IT-компаніях.

У розділі «Аналіз проблем плинності кадрів в IT-компаніях» розглянуто опис предметної області, проведено аналіз проблеми плинності кадрів в IT-компаніях.

Розділ «Методи машинного навчання для передбачення плинності кадрів» містить огляд математичного та алгоритмічного забезпечення для вирішення задачі.

У розділі «Розробка та дослідження моделей машинного навчання для передбачення плинності кадрів» наведено особливості програмної реалізації, щодо створення та дослідження моделей машинного навчання для прогнозування плинності кадрів в IT-компаніях.

У розділі «Охорона праці» розглянуто регулювання питань охорони праці на законодавчому рівні, виявлення потенційних небезпек, дослідження ризику реалізації потенційних небезпек та розробка заходів щодо їх попередження.

Ключові слова: ПЛИННІСТЬ КАДРІВ, HR-МЕНЕДЖМЕНТ, МАШИННЕ НАВЧАННЯ, МОДЕЛІ, ПРОГНОЗУВАННЯ.

ANNOTATION

Structure and scope of work. Explanatory note of the bachelor's qualification work of a student of group KN 2021-1, specialty 122 Computer Science, Gavrilov Mikhail Dmitrovich, on the topic "Research of personnel turnover in IT companies using machine learning methods" consists of 4 sections, contains 39 figures, 4 tables, 15 sources.

The bachelor's qualification work is devoted to the development and research of machine learning models for predicting personnel turnover in IT companies.

The section "Analysis of personnel turnover problems in IT companies" considers a description of the subject area, analyzes the problem of personnel turnover in IT companies.

The section "Machine learning methods for predicting personnel turnover" contains an overview of the mathematical and algorithmic support for solving the problem.

The section “Development and research of machine learning models for predicting staff turnover” presents the features of software implementation, regarding the creation and research of machine learning models for predicting staff turnover in IT companies.

The section “Occupational safety” considers the regulation of occupational safety issues at the legislative level, identification of potential hazards, research of the risk of realization of potential hazards and development of measures to prevent them.

Keywords: STAFF TURNOVER, HR MANAGEMENT, MACHINE LEARNING, MODELS, FORECASTING.

ЗМІСТ

ВСТУП 8
РОЗДІЛ 1 АНАЛІЗ ПРОБЛЕМ ПЛИННОСТІ КАДРІВ В ІТ-КОМПАНІЯХ 10
1.1 Плинність кадрів 10
1.2 Причини та наслідки плинності кадрів 13
1.3 Залученість та утримання співробітників 15
1.4 ІТ-ринок праці в Україні 16
1.5 HR-аналітика та методи штучного інтелекту 18
Висновки за розділом 21
РОЗДІЛ 2 МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ ПЕРЕДБАЧЕННЯ ПЛИННОСТІ КАДРІВ 22
2.1 Огляд методів машинного навчання 22
2.2 Загальна схема рішення 26
Висновки за розділом 28
РОЗДІЛ 3 РОЗРОБКА ТА ДОСЛІДЖЕННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ПЕРЕДБАЧЕННЯ ПЛИННОСТІ КАДРІВ 29
3.1 Використовувані програмні засоби розробки та опис датасету 29
3.2 Реалізація моделей машинного навчання 42
3.2.1 Випадковий ліс 42
3.2.2 Дерева рішень 43
3.2.3 Градієнтний бустинг 45
3.2.4 Логістична регресія 46
3.2.5 Багатошаровий персептрон 47
3.2.6 Порівняння результатів 50
3.3 Інтерфейс системи прогнозування плинності кадрів 53
3.4 Оцінка ефективності системи прогнозування, а також рекомендації для HR-менеджменту 54
Висновки за розділом 55
РОЗДІЛ 4 ОХОРОНА ПРАЦІ 56
4.1 Регулювання питань охорони праці на законодавчому рівні 56

4.2 Виявлення потенційних небезпек стосовно об’єкту проектування 58
4.3 Дослідження ризику реалізації потенційних небезпек на об’єкті проектування та розробка заходів щодо їх попередження 60
4.4 Висновки по розділу 63
ВИСНОВКИ 64
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 65

ВСТУП

Актуальність теми дослідження. В умовах стрімкого розвитку IT-індустрії та зростаючої конкуренції за кваліфіковані кадри, плинність персоналу стала однією з найгостріших проблем для IT-компаній. Через вирішальну роль, яку працівники відіграють в успіху організації, вимірювання плинності кадрів стало одним з найважливіших показників, на яких зосереджуються підприємства в сучасну епоху. Час від часу плинність кадрів може виникати через неминучі обставини, такі як переїзд у віддалене місце тощо. Але коли плинність кадрів починає створювати дірки в кишенях організації, необхідно уважно стежити за ситуацією. Часте звільнення цінних співробітників призводить до втрат знань, порушення бізнес-процесів та збільшення витрат на підбір і адаптацію нових фахівців. Застосування методів машинного навчання у HR-аналітиці дозволяє прогнозувати ризики звільнення персоналу та вчасно вживати заходів для їх утримання. Тому актуальність цього дослідження зумовлена необхідністю створення ефективних, точних і автоматизованих інструментів прогнозування плинності кадрів на основі сучасних технологій штучного інтелекту.

Мета дослідження – розробити та дослідити систему прогнозування плинності кадрів в IT-компаніях з використанням моделей машинного навчання.

Для досягнення мети необхідно виконати наступні завдання:

1 Проаналізувати причини та наслідки плинності кадрів в IT-секторі.

2 Дослідити можливості HR-аналітики та застосування методів машинного навчання для передбачення плинності.

3 Здійснити попередній аналіз даних (EDA) з наявного датасету.

4 Реалізувати моделі машинного навчання: дерево рішень, випадковий ліс, градієнтний бустинг, логістична регресія, багатошаровий персептрон.

5 Провести порівняльний аналіз ефективності моделей.

6 Розробити інтерфейс користувача для демонстрації роботи системи прогнозування.

7 Надати практичні рекомендації для HR-менеджменту на основі результатів прогнозування.

Об’єкт дослідження – процес прийняття управлінських рішень у сфері кадрового менеджменту IT-компанії, пов’язаний із прогнозуванням плинності персоналу.

Предмет дослідження – моделі машинного навчання та методи їх застосування для аналізу та прогнозування плинності кадрів на основі історичних даних.

РОЗДІЛ 1
АНАЛІЗ ПРОБЛЕМ ПЛИННОСТІ КАДРІВ В ІТ-КОМПАНІЯХ

1.1 Плинність кадрів

Плинність кадрів – це плинність працівників у компанії. Її показник виражається у % як відношення кількості припинених трудових відносин (трудових договорів) до середньої кількості персоналу за одиницю часу (зазвичай рік). Вона може бути добровільною (навмисною) , що включає працівників, які вирішили звільнитися, або вимушеною (ненавмисною) , що включає працівників, звільнених роботодавцем.

Загалом, певний рівень плинності кадрів є неминучим, природним і не повинен розглядатися як негативне явище. На його рівень впливають різні фактори, наприклад, розмір компанії та навколишнє середовище, сектор і регіон, в якому працює компанія, а також макроекономічні показники, такі як безробіття.

Природний рівень плинності кадрів низький і часто спричинений факторами, на які ми не можемо вплинути без збору та аналізу даних про компанію. Відсутність відповідних даних та аналітики для визначення можливостей для покращення підтверджує думку, що рівень плинності кадрів нижче 10%-12% є цілком прийнятним. Це правда, якщо він нижчий за середній показник по ринку. Однак, те, що добре, завжди можна покращити. За правильного підходу ми можемо отримати конкурентну перевагу в довгостроковій перспективі. Плинність кадрів не виникає сама по собі з нічого, а є наслідком того, як компанія працює протягом усього життєвого циклу співробітника (рисунок 1.1). Це результат процесів, які починаються до того, як нові співробітники приєднуються до компанії, і не закінчуються, коли вони звільняються.

Виконання вищезазначених процесів впливає на рішення працівника почати думати про роботу в певній компанії, стати її фізичною частиною,

працювати на (або навіть вище) необхідного рівня та залишатися в фірмі протягом тривалого часу, в деяких випадках до виходу на пенсію, та, як випускник, ділитися досвідом з молодим поколінням через наставництво та коучинг.

Плинність кадрів є наслідком кількох видів діяльності в компанії та залежить від рівня якості їх виконання. Кожна компанія специфічна та змінюється з часом. Це безпосередньо впливає на життєвий цикл працівника. Якщо ми зможемо збирати, аналізувати та оцінювати дані на окремих етапах життєвого циклу працівника в компанії, ми зможемо оцінити якість виконуваних процесів, а також виявити причини, які викликають невдоволення працівника та зрештою призводять до його звільнення з компанії (рисунок 1.2).

Рисунок 1.1 – Життєвий цикл працівника

Плинність кадрів є наслідком кількох видів діяльності та досвіду.

Рисунок 1.2 – Досвід працівника

Ключова відмінність між плинністю кадрів та звільненням полягає в їхніх наслідках. Хоча обидва види звільнення передбачають звільнення співробітників з компанії, звільнення в бізнесі призводить до того, що вакансії залишаються незаповненими, тоді як плинність кадрів вимагає заміни.

Плинність кадрів буває двох основних форм: добровільна та вимушена. Добровільна плинність кадрів виникає, коли працівники вирішують звільнитися за власним бажанням. З іншого боку, вимушена плинність кадрів відбувається, коли організації просять працівників звільнитися.

1. Вимушений плин кадрів. У контексті вимушеної плинності кадрів, працівники можуть звільнитися з таких причин:

– тимчасове або постійне припинення;

– скорочення або скорочення штату робочої сили;

– звільнення в компанії.

Звільнення досвідчених експертів з програмного забезпечення може порушити існуючі ініціативи, створити прогалини в спеціалізованих знаннях та вплинути на креативне середовище розробки програмного забезпечення. Крім того, через швидкі темпи технічних інновацій, втрата значних учасників може вплинути на моральний дух та згуртованість технічних команд.

2. Добровільна плинність кадрів. На противагу цьому, добровільна плинність кадрів відбувається, коли працівники вирішують звільнитися, часто з таких причин, як:

– прагнення до кращих кар'єрних можливостей;

–переїзд на роботу;

– особисті мотиви, такі як подальше навчання, подорожі, проблеми зі здоров'ям тощо;

– обираючи достроковий вихід на пенсію;

– невдоволення своїм кар'єрним шляхом.

Раптові звільнення можуть порушити комунікацію між командами розробників програмного забезпечення та технічних фахівців, затримати проекти та створити тривогу в робочому середовищі. Крім того, втрата

конкретних технічних знань, які мають співробітники, що звільняються, може загальмувати прогрес проекту та знизити загальну ефективність команди.

1.2 Причини та наслідки плинності кадрів

Відтік працівників може бути спровокований поєднанням внутрішніх та зовнішніх факторів. Усередині компанії працівники можуть звільнятися через незадоволення роботою, відсутність можливостей для зростання або конфлікти з керівництвом. Ззовні цьому може сприяти привабливість кращих пропозицій, покращений баланс між роботою та особистим життям або кар'єрний ріст в іншому місці.

Згідно зі звітом ICONIC про майбутнє праці, 83% технічних працівників звільняються через незадоволення роботою, 78% – через незадоволення кар'єрним зростанням та професійним розвитком, 75% – через проблеми із заробітною платою або компенсацією, а 73% – через професійне вигорання (рисунок 1.3).

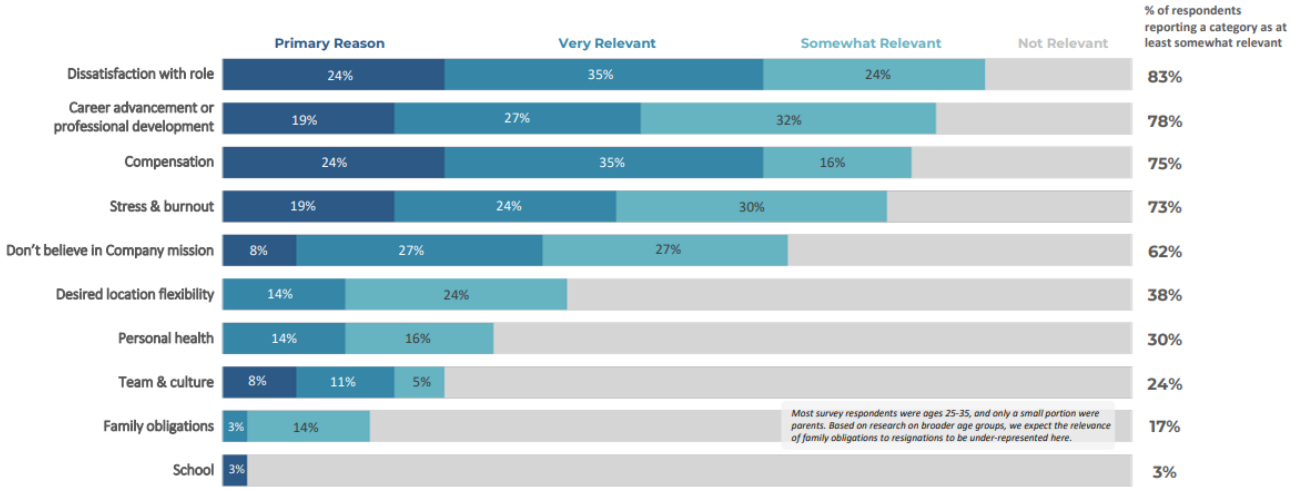

Рисунок 1.3 – Причини звільнення працівників технічних компаній*

* за даними ICONIC: Tech Employee Survey

97% працівників технологічних компаній називають кар'єрне зростання та професійний розвиток принаймні певною мірою пов'язаними з їхнім

рішенням залишитися у своїй поточній компанії, причому 53% називають це основною причиною – більше, ніж будь-яка інша названа категорія (рисунок 1.4).

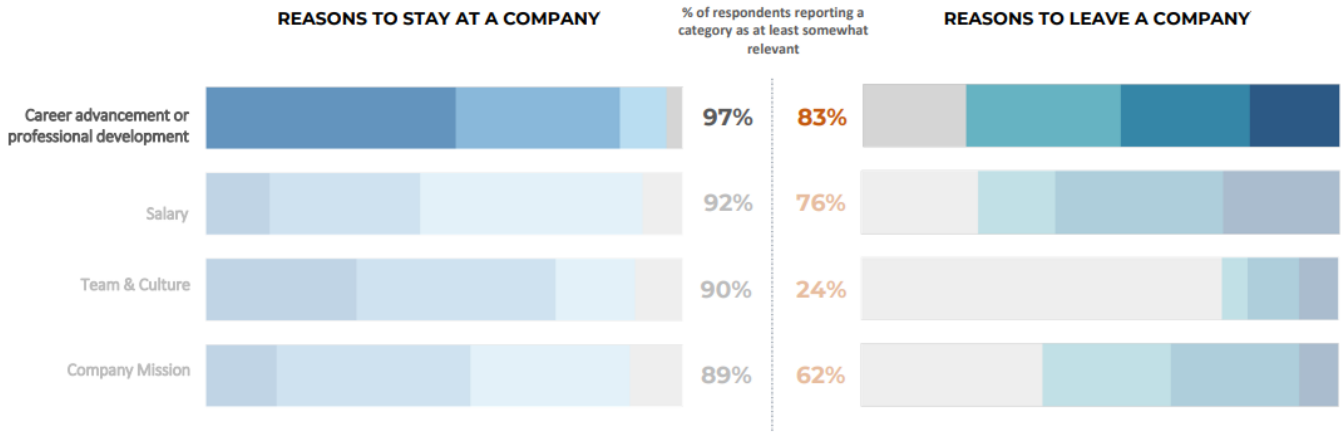

Рисунок 1.4 – Ключові причини для того, щоб залишитись або звільнитись з компанії*

* за даними ICONIC: Tech Employee Survey

Вплив звільнення працівників на результати діяльності організації

Дослідження, проведене Forbes, показало, що середній рівень відтоку кадрів в IT/технологічних компаніях становить 23% на рік, що значно вище порівняно з іншими галузями. Несприятливі показники відтоку кадрів в IT-індустрії мають великий вплив, впливаючи на різні аспекти, які можуть бути не очевидними одразу.

1. Ефективність команди: коли відбувається непередбачуване звільнення співробітників, продуктивність зазнає удару, особливо коли інші не можуть зайняти їхню посаду. Наприклад, втрата фронтенд-розробника може означати затримку у повноцінних проектах, що впливає на терміни та, своєю чергою, обтяжує продуктивність через обмежене резервне копіювання.

2. Сплеск вигорання: технічні команди, які стикаються з відтоком кадрів, часто стикаються з вигоранням, оскільки члени, що залишилися, беруть на себе додаткову роботу. Опитування показує, що 41,1% пов'язують вигорання з

нестачею персоналу. Вигорання є поширеним та дорогим явищем, яке щорічно коштує 500 мільярдів доларів США та 550 мільйонів робочих днів.

3. Втрата технологічних знань: коли член технічної команди звільняється, він забирає з собою свої унікальні знання. Цей незафіксований досвід, який іноді називають «племінним знанням», є значною втратою. Майже 42% спеціалізованої робочої експертизи належить IT-фахівцю, який звільняється. Більше того, 60% мають труднощі з доступом до критично важливої інформації від колег, що створює прогалину, для подолання якої потрібні час і зусилля.

1.3 Залученість та утримання співробітників

Щоб протидіяти негативним наслідкам відтоку працівників, технологічні компанії можуть застосувати кілька стратегій для вирішення проблем, що виникають через відтік працівників.

1. Конкурентна компенсація та пільги: пропонування конкурентних пакетів компенсації, включаючи заробітну плату, бонуси та пільги, може допомогти залучити та утримати найкращих спеціалістів в IT-індустрії.

2. Забезпечте гнучкість: дослідження, проведене Yoh, показало, що 44% жінок і 39% чоловіків розглянуть можливість прийняття нової роботи, яка пропонує більш гнучке робоче середовище. Коли організація креативно підійде до того, як команди можуть виконувати свою роботу, роботодавці зможуть врахувати та утримати більше людей.

3. Опитування співробітників: проведення регулярних опитувань для оцінки задоволеності співробітників та визначення напрямків для покращення може допомогти вирішити основні проблеми, що сприяють відтоку кадрів.

4. Вихідні співбесіди: проведення вихідних співбесід зі співробітниками, які звільняються, може дати цінну інформацію про причини звільнення та допомогти визначити напрямки для покращення.

IT-індустрія може використовувати різні технологічні інструменти для підвищення залученості та утримання співробітників . Наприклад, платформи для співпраці, такі як Slack та Microsoft Teams, сприяють безперебійній комунікації та сприяють відчуттю приналежності між віддаленими командами. Крім того, програмне забезпечення для управління ефективністю, таке як BambooHR та 15Five, забезпечує зворотний зв'язок та визнання в режимі реального часу, сприяючи задоволеності та утриманню співробітників.

