



# ДОДАТОК А

## Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ

StrikePlagiarism.com  Дата звіту 6/17/2025  
Дата редагування ---  Звіт не був оцінений

### Звіт подібності

#### метадані

Назва організації  
**Kharkiv National University of Radio Electronics**  
Заголовок  
**2025\_M\_ПІ\_ІПЗ-23-4\_Потьомкіна\_K\_O\_скорочений**  
Автор Науковий керівник / Експерт  
**Потьомкіна Катерина Олексівна Олена Олійник**  
підрозділ  
**каф. ПІ**

#### Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.

**2.98%**  
2.98% КП 1

**0.91%**  
0.91% КЦ






**25**  
Довжина фрази для коефіцієнта подібності 2

**7120**  
Кількість слів

**57457**  
Кількість символів

#### Тривога

У цьому розділі ви знайдете інформацію щодо текстових слотворень. Ці слотворення в тексті можуть говорити про МОЖЛИВІ маніпуляції в тексті. Слотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв		23
Інтервали		0
Мікропробіли		0
Білі знаки		0
Парафрази (SmartMarks)		9

#### Подібності за списком джерел

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Копір тексту означає в якому джерелі він був знайдений. Ці джерела і значення коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

#### 10 найдовших фраз

ПОРЯДКОВИЙ НОМЕР	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	Копір тексту
1	ML методи в задачі класифікації безпілотних літальних апаратів 12/21/2024 National University "Zaporizhzhia Polytechnic" (Кафедра "Системний аналіз та обчислювальна математика")	20 0.28 %
2	<a href="https://nlp.stanford.edu/~manning/courses/ling289/logistic.pdf">https://nlp.stanford.edu/~manning/courses/ling289/logistic.pdf</a>	16 0.22 %
3	<a href="https://www.raijvgopinath.com/blogs/statistics-and-data-science-hub/demystifying-shar-making-machine-learning-models-explainable-and-trustworthy">https://www.raijvgopinath.com/blogs/statistics-and-data-science-hub/demystifying-shar-making-machine-learning-models-explainable-and-trustworthy</a>	15 0.21 %

4	ML методи в задачі класифікації безпілотних літальних апаратів 12/21/2024 National University "Zaporizhzhia Polytechnic" (Кафедра "Системний аналіз та обчислювальна математика")	14 0.20 %
5	<a href="https://medium.com/@abirajna/step-by-step-guide-credit-card-fraud-detection-for-beginners-using-xgboost-22bc1a51d9c2">https://medium.com/@abirajna/step-by-step-guide-credit-card-fraud-detection-for-beginners-using-xgboost-22bc1a51d9c2</a>	13 0.18 %
6	<a href="https://medium.com/@abirajna/step-by-step-guide-credit-card-fraud-detection-for-beginners-using-xgboost-22bc1a51d9c2">https://medium.com/@abirajna/step-by-step-guide-credit-card-fraud-detection-for-beginners-using-xgboost-22bc1a51d9c2</a>	13 0.18 %
7	«Кредитний скоринг та оцінка ризиків за допомогою алгоритмів класифікації 12/24/2024 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute course papers (ФММ, К-ра економічної кібернетики)	12 0.17 %
8	<a href="https://repository.tdmu.edu.ua/bitstream/handle/123456789/11015/%D0%9C%D0%90%A2%D0%95%D0%A0%D0%86%D0%90%D0%9B%D0%98%20%D0%9A%D0%9E%D0%9D%D0%93%D0%A0%D0%95%D0%A1%D0%A3.pdf?sequence=1&amp;isAllowed=y">https://repository.tdmu.edu.ua/bitstream/handle/123456789/11015/%D0%9C%D0%90%A2%D0%95%D0%A0%D0%86%D0%90%D0%9B%D0%98%20%D0%9A%D0%9E%D0%9D%D0%93%D0%A0%D0%95%D0%A1%D0%A3.pdf?sequence=1&amp;isAllowed=y</a>	12 0.17 %
9	<a href="https://blog.onnasoft.us/es/metricas-de-evaluacion-para-modelos-de-clasificacion-con-valores-discretos/">https://blog.onnasoft.us/es/metricas-de-evaluacion-para-modelos-de-clasificacion-con-valores-discretos/</a>	11 0.15 %
10	Застосування нейронних мереж для прогнозування динаміки ринку медичних препаратів 5/13/2025 Kharkiv National University of Economics named after S.Kuznets (KNUE) (KNUE)	10 0.14 %

### з бази даних RefBooks (0.00 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
------------------	-----------	--

### з домашньої бази даних (0.00 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
------------------	-----------	--

### з програми обміну базами даних (1.32 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	ML методи в задачі класифікації безпілотних літальних апаратів 12/21/2024 National University "Zaporizhzhia Polytechnic" (Кафедра "Системний аналіз та обчислювальна математика")	34 (2) 0.48 %
2	Застосування нейронних мереж для прогнозування динаміки ринку медичних препаратів 5/13/2025 Kharkiv National University of Economics named after S.Kuznets (KNUE) (KNUE)	15 (2) 0.21 %
3	Курсова (Штучний інтелект) Гапанова І. С. ПРМ-21 11/29/2024 Lutsk National Technical University course papers (Lutsk National Technical University course papers)	14 (2) 0.20 %
4	«Кредитний скоринг та оцінка ризиків за допомогою алгоритмів класифікації 12/24/2024 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute course papers (ФММ, К-ра економічної кібернетики)	12 (1) 0.17 %
5	Система рекомендацій та прогнозування вартості нерухомого майна 6/4/2024 Nizhyn Mykola Gogol State University (Навчально-науковий інститут природничо-математичних, медико-біологічних наук та інформаційних технологій)	9 (1) 0.13 %

6	Методи штучного інтелекту для виявлення та запобігання шахрайству в банківському секторі 3/15/2025 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute (National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute)	5 (1) 0.07 %
7	Система прогнозування для підтримки промислової діяльності, з використанням машинного навчання 3/16/2025 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute (National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute)	5 (1) 0.07 %

**з Інтернету (1.66 %)**

ПОРЯДКОВИЙ НОМЕР	ДЖЕРЕЛО URL	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	<a href="https://medium.com/@abirajja/step-by-step-guide-credit-card-fraud-detection-for-beginners-using-xgboost-22bc1a51d9c2">https://medium.com/@abirajja/step-by-step-guide-credit-card-fraud-detection-for-beginners-using-xgboost-22bc1a51d9c2</a>	42 (4) 0.59 %
2	<a href="https://www.rajivgoonath.com/blogs/statistics-and-data-science-hub/demystifying-shao-making-machine-learning-models-explainable-and-trustworthy">https://www.rajivgoonath.com/blogs/statistics-and-data-science-hub/demystifying-shao-making-machine-learning-models-explainable-and-trustworthy</a>	23 (2) 0.32 %
3	