1.4 IT-ринок праці в Україні

Вперше за останні три роки кількість вакансій на DOU зросла – одразу на 31%. Найбільше додалося технічних позицій, хоча нетехнічні спеціальності показали ще швидше зростання. Продуктові компанії демонструють кращі результати, ніж сервісні, хоч і там не обійшлося без скорочень.

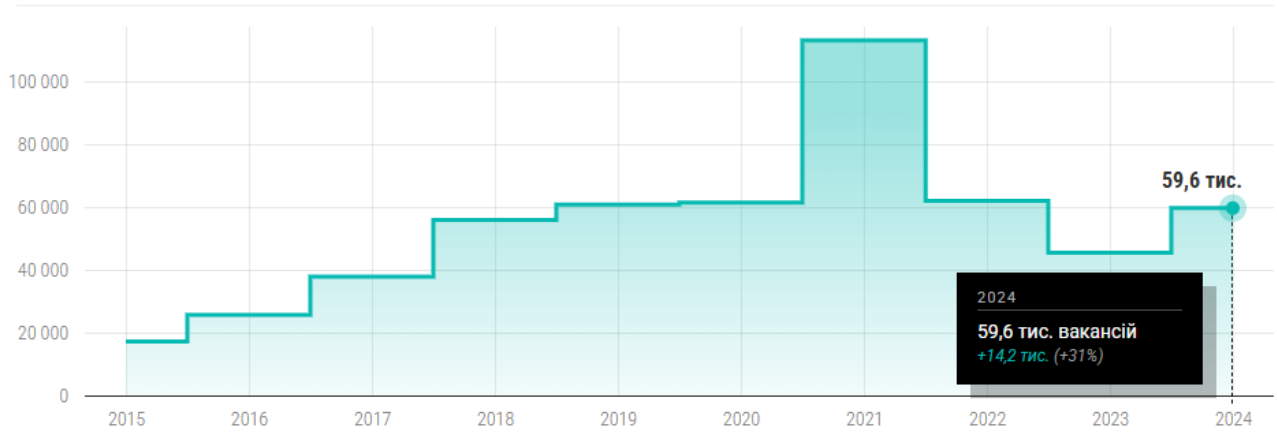

Рисунок 1.5 – Кількість вакансій на jobs.dou.ua за роками

Нетехнічні напрямки показали зростання на 40%, тоді як технічні — на 29%. Проте в абсолютних цифрах технічних вакансій додалося на 9 тисяч, а нетехнічних – на 5,4 тисячі (рисунок 1.6).

За рік змінився склад трійки найпопулярніших категорій. Перше місце, як і раніше, утримує маркетинг – 6,6 тисячі вакансій. Категорія Analyst зросла на 60% і посунула з другої позиції Sales.

Усього зросли 41 категорія вакансій, і лише 12 – зменшилися. Найбільше у числовому вимірі зросли: Marketing (+1,9 тис.), Analyst (+1,4 тис.) і QA (+1,2 тис.).

Серед лідерів за відсотковим приростом – нові категорії, які активно розвивалися у 2024 році: Hardware, AI/ML, Data Engineer. Слідом за ними – Legal (+89%) та інші нетехнічні спеціальності.

У технічній сфері найбільше зросли вакансії для Node.js (+48%), Blockchain (+45%) та Embedded (+36%).

Найсильніше падіння показала категорія C++: мінус 217 вакансій або 31%. На 33% скоротилася кількість вакансій з Unreal Engine, а також помітно зменшилися SAP, Unity та Scala – більш ніж на 20%.

Незначне зниження – на 3% – зафіксовано в популярних категоріях DevOps та Java.

Посада/Технологія	2023	2024	Зміна	Зміна, %
Marketing	4 694	6 572	+1878	+40
Analyst	2 400	3 847	+1447	+60
Sales	3 177	3 546	+369	+12
QA	2 317	3 540	+1223	+53
Front End	2 218	2 838	+620	+28
Design	1 907	2 693	+786	+41
Project Manager	1 891	2 499	+608	+32
HR	1 385	2 331	+946	+68
Support	1 395	1 990	+595	+43
PHP	1 762	1 984	+222	+13
Node.js	1 300	1 927	+627	+48
DevOps	1 895	1 835	-60	-3
Python	1 417	1 717	+300	+21
Product Manager	1 268	1 679	+411	+32
.NET	1 320	1 603	+283	+21
Finance	899	1 578	+679	+76
Java	1 486	1 438	-48	-3

Рисунок 1.6 – Кількість вакансій на DOU за категоріями

Серед Топ-50 компаній (на рисунку 1.7 Топ-10), які зростали в першому півріччі, переважають продуктові й гібридні, тоді як сервісних – лише три. Із 10 найбільших IT-компаній України лише дві показали зростання – обидві продуктові: Genesis та Ajax Systems.

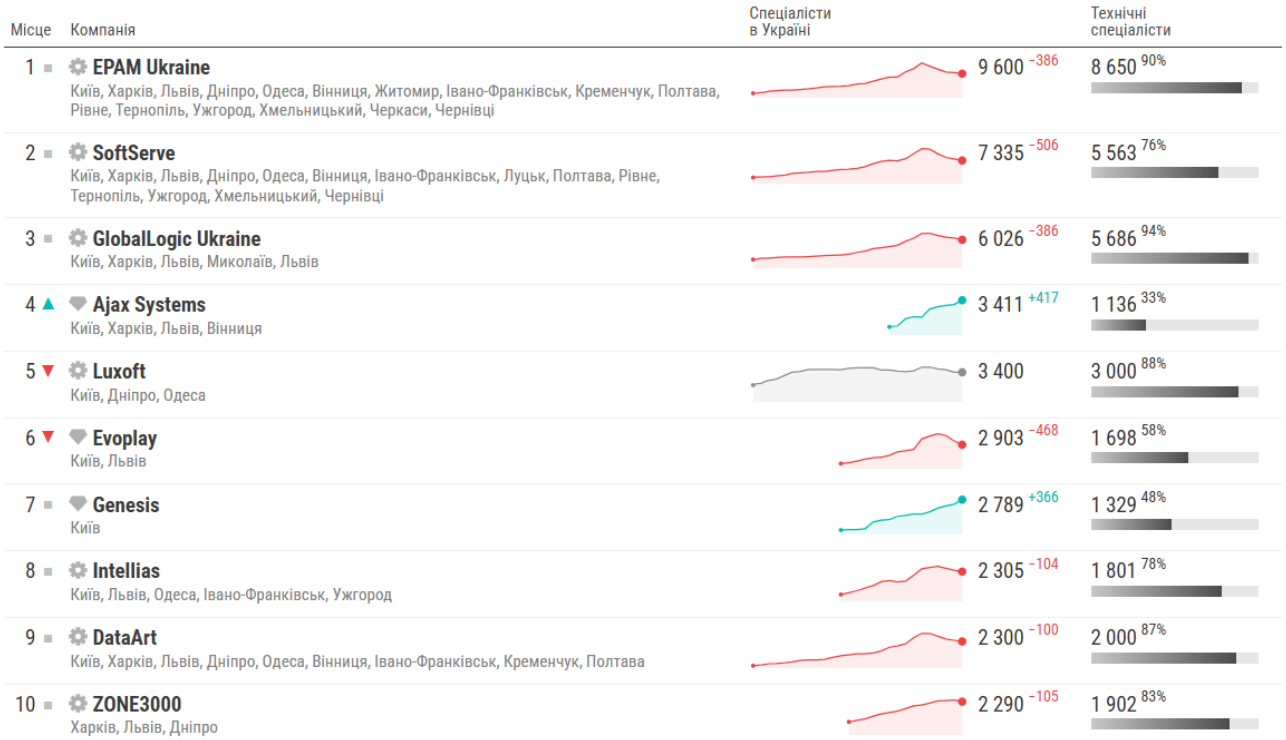

Рисунок 1.7 – Дані по спеціалістам в IT-компаніях України

1.5 HR-аналітика та методи штучного інтелекту

Штучний інтелект у сфері управління персоналом зараз є нормою. Фактично, 60% керівників вищого рівня та осіб, які приймають рішення у сфері управління персоналом, вважають, що їхні відділи управління персоналом інтегрують більше штучного інтелекту та автоматизації у свої функції та процеси протягом п'яти років, оскільки очікується, що це покращить практики управління персоналом та сприятиме конкурентоспроможності та досягненням організацій.

Штучний інтелект змінює аналітику HR , покращуючи збір, обробку та прийняття рішень даних — це оптимізує процеси управління персоналом, допомагаючи HR-командам працювати швидше та розумніше.

Компанії, що використовують штучний інтелект у сфері управління персоналом, можуть аналізувати показники плинності кадрів та дані про плани утримання персоналу для глибшого розуміння причин звільнення співробітників. Моделі штучного інтелекту також можуть аналізувати тенденції робочої сили, щоб прогнозувати області потенційного звільнення працівників. HR-команди можуть використовувати цю інформацію для покращення стратегій найму, адаптації та утримання персоналу з метою залучення та утримання найкращих талантів.

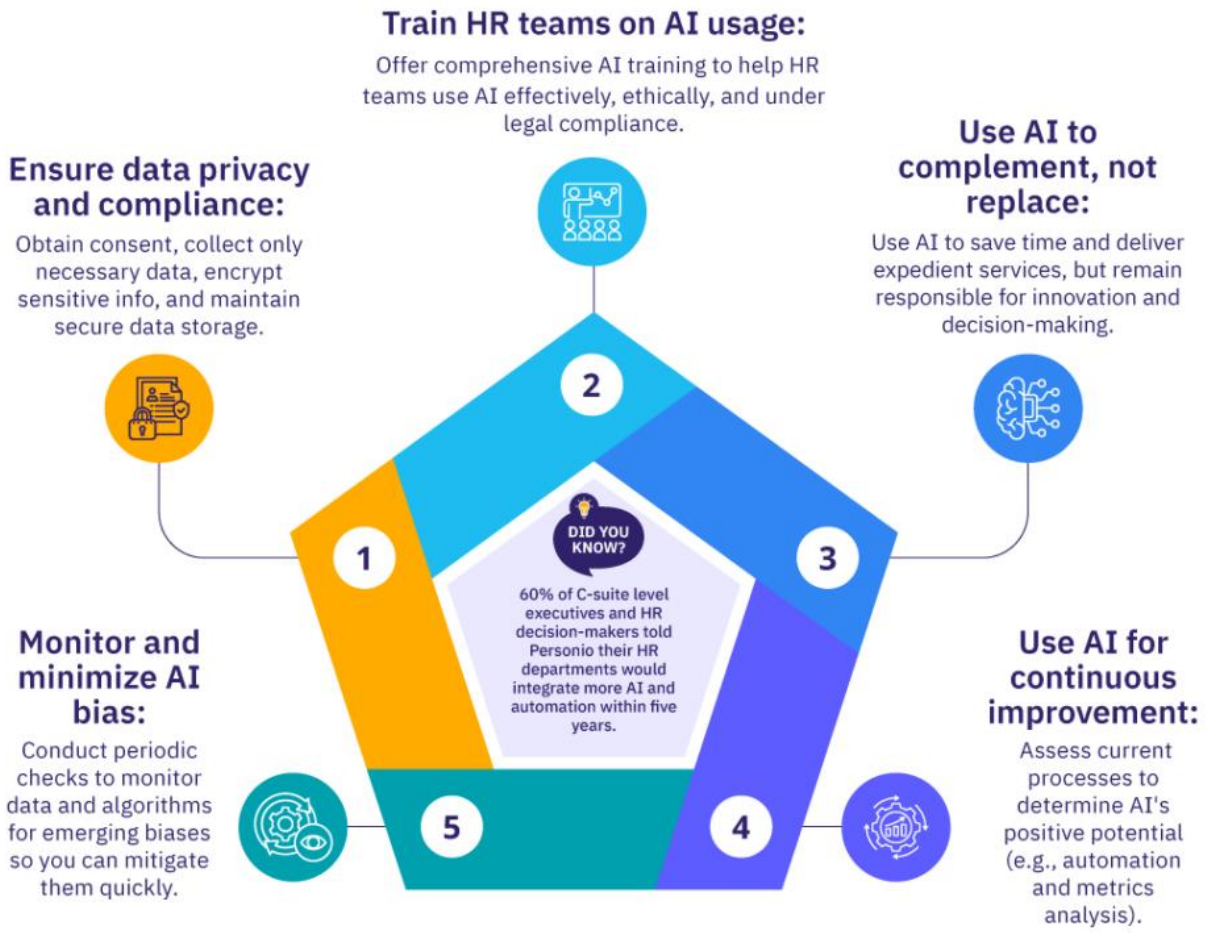

Рисунок 1.8 – Кроки для використання штучного інтелекту в HR-аналітиці

У HR-аналітиці виділяють 4 види аналітики і починаючи з передиктивної аналітики засобом досягнення мети стає ML. У таблиці 1.1 наведено цілі кожного виду аналітики, а також сфери їх застосування та методи аналізу. Ми плануємо займатись предиктивною аналітикою оскільки вона допомагає у завданні передбачення плинності кадрів в компанії.

Таблиця 1.1 – Види HR-аналітики

Назва	Ціль	Сфери застосування	Методи аналізу
Дескриптивна аналітика	Виявлення та моніторинг проблем та їх діагностика на основі даних. Відповідає н запитання «Що сталося?»	Структуровані дані та звітність, наприклад: - структура персоналу; - норми праці та нормативи чисельності; - огляди заробітних плат; - метрики ефективності процесів; - бенчмарки, Dashboard	Описова статистика
Прогнозна аналітика	Прогнозування на основі підтверджених статистичних гіпотез. Відповідь на питання «Чому і Як»	– методи прогнозування чисельності; - планування завантаження; - формування профілю успішного співробітника; – план заходів щодо підвищення залучення співробітників; - визначення норм проходження тестів	– кореляційно-регресійний аналіз; – кластерний аналіз
Предиктивна аналітика	Прогнозування на основі виявлення неочевидних залежностей та мультиваріантних сценаріїв. Відповідає на запитання «Що буде в майбутньому?»	– прогнозування звільнення працівника; - валідизація моделі компетенції; – прогнозування успішності співробітника у конкретному середовищі	Машинне навчання: - дерева рішень, - нейронні мережі
Прескриптивна аналітика	Пропозиція рішень у динамічному середовищу. Відповідь на запитання «Як змінити /поліпшити те, що буде?»	Формування рішень для зниження ймовірності звільнення конкретного співробітника	Машинне навчання: - дерева рішень, - нейронні мережі

Висновки за розділом

Проведено огляд причин плинності, її впливу на ефективність компанії, а також специфіки IT-ринку праці в Україні. Розглянуто роль HR-аналітики та потенціал використання інтелектуальних систем. Показано, що зниження плинності є критично важливим для стабільності бізнесу. Таким чином, проблема плинності кадрів в IT-сфері є актуальною та потребує науково обґрунтованих і технологічних рішень для її попередження.

РОЗДІЛ 2 МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ ПЕРЕДБАЧЕННЯ ПЛИННОСТІ КАДРІВ

2.1 Огляд методів машинного навчання

1. Логістична регресія (рисунок 2.1): загальновизнано, що логістична регресія є однією з найкорисніших доступних статистичних моделей. Крім того, це добре відомий метод аналізу даних, який використовують для дослідження як пропорційних, так і бінарних типів наборів даних. Однією з особливостей, яка відрізняє логістичну регресію від інших типів регресії, є те, що її можна використовувати для аналізу даних з більш ніж одного класу. Це один з найчастіше використовуваних алгоритмів класифікації.

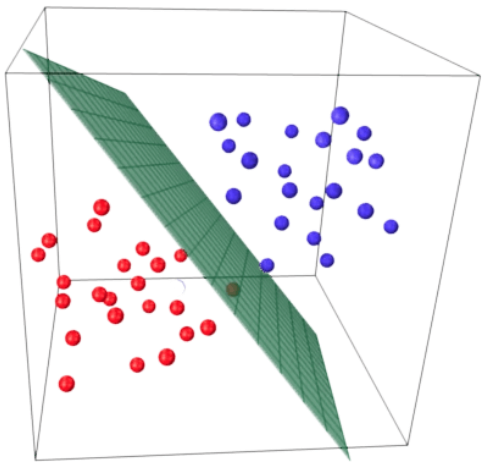

Рисунок 2.1 – Логістична регресія

2. Дерева рішень – це алгоритм ML, який використовує структуру дерева для ухвалення рішень на основі вхідних даних. Корінь, гілки та листя складають компоненти дерева рішень. Загальною практикою є позначення кореневого вузла батьківським вузлом. Вузли використовуються для представлення кожної характеристики, тоді як гілки використовуються для позначення зв'язків, що встановлюються між вузлами. Правила або вибір містяться в цих гілках, як показано на рисунку 2.2.

Корінь, гілки та листя складають компоненти дерева рішень.

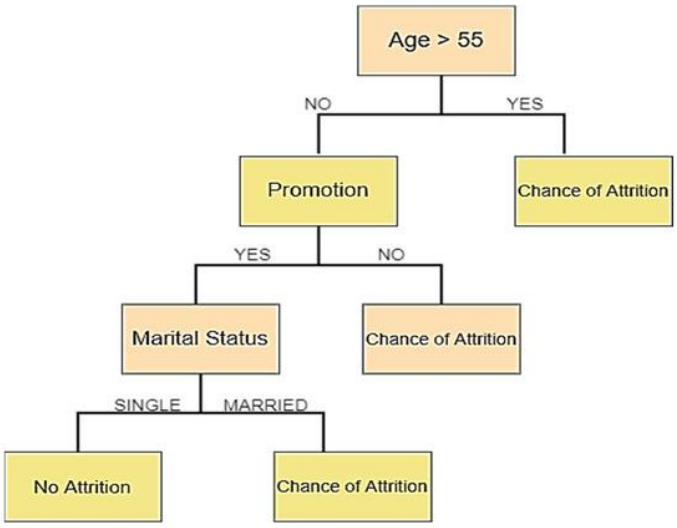

Рисунок 2.2 – Приклад дерева рішень

Очікується, що листок слугуватиме результатом. ID3 та CART є прикладами деяких алгоритмів дерева рішень, які використовуються найчастіше. Цей алгоритм застосовується до задач, пов'язаних з класифікацією та регресією, і може працювати як з безперервною, так і з категоріальною інформацією.

3. K–найближчі сусіди (KNN) – це алгоритм машинного навчання з учителем, який можна застосовувати до задач, пов'язаних як з класифікацією, так і з регресією. Алгоритм KNN робить прогнози щодо результату, використовуючи знання про вхідні дані. Потім вхідні дані поділяються на відповідні категорії. Алгоритм прагне шукати позицію, яка забезпечить найкращі результати для нової точки даних. Після аналізу точок даних, що були надані як вхідні дані, приймається рішення щодо розташування нової точки (рисунок 2.3). Алгоритм, який використовувався в цьому дослідженні, виглядає наступним чином:

Крок 1: вибір числа K, тобто околиці.

Крок 2: обчислення відстані (зазвичай евклідова).

Крок 3: використовуючи положення точок даних, знайти приклади, які найближче розташовані до заданого положення.

Крок 4: підрахувати кількість балів, набраних у кожній категорії.

Крок 5: віднесення щойно отриманих точок до категорії, в якій кількість навколишніх точок більша.

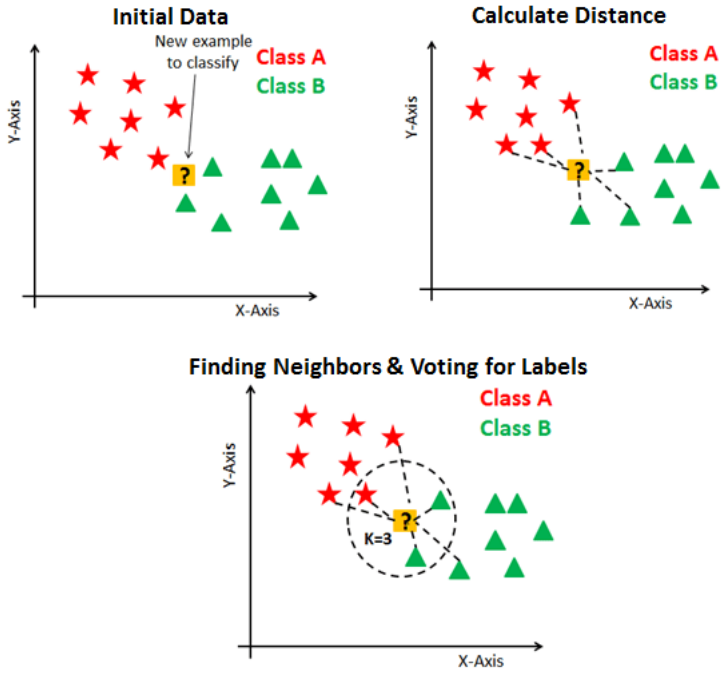

Рисунок 2.3 – K-найближчі сусіди

4. Машини опорних векторів (SVM) – ще один популярний різновид моделей машинного навчання з учителем. Його основне застосування полягає у вирішенні проблем класифікації, хоча його також можна використовувати для вирішення регресійних задач (рисунок 2.4).

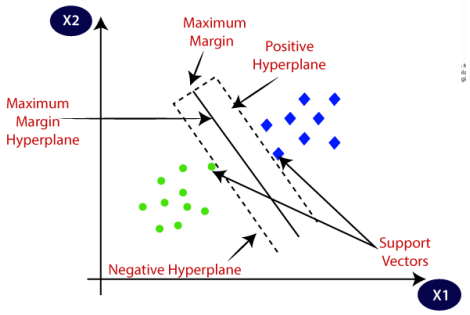

Рисунок 2.4 – Машини опорних векторів

Фундаментальна концепція цього методу полягає в проведенні лінії або встановленні межі, яка окреслює n окремих груп або класифікацій у загальному просторі. Коли до цього простору додається нова точка даних, вона може швидко знайти своє правильне місце в сформованих категоріях. Гіперплощина – це інша назва лінії, яка проходить через середину цих класів. Кажуть, що задача класифікації має лінійний алгоритм, коли її можна вирішити, провівши одну пряму лінію. Її називають нелінійною машиною опорних векторів (SVM), коли прямої лінії недостатньо, і в цьому випадку замість неї отримується крива лінія.

5. Випадковий ліс – це метод машинного навчання, який використовується для вирішення задач, пов'язаних з регресією та класифікацією. Він походить від ідеї ансамбльового навчання (рисунок 2.5). Набір даних розділяється на численні підмножини, щоб цей метод міг враховувати різні дерева рішень. Завдяки цій процедурі кілька результатів об'єднуються в один результат, який потім обчислюється як середнє значення всіх результатів. Точність алгоритму підвищується пропорційно до кількості дерев та піднаборів даних, що враховуються. Завдяки такій поведінці він може керувати надзвичайно великою кількістю наборів даних.

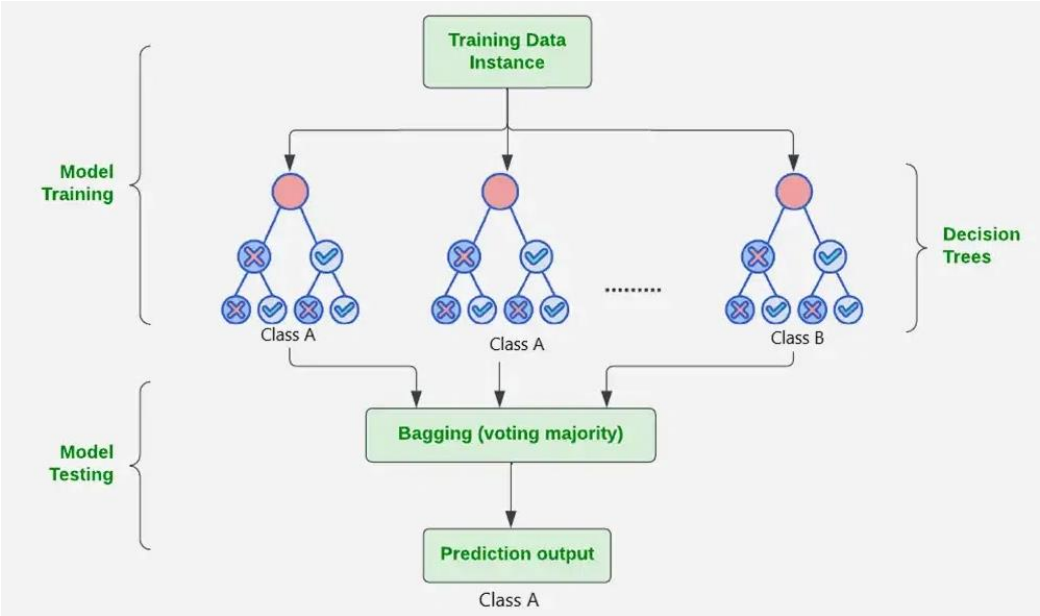

Рисунок 2.5 – Random Forest

6. Наївний Байєс – це добре відомий підхід до машинного навчання з учителем, який був розроблений шляхом застосування теореми Баєса. Через його ймовірнісний характер, робота цього алгоритму базується на ймовірності того, що елемент буде знайдено. Хоча він здебільшого використовується для задач, що потребують категоризації тексту, він достатньо адаптований для використання в широкому спектрі задач класифікації.

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$

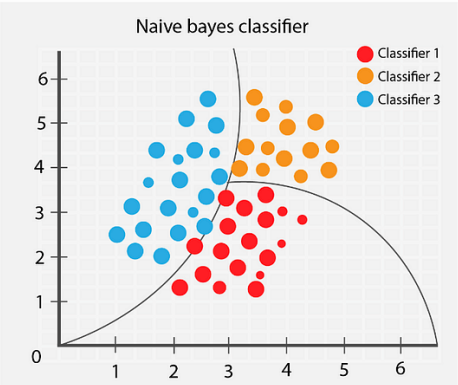

Рисунок 2.6 – Наївний Байєс

2.2 Загальна схема рішення

Загальна схема рішення представлена на рисунку 2.7.

1. Збір даних: основою цього дослідження є комплексний датасет з різноманітних наборів даних для охоплення безлічі факторів, що впливають на плинність кадрів. Первинні джерела включають історичні записи про працівників, що охоплюють такі деталі, як стаж роботи, посада, підвищення по службі та попередні випадки плинності кадрів. Метрики ефективності, що охоплюють індивідуальні та командні досягнення, а також механізми зворотного зв'язку, сприяють тонкому розумінню внеску та задоволеності працівників. Крім того, дані, пов'язані з балансом між роботою та особистим життям, можливостями кар'єрного розвитку та міжособистісними стосунками, включені для забезпечення цілісного уявлення про робоче середовище.

Опитування та механізми зворотного зв'язку використовуються для збору суб'єктивних висновків, додаючи якісні виміри до кількісних даних.

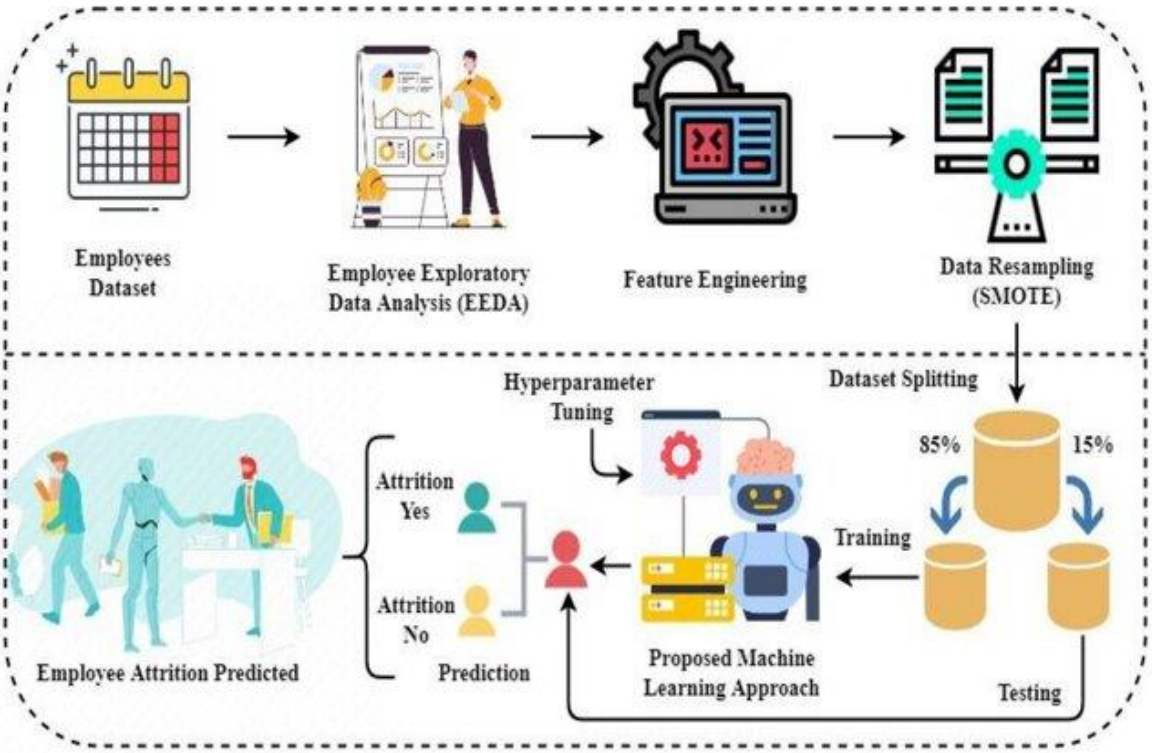

Рисунок 2.7 – Використання машинного навчання для прогнозного аналізу стратегій плинності та утримання кадрів

2. Попередня обробка даних: Необроблені дані проходять ретельну фазу попередньої обробки, щоб забезпечити точність та надійність у подальших аналізах. Це включає обробку відсутніх даних, викидів та аномалій для підтримки цілісності даних. Категоричні змінні кодуються, а числові змінні стандартизуються для забезпечення узгодженості в усьому наборі даних. Методи нормалізації даних застосовуються для приведення різнорідних шкал до стандартизованого діапазону. Крім того, набір даних розділений на навчальні та тестові набори для полегшення оцінки моделей машинного навчання. Ця фаза попередньої обробки є вирішальною для підвищення якості вхідних даних, дозволяючи алгоритмам машинного навчання ефективно розрізняти закономірності та кореляції.

3. Алгоритми машинного навчання. Для прогнозування плинності кадрів розглядається кілька алгоритмів машинного навчання, кожен з яких привносить свої сильні сторони в аналіз. Ми вже згадували їх. Логістична регресія забезпечує просту модель для бінарної класифікації, тоді як дерева рішень пропонують інтерпретованість та розуміння важливості ознак. Ансамблеві методи, такі як випадкові ліси та градієнтне підвищення, поєднують кілька моделей для підвищення точності прогнозування. Методи опорних векторів досліджуються на предмет їхньої ефективності в обробці нелінійних зв'язків у даних. Вибір алгоритмів базується на їх придатності для набору даних та інтерпретованості, необхідній для того, щоб зробити практичні висновки. Для оцінки ефективності та узагальнюваності моделі використовуються методи перехресної перевірки. Такий диверсифікований підхід забезпечує ретельне дослідження методів машинного навчання, що дозволяє визначити найефективніші моделі для прогнозування плинності кадрів та розробки стратегій утримання персоналу.

Висновки за розділом

Описано загальну схему побудови системи прогнозування та проаналізовано основні алгоритми машинного навчання, які можуть бути застосовані для цієї задачі. Визначено найбільш релевантні підходи, які поєднують точність та інтерпретованість. Таким чином, методи машинного навчання, зокрема ансамблеві моделі та нейронні мережі, мають високий потенціал для вирішення задачі передбачення плинності персоналу.

РОЗДІЛ 3
РОЗРОБКА ТА ДОСЛІДЖЕННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ПЕРЕДБАЧЕННЯ ПЛИННОСТІ КАДРІВ

3.1 Використовувані програмні засоби розробки та опис датасету

Спершу потрібно обрати відповідні інструменти та дані, адже саме від їх правильного підбору залежить успішність результатів дослідження. Наразі в більшості робіт, присвячених розробці моделей машинного навчання, активно застосовуються такі інструменти:

– Python – мова програмування для написання програмного коду;

– Numpy – бібліотека, яка забезпечує можливість роботи з багатовимірними масивами та матрицями;

– Pandas – бібліотека, призначена для обробки та аналізу даних;

– Sklearn (Scikit-learn) – бібліотека ML, що надає алгоритми та інструменти для аналізу даних та побудови моделей ML;

– Matplotlib, Seaborn – бібліотеки візуалізації даних, що дозволяють створювати різноманітні графіки та діаграми;

– Google Colab – це веб-сервіс від Google, призначений для написання та виконання коду мовою Python.

Python був обраний через те, що для нього значна кількість необхідних бібліотек, наприклад, sklearn, яка надає необхідні алгоритми для роботи з ML.

Вхідні дані взято на платформі Kaggle. Вхідні дані являють собою cvs-файл, що містить дані по працівникам різних відділів. У таблиці 3.1 дається опис кожній ознаки.

Можна помітити, що останні дві колонки є категоріальними. Це означає, що дані у цих стовпцях слід замінити на числові чи провести їм бінаризацію.

Наступним кроком дізнаємося загальну кількість записів, кількість ненульових значень, типи даних кожного стовпця. Отримані дані проілюструємо на рисунку 3.2.

Таблиця 3.1 – Опис ознак датасету

Ознака	Опис
satisfaction level	Рівень задоволеності працівника роботою у компанії. Значення знаходиться в діапазоні від 0 до 1, де 0 – повне незадоволення, а 1 – повне задоволення
last evaluation	Оцінка роботи співробітника на останній здійсненій оцінці. Значення в діапазоні від 0 до 1
number project	Кількість проектів, над якими працював співробітник
average montly hours	Середня кількість годин, яку працівник витрачає на роботу на місяць
time spend company	Загальний час, який співробітник вже пропрацював у компанії у роках
work accident	Чи був у працівника нещасний випадок на роботі Значення 1 вказує на наявність, а 0 – її відсутність
left	Чи звільнився співробітник. Значення 1 вказує на звільнення, а 0 - на тих, що залишаються в компанії
promotion last 5 years	Чи був поступ у співробітника за останні 5 років. Значення 1 вказує на наявність просування, а 0 – на його відсутність
department	Відділ, у якому працює співробітник
salary	Рівень зарплати працівника. Можливі значення: "low" (низький), "medium" (середній), "high" (високий)

Імпортуємо вхідні дані та для отримання уявлення про використовувані дані. На рисунку 3.1 показані фрагмент із файлу з даними про співробітників.

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Department	salary
0	0.38	0.53	2	157	3	0	1	0	sales	low
1	0.80	0.86	5	262	6	0	1	0	sales	medium
2	0.11	0.88	7	272	4	0	1	0	sales	medium
3	0.72	0.87	5	223	5	0	1	0	sales	low
4	0.37	0.52	2	159	3	0	1	0	sales	low
5	0.41	0.50	2	153	3	0	1	0	sales	low
6	0.10	0.77	6	247	4	0	1	0	sales	low
7	0.92	0.85	5	259	5	0	1	0	sales	low
8	0.89	1.00	5	224	5	0	1	0	sales	low
9	0.42	0.53	2	142	3	0	1	0	sales	low

Рисунок 3.1 – Дані про співробітників

```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   satisfaction_level    14999 non-null  float64
 1   last_evaluation       14999 non-null  float64
 2   number_project        14999 non-null  int64
 3   average_montly_hours  14999 non-null  int64
 4   time_spend_company    14999 non-null  int64
 5   Work_accident         14999 non-null  int64
 6   left                  14999 non-null  int64
 7   promotion_last_5years 14999 non-null  int64
 8   Department            14999 non-null  object
 9   salary                14999 non-null  object
```

Рисунок 3.2 – Загальна інформація про набір даних

Отримана інформація про датасет говорить про відсутність нульових значень та наявність 14999 прикладів.

Для отримання більш повної картини набору даних побудуємо діаграми boxplot (рисунок 3.3).

Box Plot – це графічний метод візуалізації розподілу даних для отримання розуміння та прийняття обґрунтованих рішень. Box Plot відображає групу числових даних через їхні квартилі.

Box Plot дає зведення з п’яти чисел набору даних, якими є:

– мінімум – це мінімальне значення в наборі даних без урахування викидів;

– перший квартиль (Q1) – 25% даних знаходяться нижче першого (нижнього) квартиля;

– медіана (Q2) – це середня точка набору даних. Половина значень лежить під ним, а половина – вище;

– третій квартиль (Q3) – 75% даних знаходяться нижче третього (верхнього) квартиля;

максимум – це максимальне значення в наборі даних без урахування викидів.

Виходячи з цих діаграм, можна зробити висновок, що дані не містять аномалій, оскільки відсутні викиди даних.

Визначимо перелік відділів, які у стовпці «department». У нас виявлено 10 унікальних відділів (рисунок 3.4). Надалі для цієї ознаки буде проведено кодування.

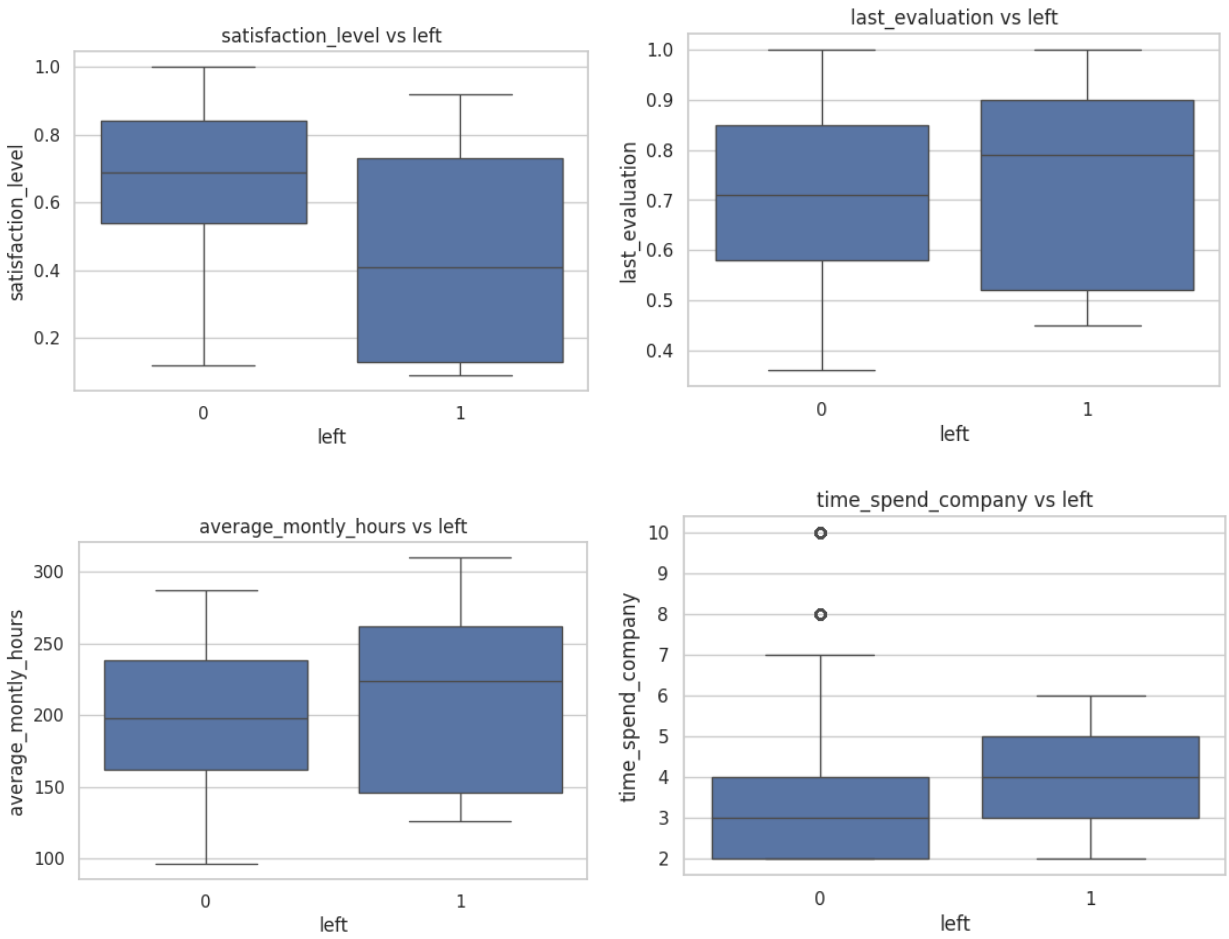

Рисунок 3.3 – Діаграми boxplot

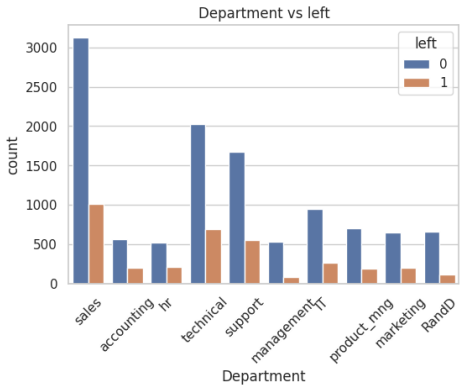

Рисунок 3.4 – Кількість співробітників, що залишаються та звільнились по відділах

Можна також помітити, що значення ознак для кожного відділу приблизно однакові, і вони на хорошому рівні, що свідчить про успішне управління. Це показує, що управління відділом у контексті утримання співробітників має важливе значення. За представленою діаграмою можна проаналізувати компанію. Зауважимо, що компанія успішно розвиває свою діяльність у сфері продажу завдяки активній участі великої кількості співробітників у клієнтській підтримці. Зважаючи на все, співробітники у цій сфері відповідають на запитання клієнтів, у той час як технічні фахівці фокусуються на підтримці програмного забезпечення та його розробці.

Продовжуючи аналіз даних, дізнаємося скільки співробітників звільнилося і залишилося у компанії. Покажемо це на рисунку 3.5.

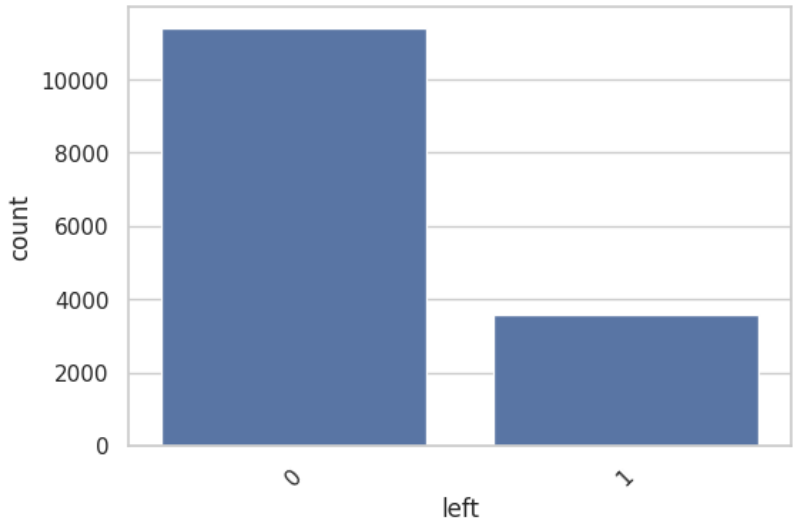

Рисунок 3.5 – Кількість співробітників, що звільнилися і залишилися

З 14 999 співробітників 3571 звільнилися з компанії, а 11 428 залишилося.

Помітна різниця для двох груп спостерігається в ознаці «рівень задоволеності» (рисунок 3.6).

Цю ознаку складно виразити у числовій формі, тому, ймовірно, вона була сформована на основі інших, невідомих нам тестів чи оцінок, які використовувалися для кількісного вимірювання продуктивності працівників.

Також помітно, що ті співробітники, які отримували підвищення, значно рідше звільнялися з роботи — їх залишалося майже в п’ять разів більше, ніж тих, хто не отримував просування.

Це підкреслює важливість можливостей кар’єрного розвитку для збереження кадрів.

Ще один цікавий факт: працівники, які пережили нещасний випадок на роботі, майже в чотири рази частіше залишалися в компанії, ніж ті, хто не мав таких інцидентів.

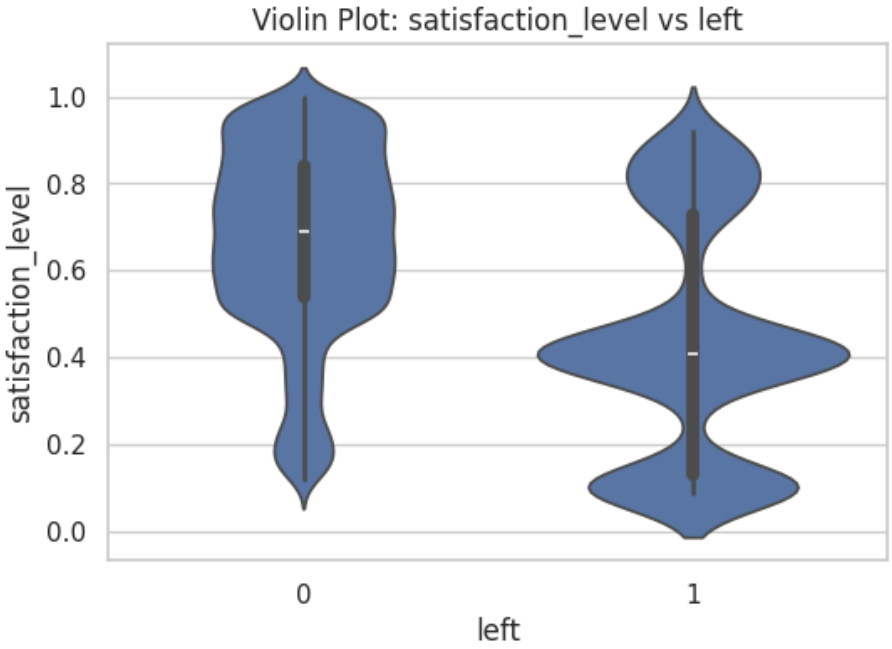

Рисунок 3.6 – Кількість співробітників, що звільнилися і залишилися в залежності від рівня задоволеності

Після аналізу однієї з категоріальних ознак satisfaction_level переходимо до наступній – salary, яка класифікує працівників на три рівні зарплати: high, middle, low (рисунок 3.7).

Зауважимо кореляцію, співробітники, які мають низьку заробітну плату та відсутнє підвищення, частіше залишають компанію, про що свідчить діаграма, зображена на рисунку 3.8.

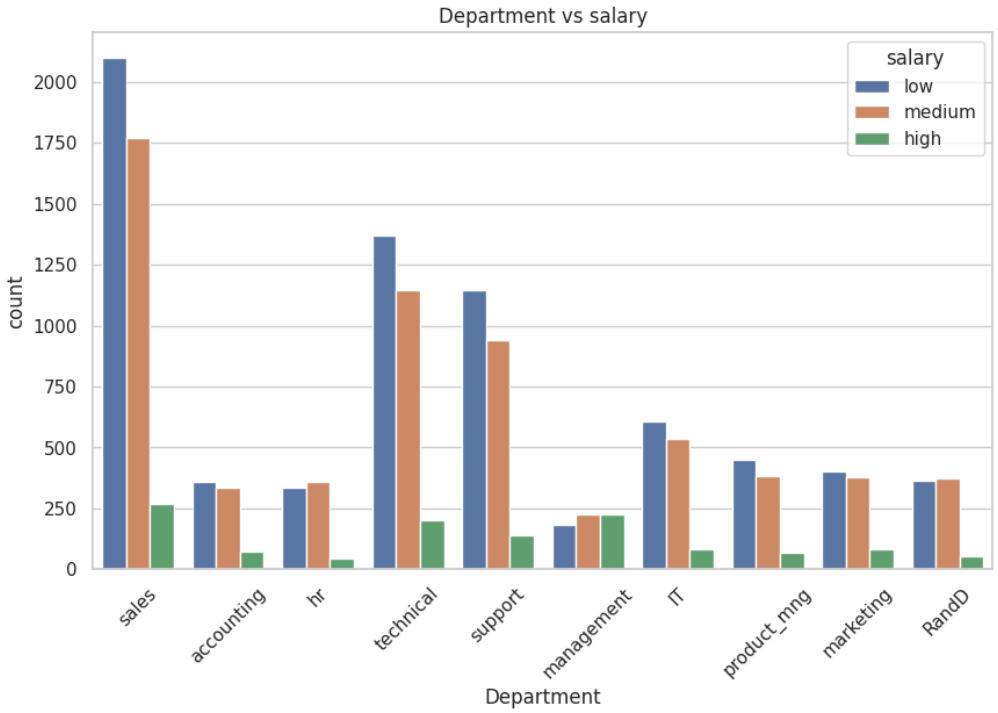

Рисунок 3.7 – Кількість співробітників по відділах відносно рівня зарплати

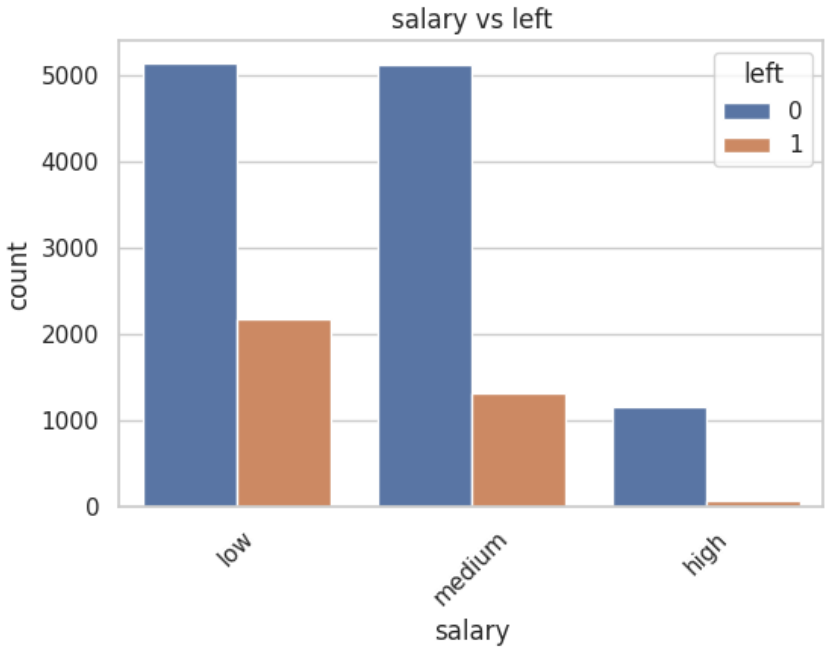

Рисунок 3.8 – Рівень зарплати порівняно з плинністю кадрів

Обчислимо відносні пропорції, розділивши кожне значення у рядку таблиці на суму значень у цьому рядку. Це перетворення дозволило отримати частку співробітників, що звільнилися і залишилися, для кожного рівня

зарплати. Відобразимо розподіл працівників залежно від рівня зарплати на рисунку 3.9.

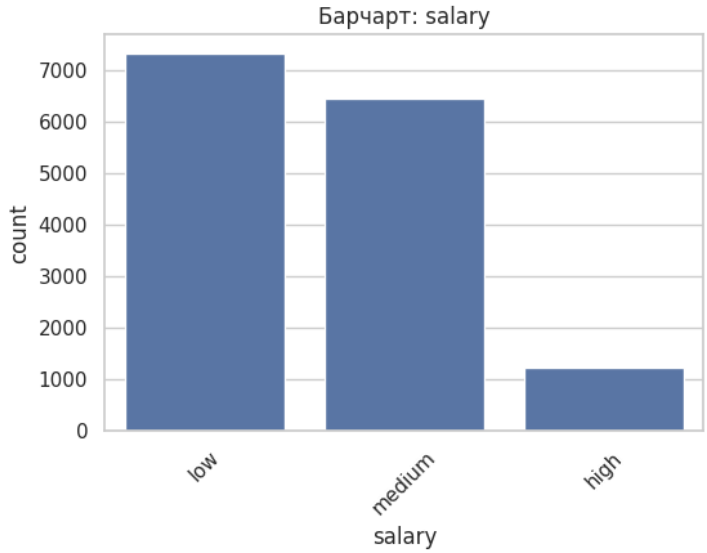

Рисунок 3.9 – Розподіл працівників залежно від рівня зарплати

У компанії переважає низький та середній рівень заробітної плати, у той час як високий рівень зарплати надано обмеженій кількості працівників.

Якщо розрахувати відсоток плинності кадрів за весь час, то вийде, що вона дорівнює 23,8%. Це вдалося вирахувати завдяки формулі

$$Кп=Сзв\cdot 100\% / Сспр, \quad (3.1)$$

де Кп – коефіцієнт плинності; Сзв – число вибулих з причин плинності; Сспр – середня облікова кількість працівників протягом того ж періоду.

Плинність рівна 23,8% і це є великий показник, оскільки нормальним рівнем плинності кадрів вважається 3-5%.

Розуміючи, що цей показник може змінюватись у залежність від сфери діяльності підприємства, у разі відзначається істотне перевищення середніх значень.

Побудуємо серію гістограм для кожного числового стовпця з візуалізації розподілу даних.

На осі X відображатимуться значення ознак, а на осі Y – кількість співробітників. Гістограми показані на рисунку 3.10.

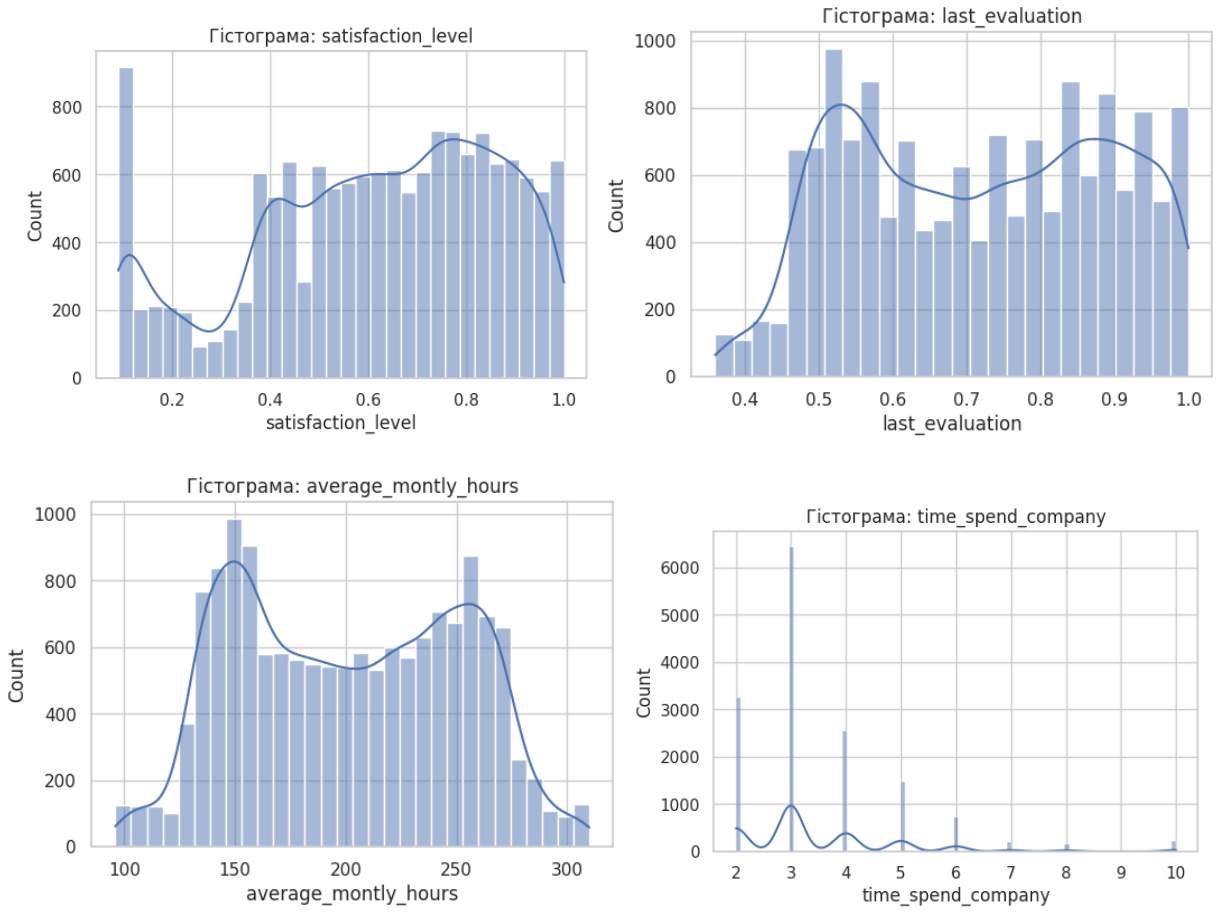

Рисунок 3.10 – Розподіл даних для кожної числової ознаки

На графіку satisfaction_level по підрозділах (рисунок 3.11) можна побачити, як рівень задоволеності працівників розподіляється залежно від відділу. Ось ключові спостереження:

1. Sales, Technical, Support: у цих великих підрозділах розподіл задоволеності доволі рівномірний, але видно деякий спад на низькому рівні (< 0.3) і пік на високому (> 0.7). Це свідчить, що незадоволені працівники трапляються рідше, однак вони все ж є.

2. HR, Accounting, Product Management, Management: розподіли менш виражені, але в них є суттєвий пік у середньому та високому діапазоні

задоволення. Це може свідчити про стабільніше середовище в цих департаментах або про меншу кількість працівників.

3. IT, Marketing, RandD: IT має пік у районі 0.7–0.8, тобто загалом високий рівень задоволення. В RandD та Marketing розподіл ближчий до нормального – більшість працівників мають середній рівень задоволеності.

Загалом видно, що найбільші відділи мають широкий розкид рівня задоволення, тоді як менші підрозділи демонструють концентрованіші значення, часто ближчі до високих рівнів.

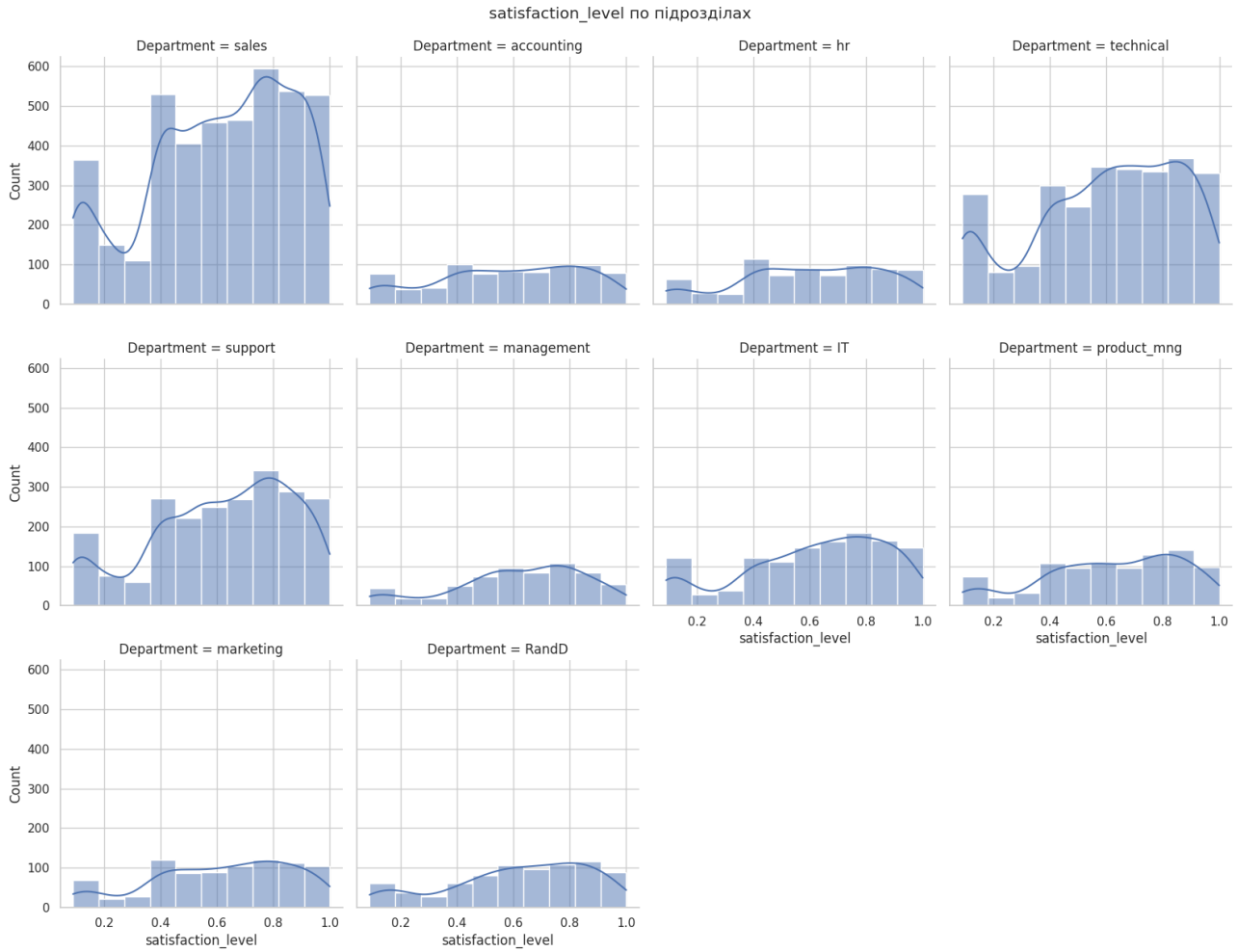

Рисунок 3.11 – Розподіл рівню задоволеності працівників залежно від відділу

На парних графіках числових змінних (рисунок 3.12) можна побачити цікаві закономірності, що стосуються співробітників, які залишили компанію (left = 1), порівняно з тими, хто залишився (left = 0):

1. satisfaction_level: більшість тих, хто залишив компанію, мали або дуже низький, або дуже високий рівень задоволення. Низький рівень задоволення – очевидний чинник плинності.

2. last_evaluation: у тих, хто залишив компанію, можна побачити дві групи: одні з низькою оцінкою, інші – з високою, що може вказувати або на незадоволеність, або на вигорання.

3. average_montly_hours: є тенденція, що звільнення частіше траплялись при дуже великій або дуже малій кількості робочих годин.

4. time_spend_company: сплеск звільнень спостерігається після 3–4 років роботи, а також для працівників із довгим стажем (6–7 років).

5. Кореляції між змінними: satisfaction_level та left візуально мають сильну зворотну залежність – чим нижче задоволення, тим вища ймовірність звільнення. Також average_montly_hours і last_evaluation мають певну позитивну залежність – вищі оцінки зазвичай пов'язані з більшим обсягом роботи.

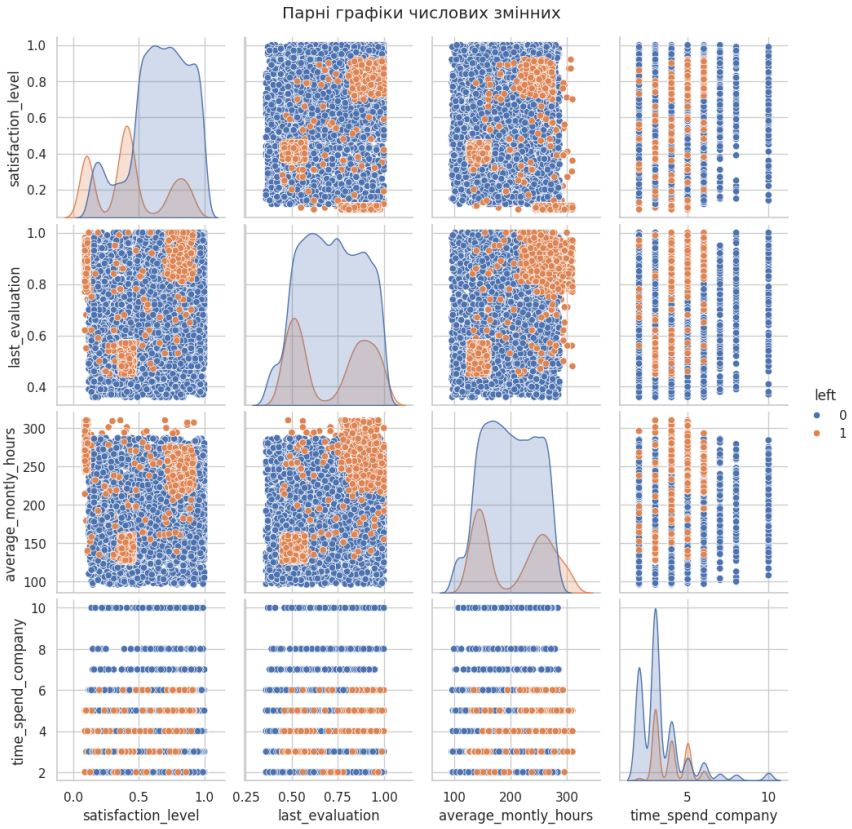

Рисунок 3.12 – Парні графіки числових змінних

Розглянемо кореляцію ознак. Результати наведено у вигляді теплової карти на рисунку 3.13. Бачимо значну негативну кореляцію між ознакою «satisfaction_level» та цільовою ознакою. Позитивна кореляція з найбільшим коефіцієнтом у ознаки «time_spend_company» та цільовим полем. Зауважимо також кореляцію, що показує ситуацію коли співробітники, які мають низьку заробітну плату та відсутнє підвищення, частіше залишають компанію.

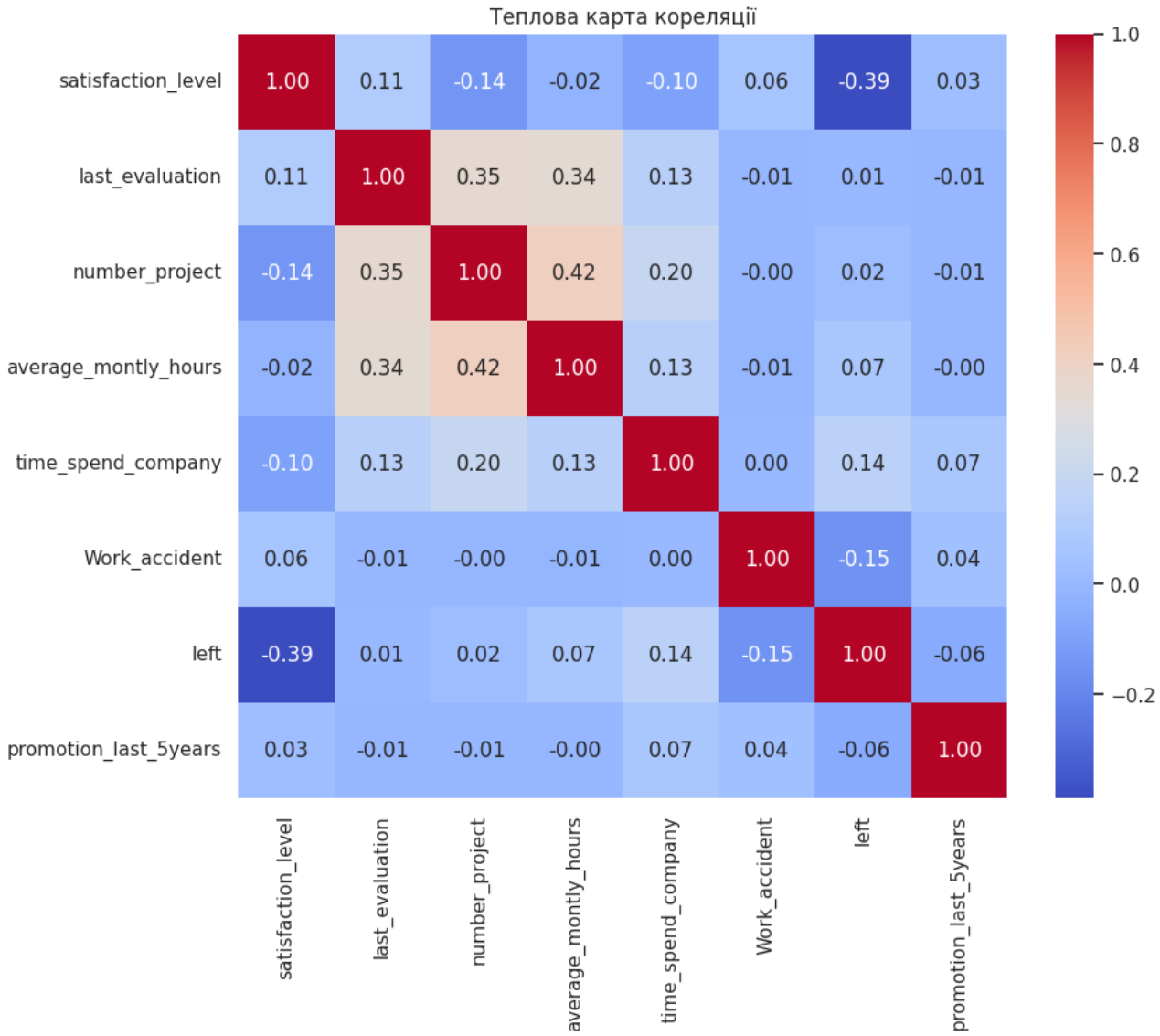

Рисунок 3.13 – Кореляція ознак

Перейдемо до аналізу та нормалізації даних. У ході цього процесу здійснимо безпосередньо сам аналіз даних та здійснимо перевірку на наявність пропущених значень, аномалій та інших помилок у даних. Нормалізація

включає масштабування і перетворення категоріальних ознак у числові значення.

Масштабування – це процес приведення значень ознак до того самого діапазону. Це дозволяє моделі більш ефективно використовувати ознаки та уникнути можливих спотворень результатів, які могли б виникнути через різні масштаби ознак.

Також слід враховувати наявність категоріальних ознак. Категоріальні змінні представляють кількісні дані, записані у вигляді тексту. Якщо даних присутні категоріальні змінні, їх необхідно перетворити на числові значення, оскільки більшість алгоритмів ML приймають лише числові дані.

Попередня обробка даних також включає поділ даних на дві частини: навчальний набір даних і тестовий набір даних.

Навчальний набір даних використовується для навчання моделі, а тестовий набір даних використовується для оцінки якості моделі нових, раніше не бачених даних. Зазвичай дані поділяють у співвідношенні, наприклад, 80% для навчання та 20% для тестування.

Навчання моделі включає передачу навчальних даних у модель та налаштування її параметрів з використанням алгоритму навчання. Модель аналізує навчальні дані та «навчається» знаходити певні закономірності або патерни в даних.

Оскільки в даних присутні категоріальні ознаки salary та department, їх необхідно перетворити на числові перед моделюванням. Для цього проведемо бінаризацію, створивши фіктивні змінні.

Крім того, виконаємо масштабування в діапазоні від 0 до 1 для ознак number_project, average_montly_hours, time_spend_company, скориставшись методом Min-Max Scaling із бібліотеки scikit-learn. Застосування масштабування зробить значення ознак порівнянними, що підвищить стабільність навчання алгоритму.

Зрештою, вийшла таблиця з даними в діапазоні від 0 до 1, і в ній є 21 стовпець (ознака). Нормалізація даних для моделювання виконана. Наступним

кроком здійснимо розбиття даних на навчальну та тестову вибірку. Розділимо дані у співвідношенні 80% для навчання та 20% для тестування. Навчання алгоритму включає передачу навчальних даних та налаштування параметрів. Алгоритм аналізує навчальні дані та «навчається» знаходити певні закономірності або патерни в даних.

```
# Застосуємо one-hot encoding до категоріальної ознаки "department"
dataset = pd.get_dummies(dataset_s, columns=['Department'], prefix='department')
dataset = pd.get_dummies(dataset, columns=['salary'], prefix='salary')
print(dataset.head(5))
# Вибираємо ознаки, які потребують масштабування
features_to_scale = ['number_project', 'average_montly_hours', 'time_spend_company']
# Створюємо об'єкт MinMaxScaler
scaler = MinMaxScaler()
# Застосовуємо масштабування лише до вибраних ознак
dataset[features_to_scale] = scaler.fit_transform(dataset_s[features_to_scale])
# Результат: ознаки тепер знаходяться в діапазоні від 0 до 1
dataset.head(5)
```

3.2 Реалізація моделей машинного навчання

Перед реалізацією алгоритму випадковий ліс необхідно визначитися із значеннями гіперпараметрів.

3.2.1 Випадковий ліс

Налаштування проводиться з метою досягти оптимального балансу між узагальнюючою здатністю та запобіганням перенавчанню.

```
# Випадковий ліс
rf = RandomForestClassifier()
rf.fit(X_train, y_train)
print('Випадковий ліс:')
predictions_rf = rf.predict(X_test)
print(classification_report(y_test, predictions_rf))
print(confusion_matrix(y_test, predictions_rf))
print('Випадковий ліс =', accuracy_score(y_test, predictions_rf))
```

Після створення алгоритму та його навчання зробимо прогноз на тестовому наборі даних. Результати представлено у звіті та матриці помилок (рисунок 3.14).

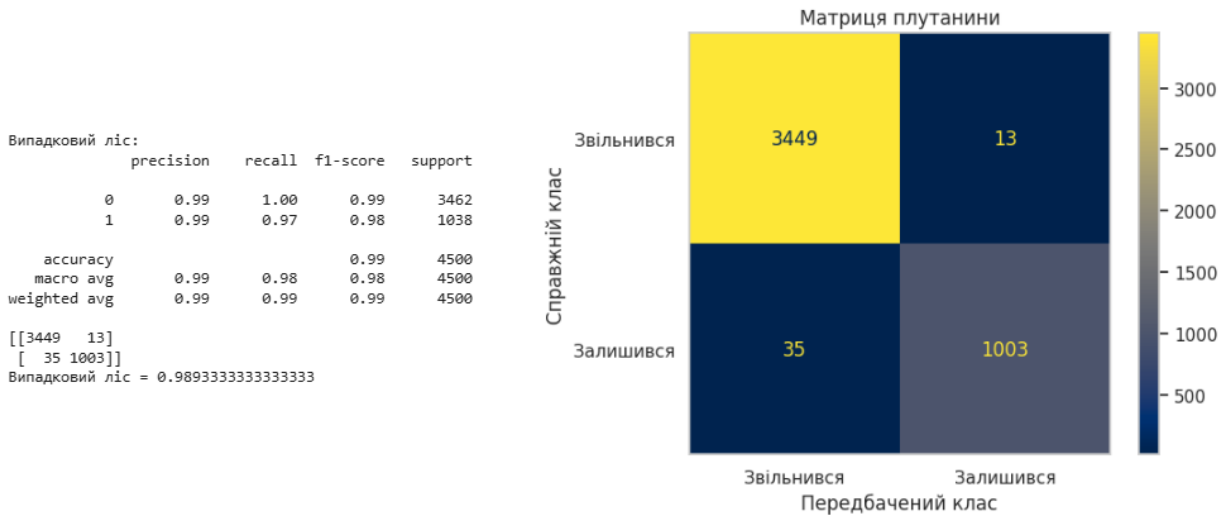

Рисунок 3.14 – Результат передбачення алгоритму випадковий ліс

Навіть без налаштування гіперпараметрів точність передбачення алгоритму виявилася дуже високою. Не зайвим буде згадати, що досягти цього вдалося завдяки ефективній нормалізації даних. Алгоритм має дуже хорошу здатність класифікації та виявляє як позитивні, так і негативні випадки з високою точністю.

3.2.2 Дерева рішень

Наступним алгоритмом буде дерево рішень.

```
# Дерево рішень
dt = DecisionTreeClassifier()
dt.fit(X_train, y_train)
print('Дерево рішень:')
predictions_dt = dt.predict(X_test)
print(classification_report(y_test, predictions_dt))
print(confusion_matrix(y_test, predictions_dt))
print('Дерево рішень =', accuracy_score(y_test, predictions_dt))
```

Зробимо прогноз на тестовому наборі даних і зведемо результати до звіту (рисунок 3.15).

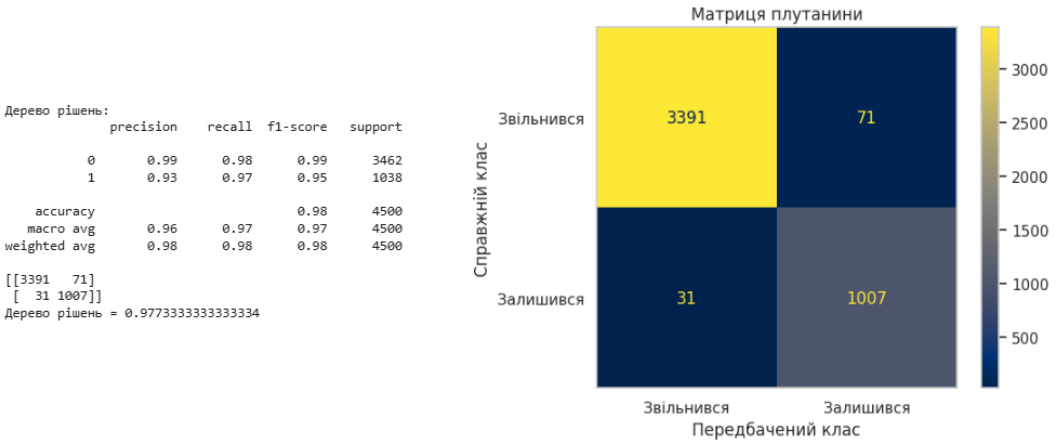

Рисунок 3.15 – Результат передбачення алгоритму дерева рішень

Даний алгоритм демонструє високу точність передбачення, але, незважаючи на це, він програє за результатами випадкового лісу по всіх метриках, хоч і незначно. Алгоритм дерево рішень робить більше помилок при пророкуванні негативних випадків, тобто алгоритм передбачив, що співробітник не буде звільнений, у той час як він фактично звільнився порівняно з випадковим лісом – 71 проти 13.

Фрагмент дерева зображений на рисунку 3.16. Це дає можливість проаналізувати його з точки зору виявлення асоціативних залежностей. Бачимо, що перший вузол використовує правило прийняття рішень (satisfaction_level <= 0.465).

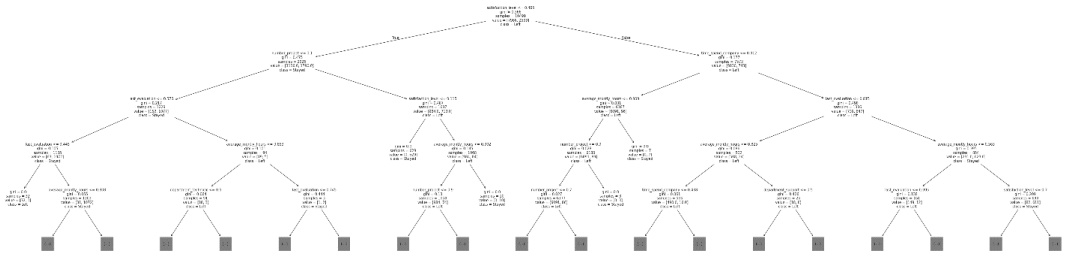

Рисунок 3.16 – Фрагмент дерева рішень

Далі, якщо «так», то аналізується ознака «number_project», а якщо «ні», то ознака «time_spend_company». Це має свій сенс, тож метод дерева рішень отримав свою популярність тому, що він імітує людський процес прийняття рішень, є наочним та добре інтерпретується.

3.2.3 Градієнтний бустинг

Далі навчимо модель градієнтного бустингу. Результати передбачення зведемо до таблиці та матриці помилок (рисунок 3.17).

```
# Створюємо модель градієнтного бустингу
gb = GradientBoostingClassifier()
# Навчаємо модель на навчальних даних
gb.fit(X_train, y_train)
print('Градієнтний бустинг:')
predictions_gb = gb.predict(X_test)
print(classification_report(y_test, predictions_gb))
print(confusion_matrix(y_test, predictions_gb))
print('Градієнтний бустинг =', accuracy_score(y_test, predictions_gb))
```

Градієнтний бустинг також демонструє високу точність прогнозування. Тим не менш, мінімальна точність прогнозу, що оцінюється за розглянутими метриками становить 93%, що є найменшим показником серед раніше проаналізованих алгоритмів.

В іншому градієнтний бустинг і дерево рішень показують схожі результати програючи при цьому випадковому лісі.

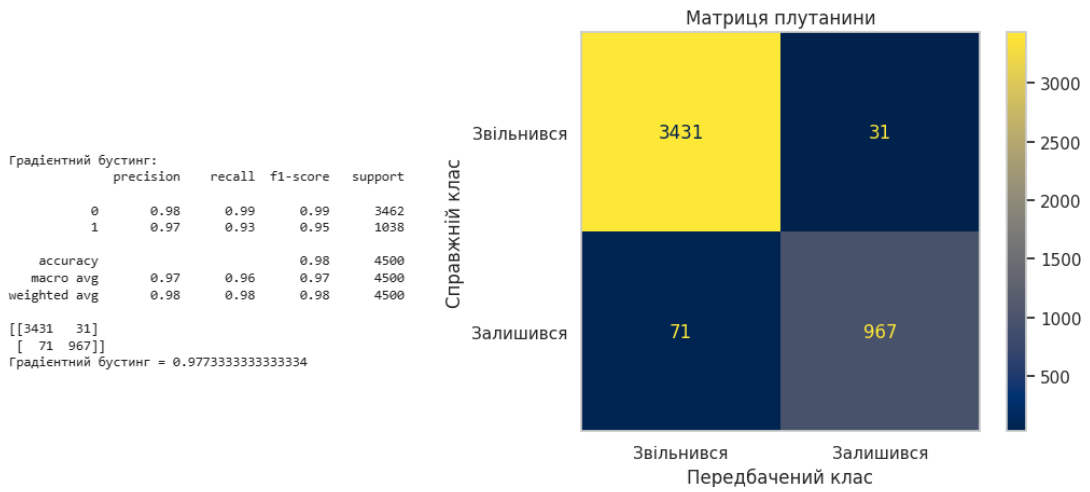

Рисунок 3.17 – Результат передбачення алгоритму градієнтний бустинг

Справді, градієнтний бустинг частіше помиляється у передбаченні звільнення співробітника, коли насправді не звільнився, проти іншими розглянутими алгоритмами.

3.2.4 Логістична регресія

Наступний алгоритм, який буде розглянуто у цій роботі, це логістична регресія. Дані, отримані в результаті прогнозування, зведемо до таблиці та матриці помилок (рисунок 3.18).

```
# Логістична регресія
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
lr = LogisticRegression()
lr.fit(X_train, y_train)
print('Логістична регресія:')
predictions_lr = lr.predict(X_test)
print(classification_report(y_test, predictions_lr))
print(confusion_matrix(y_test, predictions_lr))
print('Логістична регресія =', accuracy_score(y_test, predictions_lr))
```

Алгоритм показав найгірші результати, порівняно з попередніми алгоритмами. Виведемо коефіцієнти логістичної регресії, які були отримані з навченого алгоритму та оцінимо їх вплив на результат. Графік показаний на рисунку 3.19.

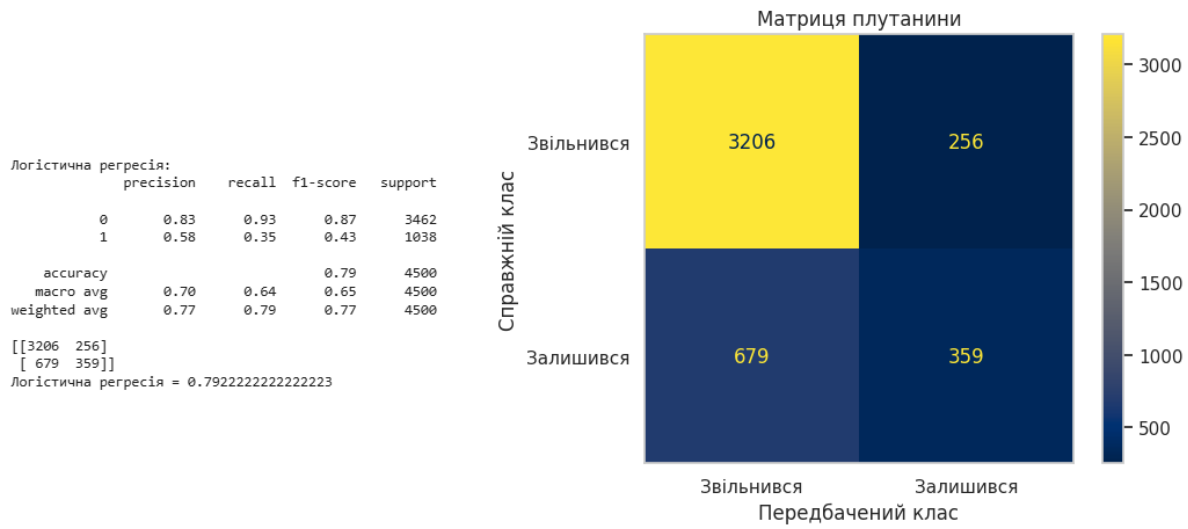

Рисунок 3.18 – Результат передбачення алгоритму логістична регресія

Отримані коефіцієнти логістичної регресії є показниками інтенсивності впливу кожної ознаки результат класифікації. Аналіз показує, що рівень задоволеності істотно впливає прийняття рішення віднести співробітника до класу 0 (не звільниться). Це пояснює низькі значення точності й повноти алгоритму, підкреслюючи, що алгоритм робить значну кількість помилок у передбаченнях, ґрунтуючись на даній конкретній ознаці.

Як можна помітити, алгоритм робить велику кількість помилок, як при визначенні співробітників, які можуть звільнитися, так і тих, що залишаться. Цей алгоритм явно не підходить для вирішення задачі прогнозування.

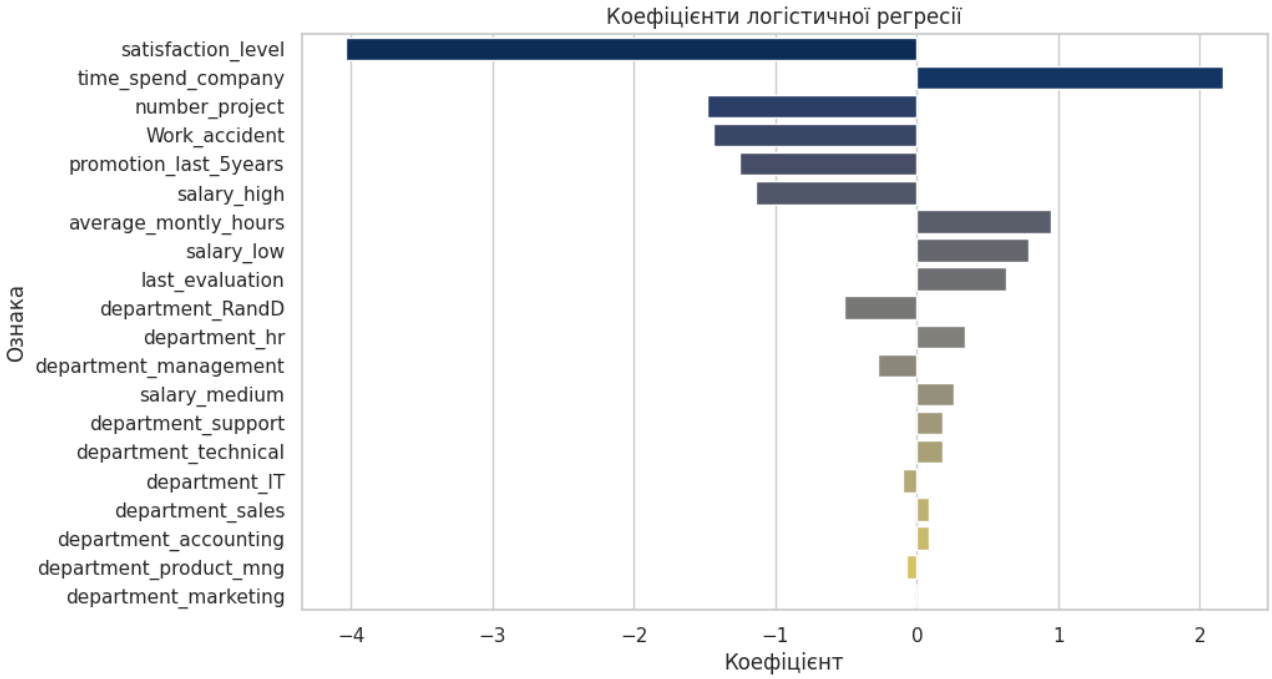

Рисунок 3.19 – Графік коефіцієнтів логістичної регресії

3.2.5 Багатошаровий персептрон

Наступний алгоритм, який буде розглянуто у цій роботі, це нейронна мережа багатошарового персептрону.

Архітектуру нейронної мережі показано на рисунку 3.20. Загальна кількість параметрів для навчання – 3,137.

```python
# додаємо шари
model = models.Sequential([
    layers.Input(shape=(X_train.shape[1],)),  # кількість ознак

    layers.Dense(64, activation='relu'), # другий шар
    layers.BatchNormalization(),
    layers.Dropout(0.3), # Dropout для уникнення перенавчання

    layers.Dense(32, activation='relu'), # третій шар
    layers.BatchNormalization(),
    layers.Dropout(0.3), # Dropout для уникнення перенавчання

    layers.Dense(1, activation='sigmoid')  # для бінарної класифікації
])

# Компілюємо модель
model.compile(
    optimizer='adam',
    loss='binary_crossentropy',
    metrics=['accuracy']
)
```

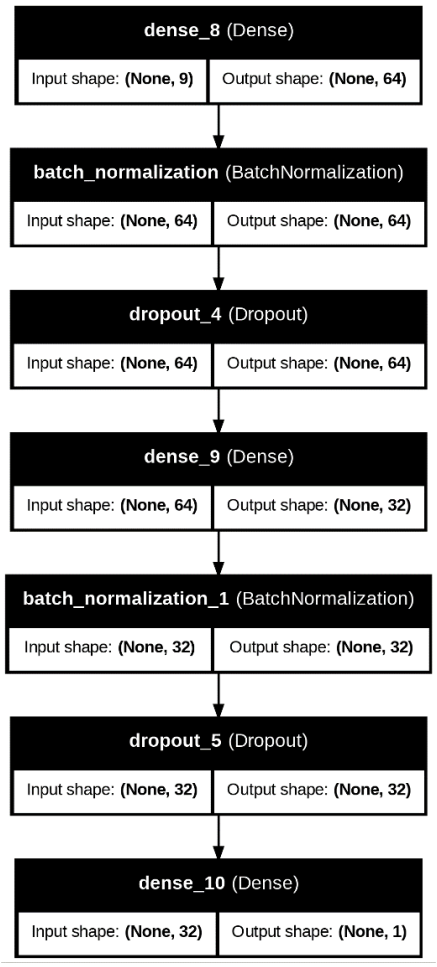

Рисунок 3.20 – Архітектура нейронної мережі

Далі навчаємо модель на протязі 100 епох. Точність на тестових даних: 96.78%.

Далі побудуємо графіки втрат і точності під час тренування (рисунок 3.21).

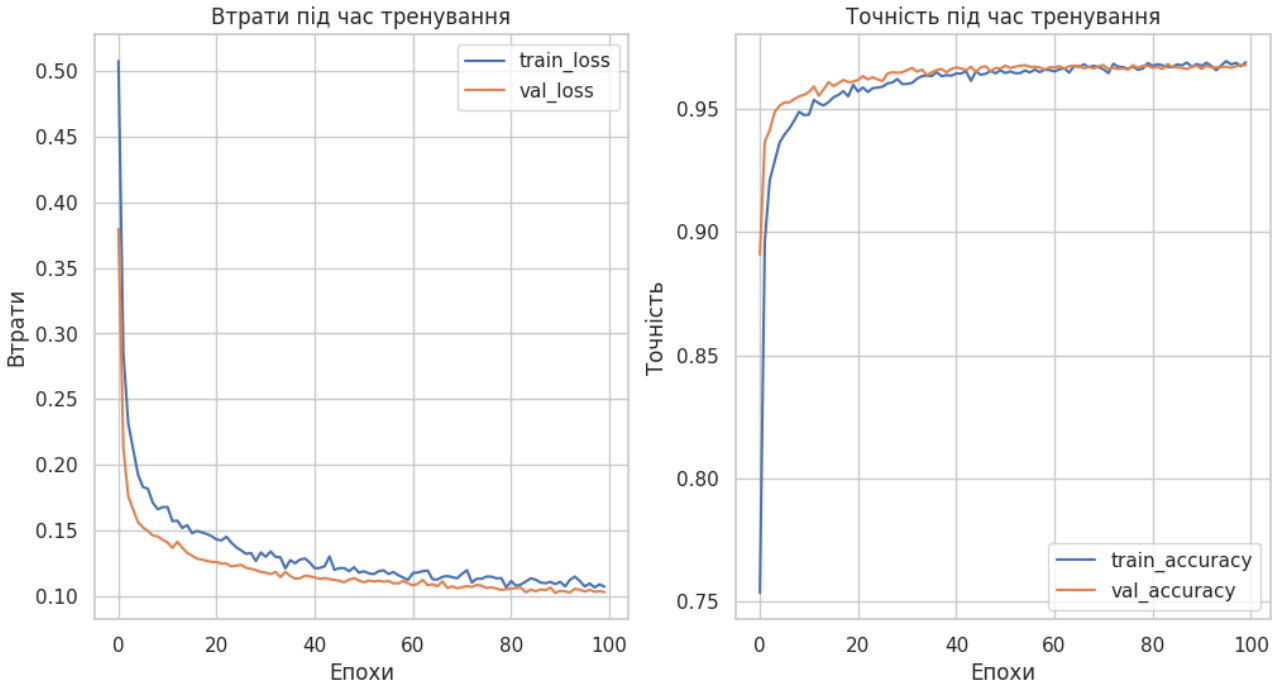

Рисунок 3.21 – Графіки втрат і точності під час тренування

І тренувальні (train_loss), і валідаційні (val_loss) втрати постійно зменшуються і сходяться, тобто модель не перенавчається. Значення val_loss трохи нижче за train_loss і це ознака хорошої узагальненості. І тренувальна, і валідаційна точність швидко зростають і досягають ≈97%, що свідчить модель добре навчається. Значення val_accuracy трішки вище за train_accuracy – це також говорить за те, що немає ознак перенавчання або недонавчання. Таким чином вибрана архітектура MLP дуже ефективна для цього датасету, модель добре узагальнює.

Дані, отримані в результаті прогнозування, зведемо до таблиці та матриці помилок (рисунок 3.22). Нейронна мережа також демонструє високу точність прогнозування, хоча це і не найкращій показник серед раніше проаналізованих алгоритмів.

Звіт по класифікації:

	precision	recall	f1-score	support
0	0.97	0.99	0.98	3428
1	0.97	0.90	0.93	1072
accuracy			0.97	4500
macro avg	0.97	0.94	0.95	4500
weighted avg	0.97	0.97	0.97	4500

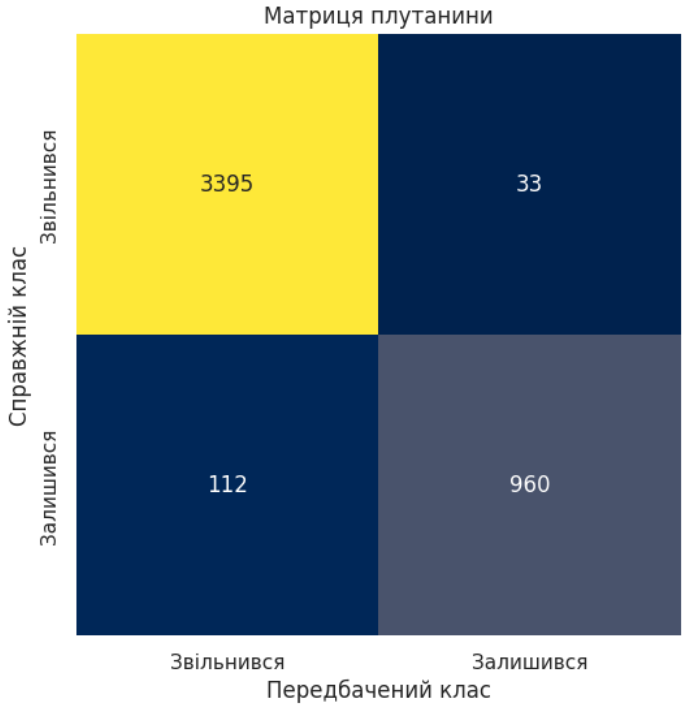

Рисунок 3.22 – Результат передбачення нейронної мережі

3.2.6 Порівняння результатів

Покажемо ROC-криву для 4 алгоритмів (рисунок 3.23).

Графік показує, що всі 4 алгоритми (для нейронної мережі не будуємо) демонструють майже ідеальні показники, адже ідеальний алгоритм – це алгоритм, який робить усі правильні передбачення і завдяки цьому матиме крапку у верхньому лівому куті графіка (TPR = 1, FPR = 0).

Тобто чим ближче крива ROC до верхнього лівого кута, тим краще алгоритм розділяє класи і робить більш точні передбачення.

Зробимо порівняльний аналіз, щоб визначити, який алгоритм підходить найбільше для вирішення задачі прогнозування плинності кадрів. Наведемо порівняльну таблицю 3.2.

Зведемо дані матриці помилок для 4 алгоритмів таблицю 3.3.

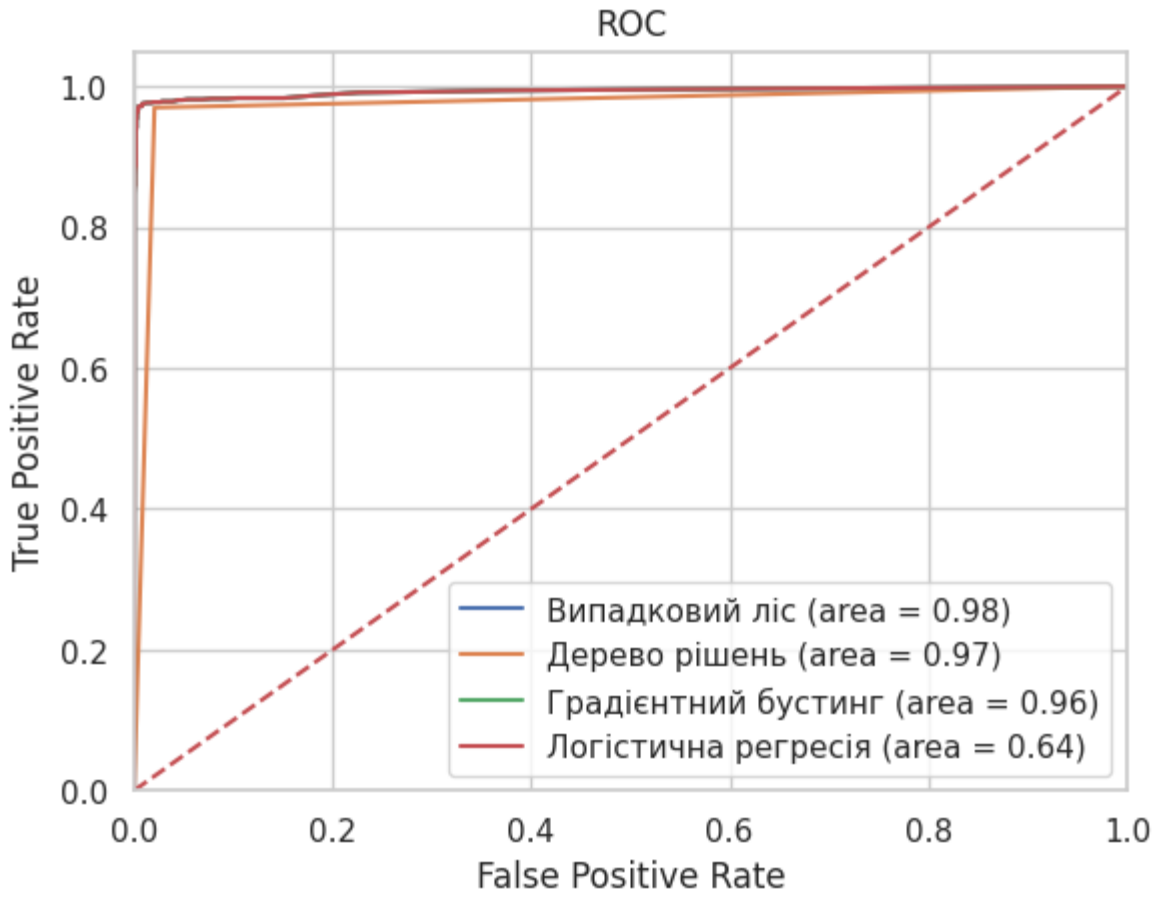

Рисунок 3.23 – ROC-крива для 4 алгоритмів

Таблиця 3.2 – Порівняльна таблиця за метриками

Алгоритм	Класифікація	Precision	Recall	F-score	Accuracy
Випадковий ліс	Звільнився	0,99	1,00	0,99	0,99
	Залишився	0,99	0,97	0,98	
Дерево рішень	Звільнився	0,99	0,98	0,99	0,98
	Залишився	0,93	0,97	0,95	
Градієнтний бустинг	Звільнився	0,98	0,99	0,99	0,98
	Залишився	0,97	0,93	0,95	
Логістична регресія	Звільнився	0,83	0,93	0,87	0,79
	Залишився	0,58	0,35	0,43	
Нейронна мережа	Звільнився	0,97	0,99	0,98	0,97
	Залишився	0,97	0,90	0,93	

Таблиця 3.3 – Порівняльна таблиця по матриці помилок

Алгоритм	True Positive	False Positive	False Negative	True Negative
Випадковий ліс	3449	13	35	1003
Дерево рішень	3391	71	31	1007
Градієнтний бустинг	3431	31	71	967
Логістична регресія	3206	256	679	359
Нейронна мережа	3395	33	112	960

З аналізу метрик за розглянутими алгоритмами, можна дійти висновку про перевагу випадкового лісу. Даний алгоритм демонструє високі показники з більшості метрик, що підтверджує його ефективність. Також зазначимо, що випадковий ліс робить менше помилок у передбаченнях у порівнянні з іншими алгоритмами. Відобразимо важливість ознак цього алгоритму на рисунку 3.24.

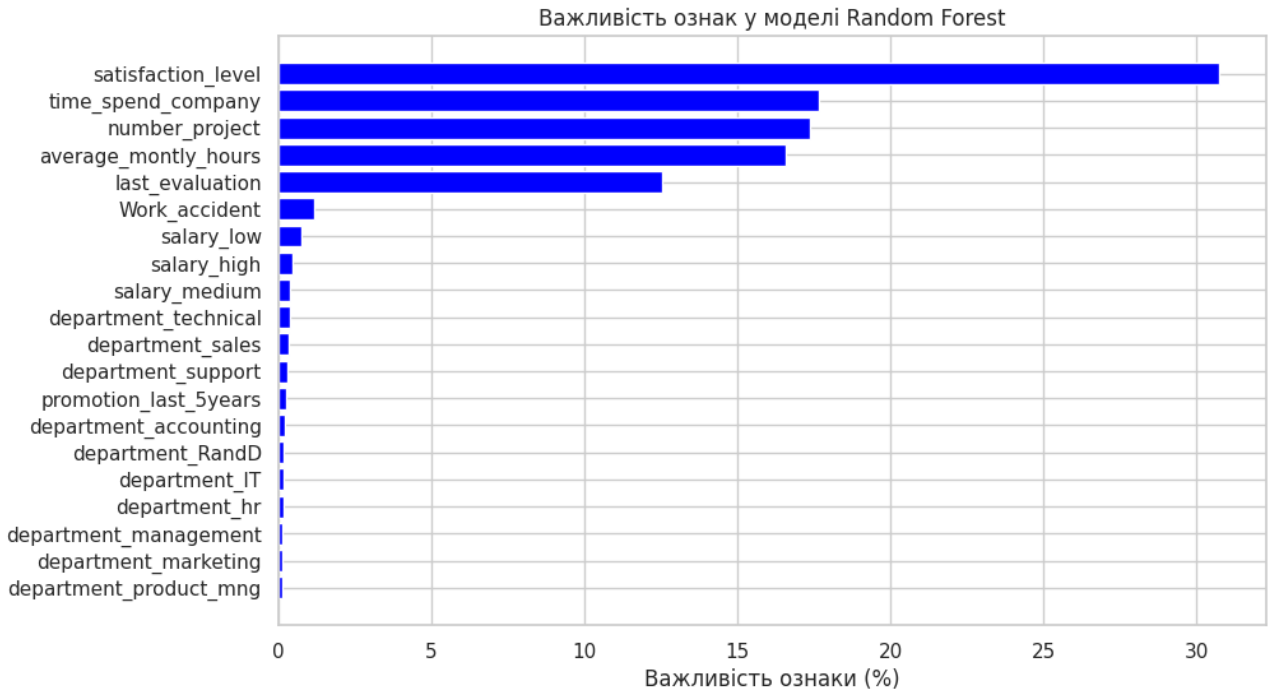

Рисунок 3.23 – Важливість ознак

Згідно з алгоритмом, єдиною найбільш важливою ознакою, що визначає, чи залишиться співробітник чи піде, є задоволеність.

Якщо співробітник задоволений своєю роботою, ймовірність того, що він покине компанію, є малоймовірною.

3.3 Інтерфейс системи прогнозування плинності кадрів

Систему передбачення плинності кадрів розроблено в Node.js, використаний фреймворк Vue для створення клієнтської частини (рисунок 3.24).

При натисканні кнопки «Прогнозування» ми викликаємо передбачення на попередньо навченій моделі машинного навчання. Ми можемо задати параметри та зробити прогноз.

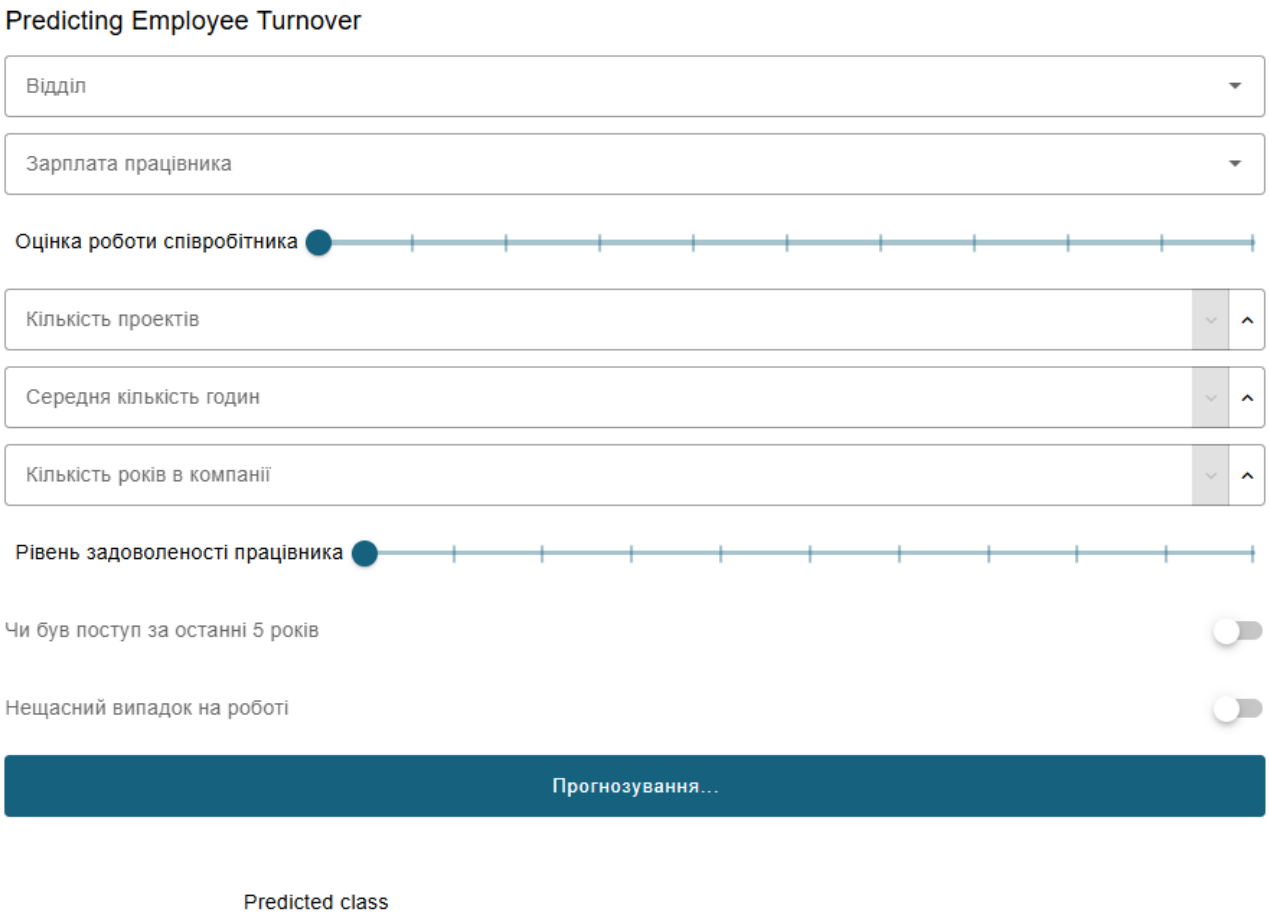

Рисунок 3.24 – Інтерфейс системи прогнозування плинності кадрів

3.4 Оцінка ефективності системи прогнозування, а також рекомендації для HR-менеджменту

На основі проведеного дослідження, яке включало EDA (Exploratory Data Analysis), машинне навчання та побудову кількох моделей, можна зробити такі ключові висновки:

1 EDA виявив низку значущих чинників плинності кадрів, серед яких: кількість проєктів, кількість годин на тиждень, рівень задоволеності та оцінка ефективності роботи працівника;

2 низька задоволеність та високе навантаження корелюють з підвищеною ймовірністю звільнення;

3 моделі машинного навчання показали дуже високий рівень точності: випадковий ліс: 0.99; дерева рішень: 0.98; градієнтний бустинг: 0.98; нейронна мережа (MLP): 0.97; логістична регресія: 0.79;

4 найкращу точність показала модель Random Forest, що дозволяє ефективно враховувати нелінійні залежності та взаємодії між змінними;

5 нейронна мережа (MLP) також виявилася стабільною та точною, з хорошим узагальненням і майже відсутністю перенавчання.

Таким чином, проводячи оцінку ефективності системи прогнозування плинності кадрів на базі моделей машинного навчання можна відмітити точність на рівні ≥97%, що говорить про те, що її можна використовувати в практиці IT-компаній.

Використання системи надає можливість завчасно ідентифікувати «ризикованих» працівників та надає HR-відділу об’єктивний інструмент для ухвалення рішень.

Загалом на основі дослідження можна сформувати такі рекомендації для HR-стратегії в IT-компанії

1. Моніторинг рівня задоволеності. Регулярне анкетування (опитування про добробут, навантаження, менеджмент тощо). Виявлення груп ризику для індивідуального опрацювання (коучинг, зміна ролі).

2. Зменшення перевантаження. Обмеження кількості проєктів для одного працівника. Визначення оптимального робочого навантаження на основі історичних даних.

3. Прогнозування та раннє втручання. Використання моделі в HR-системі для щотижневого оновлення ризиків. Реалізація “м’яких” інтервенцій (розмови, підтримка, бонуси).

4. Індивідуальні траєкторії розвитку. Запровадження персоналізованих планів кар’єрного зростання для високоефективних працівників із високим ризиком плинності.

5. Аналіз причин звільнень. Інтерв’ю “на виході” мають бути збережені як частина даних для подальшого навчання моделей.

Висновки за розділом

Проведено підготовку даних, побудовано моделі: дерево рішень, випадковий ліс, градієнтний бустинг, логістична регресія, нейронна мережа. Найкращий результат показав випадковий ліс – точність 99%. Побудовано інтерфейс користувача та здійснено оцінку ефективності системи. Побудовані моделі досягли високої точності. Найкращі результати отримано за допомогою ансамблевих методів. Система може ефективно застосовуватися в реальних HR-процесах для прогнозування ризику звільнення.

РОЗДІЛ 4
ОХОРОНА ПРАЦІ

4.1 Регулювання питань охорони праці на законодавчому рівні

Законодавство з охорони праці має важливе значення як перший крок для впровадження вдосконалених заходів з безпеки та охорони праці, а також для запобігання нещасним випадкам та захворюванням у сфері праці.

Законодавство з питань охорони праці вирішує завдання із створення безпечного та здорового робочого середовища як на національному рівні, так і на рівні підприємства.

На національному рівні воно включає як загальні закони (для України це, в першу чергу, закон «Про охорону праці», Кодекс законів про працю, Закон України «Про загальнообов'язкове державне соціальне страхування»), так і інші законодавчі акти, які розкривають загальні вимоги з охорони праці у галузевому контексті (наприклад, для IT-сфери це ДСТУ EN 60950-1:2015 Обладнання інформаційних технологій. Безпека. Частина 1. Загальні вимоги (EN 60950-1:2006; A11:2009; A1:2010; A12:2011; AC:2011; A2:2013, IDT), ДСТУ EN 62368-1:2017 Обладнання аудіо-, відео-, інформаційних та комунікаційних технологій. Частина 1. Вимоги щодо безпеки (EN 62368-1:2014; AC:2015-05; AC:2015-02; AC:2015-11; AC:2017; A11:2017; IDT; IEC 62368-1:2014, MOD; Cor 1:2014; Cor 2:2015, IDT) та ДСТУ EN 61140:2019 Захист від ураження електричним струмом. Загальні аспекти щодо установок та обладнання (EN 61140:2016, IDT; IEC 61140:2016, IDT)).

На рівні підприємства законодавство може охоплювати питання від етапу планування та проектування підприємства та постачання обладнання та речовин до питань технічного обслуговування, що проходять через обов'язки роботодавця, права працівників та екологічні міркування.

Охорона праці та гігієна праці – це широка сфера, законодавчі підстави якої формуються із урахуванням діяльності міжнародних організацій та кращих міжнародних практик.

Відповідно до Рекомендації Міжнародної організації праці 2006 року про основи сприяння безпеці та гігієні праці (№ 197), національний профіль з охорони праці повинен включати, серед іншого, інформацію з таких елементів: закони та нормативні акти, відповідальні органи або організації, механізми забезпечення дотримання національних законів та нормативних актів та послуги відповідно до національного законодавства та практики.

Багато країн визначили безпеку та охорону праці як один із пріоритетів у своїй Програмі гідної праці в країнах. Зміцнення умов безпеки та охорони праці на робочому місці та дотримання відповідного законодавства з охорони праці часто є ключовими аспектами цих планів. Ця база даних допомагає країнам забезпечувати безпечні та здорові умови праці та дозволяє їм навчатися на більш передових законах та критеріях. Вона допомагає формувати національні стратегії, політику та програми з охорони праці.

Вимоги законодавства, які формують інституційні програми управління та інститути, пов'язані з адмініструванням чи забезпеченням охорони праці, мають прямі наслідки для працівників та підприємств. Поліпшення охорони та гігієни праці для працівників та підприємств є провідною опорою та центральною метою у створенні надійного законодавства з охорони праці.

Завдяки сьогоднішнім досягненням у галузі технологій люди все частіше працюють з обладнанням з дисплеєм у своїй повсякденній роботі. Роботодавець завжди повинен робити все можливе, щоб гарантувати, що його співробітники не ризикують своїм здоров'ям під час роботи з технологічним обладнанням, таким як комп'ютери, ноутбуки, планшети.

Індустрія інформаційних технологій (IT) відома своїм динамічним середовищем, швидким технологічним прогресом та висококваліфікованою робочою силою. Хоча основна увага в секторі IT приділяється інноваціям та наданню послуг, зростає визнання важливості охорони праці та безпеки у цій галузі. Традиційно проблеми охорони праці та безпеки були помітнішими у таких галузях, як виробництво, будівництво та гірничодобувна промисловість,

де фізичні небезпеки більш очевидні, однак у ІТ-галузі є свій унікальний набір проблем, які потребують надійної структури охорони праці.

4.2 Виявлення потенційних небезпек стосовно об’єкту проектування

У дипломній роботі бакалавра розглянуто проблему плинності кадрів в ІТ-компаніях та можливості її прогнозування з використанням методів машинного навчання. ІТ-сфера, незважаючи на «безпеку» офісної роботи, характеризується низкою факторів, які можуть несприятливо впливати на здоров'я співробітників. Умови праці залежать від ролі (розробник, аналітик, менеджер, адміністратор тощо), але загальна проблема для більшості працівників цієї сфери – тривале перебування за комп'ютером, високе розумове навантаження і обмежена фізична активність.

Таблиця 4.1 – Аналіз небезпечних та шкідливих виробничих факторів

Фактор	Прояв	Наслідки
1	2	3
Підвищене зорове навантаження	Робота за екраном 6-10 годин на день	Зорова втома, розвиток синдрому сухого ока, погіршення зору
Неправильне освітлення	Відсутність природного освітлення, мерехтіння ламп штучного освітлення	Головний біль, стомлюваність, погіршення зору
Шум	Робота в опенспейсах, дзвінки, робота вентиляторів, техніки	Дратівливість, погіршення концентрації уваги
Неправильна ергономіка робочого місця	Невідповідний стілець, висота монітора, клавіатури	Порушення постави, болі в спині та шиї, тунельний синдром
Вплив офісної хімії	Засоби для чищення, картриджі, пил	Алергічні реакції, подразнення дихальних

		шляхів
Високі розумові навантаження	Стрес через дедлайни, помилки, перевантаження	Хронічний стрес, синдром вигоряння, тривожні розлади
Ненормований робочий день	Понаднормова робота, чергування, віддалений формати роботи	Порушення сну, втома, погіршення Work-Life Balance
Інформаційне навантаження	Потік повідомлень, листів, повідомлень	Розсіяність, тривожність, зниження продуктивності
Відсутність гігієни в офісі	Загальні клавіатури, миші, кондиціонери	Простудні захворювання, передача вірусів
Ізоляція при віддаленій роботі	Відсутність живого спілкування	Відчуття самотності, депресія

Накопичувальний вплив шкідливих факторів на працівників IT-сфери може призвести до:

- остеохондрозу та сколіозу;
- синдрому зап'ясткового каналу (тунельного синдрому);
- синдрому сухого ока та зниження гостроти зору;
- неврозам, тривожності, вигорянню;
- хронічної втоми та порушення сну;
- гіподинамії та пов'язаних з нею хвороб (ожиріння, гіпертонія).

Окрім того, потрібно ураховувати небезпеки і більш загального характеру, які також можуть виникати на робочих місцях працівників, наприклад, пожежа, яка може завдавати значної шкоди людям та майну.

4.3 Дослідження ризику реалізації потенційних небезпек на об’єкті проектування та розробка заходів щодо їх попередження

Оцінка ризиків являє собою процедуру, яка застосовується в рамках організації для встановлення структури небезпек, які впливають на працівників, виявлення тих небезпек, які становлять загрозу найбільшою мірою та, як результат, розробки заходів щодо зменшення рівня їх ризику.

Для оцінки ризиків організація може використовувати будь-який доступний їй метод, адаптувати його під свої можливості, ресурси та потреби. Одним із простих та доступних методів є матриця оцінки ризиків.

Матриця оцінки ризиків, також відома як Матриця ризиків або Матриця контролю ризиків, є візуальним інструментом, який широко використовується в управлінні проектами та інших областях. Вона діє як сіткова діаграма для оцінки потенційних ризиків на основі двох ключових критеріїв:

Можливість ризику (імовірність): наскільки ймовірно, що ризик здійсниться? (можливо, можливо, малоймовірно, дуже малоймовірно).

Вплив ризику (серйозність): наскільки серйозними можуть бути наслідки? (дуже високі, високі, помірні, низькі, дуже низькі).

Нанесення кожного виявленого ризику на матрицю, виходячи з його ймовірності та впливу, дає чітке уявлення про потенційні загрози. Це дозволяє зрозуміти різні типи ризиків, зосередитися на тих, які вимагають найбільшої уваги при плануванні подальших дій, визначити відповідні дії для кожного виявленого ризику, щоб пом'якшити чи усунути його вплив, а також візуалізувати відносну важливість різних ризиків, полегшуючи прийняття обґрунтованих рішень.

Матриці ризику зазвичай використовують квадратну сітку зі шкалами ймовірності та впливу. Звичайні розміри матриць включають:

Матриця 3×3 (проста та ефективна): це найбільш базовий та широко використовуваний формат. Така матриця доцільна для використання для

проектів з керованою кількістю ризиків або для проектів, де необхідний швидкий та чіткий огляд основних категорій ризиків.

Матриця 4×4 (баланс між деталізацією та простотою): порівняно з матрицею 3×3 ця сітка забезпечує більш детальне представлення ризиків, пропонуючи додаткові категорії як ймовірності, так впливу. Це дозволяє охопити ширший спектр ризиків та рівнів їх серйозності, що робить її придатною для проектів середньої складності.

Матриця 5×5 (дуже докладна та точна): ця сітка забезпечує найбільш докладну оцінку ризиків. Вона найкраще підходить для складних проектів з безліччю потенційних ризиків та необхідністю точної оцінки ризиків.

Таблиця 4.4 – Матриця оцінювання ризиків

Неправильна ергономіка робочого місця				
Визначення категорії серйозності небезпеки		Визначення рівня ймовірності небезпеки		Індекс ризику небезпеки
1		2		
Вид, категорія	Опис	Вид, рівень	Опис	
III – гранична	Порушення постави, болі в спині та шиї, тунельний синдром	А	Ситуація, яка має місце у значній кількості компаній	3 А – неприпустимий (надмірний) рівень ризику

Ергономічні небезпеки часто зустрічаються в діяльності організацій, де працівники за характером своєї роботи багато працюють за комп’ютером. Тому контроль стану та обладнання на їх робочих місцях – важливе завдання.

1. Робочий стілець або робоче крісло. Гарний стілець – основа ергономічного робочого місця.

Регульована висота: сидіння повинно дозволяти стопам повністю стояти на підлозі, а коліна бути під кутом ~90°.

Підтримка попереку: наявність вбудованої або додаткової поперекової підтримки.

Спинка з нахилом: регульована по куту нахилу та підтримуюча хребет.

Підлокітники: регульовані по висоті та ширині, щоб плечі залишалися розслабленими.

Матеріал: тканина, що дихає, зручна для тривалого сидіння.

2. Положення монітора. Неправильна висота або дистанція до екрану – часта причина втоми очей та болю у шиї.

Висота: верхня межа екрана на рівні очей або трохи нижче.

Відстань: 50-70 см від очей до екрана (приблизно витягнута рука).

Нахил екрана: легкий нахил назад (~10-20°).

Кількість моніторів:

При двох екранах – основний монітор центром, другий під кутом.

При використанні ноутбука – зовнішній монітор чи підставка.

3. Клавіатура та миша. Положення рук та зап'ясть критично для запобігання тунельному синдрому.

Висота клавіатури: лікті зігнуті під кутом ~90°, зап'ястя у нейтральному положенні.

Підставка під зап'ястя: м'яка, що підтримує, але не притискає зап'ястя до столу.

Положення миші: поруч із клавіатурою, одному рівні.

Альтернативи: вертикальні миші, трекболи зменшують навантаження на зап'ястя.

4. Робочий стіл. Стіл повинен підтримувати правильну поставу та бути функціональним.

Висота: 70-75 см для сидячої роботи (або регульована).

Глибина: щонайменше 80 см, щоб екран був на безпечній відстані.

Ширина: достатньо простору для обладнання та паперів (мінімум 120 см).

Покриття: матове, щоб уникнути відблисків.

5. Освітлення. Правильне освітлення знижує навантаження очі.

Природне світло: розташування робочого місця біля вікна (але без відблисків на екрані).

Штучне освітлення: настільна лампа з регулюванням яскравості та напряму.

Температура світла: нейтрально-біла (4000-5000 К) для забезпечення комфорту роботи.

6. Організація робочого ритму.

Правило 20-20-20: кожні 20 хвилин дивитися на об'єкт за 6 м на 20 секунд.

Перерви: кожні 60 хвилин – встати, розім'ятися, пройтися.

Розминка: обертання плечима, нахили голови, розтяжка рук та зап'ясть.

4.4 Висновки по розділу

У розділі з охорони праці розглянуто основні положення законодавства з охорони праці, формування якого на національному рівні здійснюється із урахуванням Декларацій, Рекомендацій та Конвенцій, розробниками яких є міжнародні та регіональні організації з питань охорони праці та безпеки, положення яких повинні використовувати країни-члени цих організацій, в тому числі і Україна, чиє законодавство, зокрема, закон «Про охорону праці», відображає найбільш актуальні та ефективні положення.

Також у розділі приділено увагу аналізу умов праці працівників ІТ-сфери, визначено основні шкідливі фактори, які впливають на них та їх наслідки та розглянуто рекомендації із покращення умов праці за рахунок оптимізації робочого місця працівника, що є особливо важливим, враховуючи те, що такі працівники багато часу проводять за комп’ютером.

ВИСНОВКИ

Роботу присвячено розробці системи прогнозування плинності кадрів в IT-компаніях з використанням моделей машинного навчання.

В ході виконання роботи було вирішені наступні задачі:

1 Проведено аналіз причин та наслідків плинності кадрів в IT-секторі.

2 Досліджено можливості HR-аналітики та застосування методів машинного навчання для передбачення плинності.

3 Здійснено вибір датасету та попередній аналіз даних.

4 Реалізовано моделі машинного навчання та проведено порівняльний аналіз ефективності моделей.

5 Розроблено інтерфейс користувача для демонстрації роботи системи прогнозування.

6 Надано практичні рекомендації для HR-менеджменту на основі результатів прогнозування..

Таким чином, у межах роботи здійснено дослідження проблеми плинності кадрів в IT-компаніях із використанням сучасних методів машинного навчання. Побудовані моделі продемонстрували високу точність, а система прогнозування, розроблена на їх основі, може стати дієвим інструментом для HR-відділу. Запропоновано рекомендації, що базуються на аналітичних результатах та можуть сприяти зниженню плинності та покращенню взаємодії з персоналом. Робота доводить ефективність інтеграції аналітичних підходів у процес управління людськими ресурсами.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1 Raza A, Munir K, Almutairi M, Younas F, Fareed MMS. Predicting Employee Attrition Using Machine Learning Approaches. Applied Sciences. 2022; 12(13):6424. https://doi.org/10.3390/app12136424

2 Chang V. et al. Job satisfaction and turnover decision employees in the Internet sector в США //Enterprise Information Systems. - 2023. - Т. 17. - №. 8. 2130013.

3 Дослідженні IT Research Ukraine 2023: Адаптивність та стійкість під час війни – Режим доступу: https://itcluster.lviv.ua/projects/it-research-ukraine/.

4 Fallucchi F, Coladangelo M, Giuliano R, William De Luca E. Predicting Employee Attrition Using Machine Learning Techniques. Computers. 2020; 9(4):86. https://doi.org/10.3390/computers9040086

5 Kumar P, Gaikwad SB, Ramya ST, Tiwari T, Tiwari M, Kumar B. Predicting Employee Turnover: A Systematic Machine Learning Approach for Resource Conservation and Workforce Stability. Engineering Proceedings. 2023; 59(1):117. https://doi.org/10.3390/engproc2023059117

6 Guerranti F, Dimitri GM. A Comparison of Machine Learning Approaches for Predicting Employee Attrition. Applied Sciences. 2023; 13(1):267. https://doi.org/10.3390/app13010267

7 Chakraborty R. et al. Study and prediction analysis of the employee turnover using machine learning approaches //2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON). - IEEE, 2021. - С. 1-6.

8 Marín Díaz G, Galán Hernández JJ, Galdón Salvador JL. Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making. Mathematics. 2023; 11(22):4677. https://doi.org/10.3390/math11224677

9 Liu M, Yang B, Song Y. Research on Predicting the Turnover of Graduates Using an Enhanced Random Forest Model. Behavioral Sciences. 2024; 14(7):562. https://doi.org/10.3390/bs14070562

10 Aham Edward Kanuto, 2024. "Identifying Patterns and Predicting Employee Turnover Using Machine Learning Approaches," International Journal of Science and Business, IJSAB International, vol. 36(1), pages 20-35.

11 D. Alao and A. Adeyemo, “Analyzing employee attrition using decision tree algorithms,” Computing, Information Systems, Development Informatics and Allied Research Journal, vol. 4, no. 1, pp. 17–28, 2013.

12 R. Jain and A. Nayyar, “Predicting employee attrition using xgboost machine learning approach,” in 2018 international conference on system modeling & advancement in research trends (smart). IEEE, 2018, pp. 113–120.

13 Al-Darraji S, Honi DG, Fallucchi F, Abdulsada AI, Giuliano R, Abdulmalik HA. Employee Attrition Prediction Using Deep Neural Networks. Computers. 2021; 10(11):141. https://doi.org/10.3390/computers10110141

14 Tanasescu LG, Vines A, Bologa AR, Vîrgolici O. Data Analytics for Optimizing and Predicting Employee Performance. Applied Sciences. 2024; 14(8):3254. https://doi.org/10.3390/app14083254

15 Датасет Human resource на Kaggle // Режим доступу: https://www.kaggle.com/datasets/kuniowu/human-resource

16. Закон України «Про охорону праці». – Офіційний сайт Верховної Ради України. – Режим доступу: https://zakon.rada.gov.ua/laws/show/2694-12.

17. Офіційний сайт Міжнародної організації праці. – Режим доступу: https://www.ilo.org/.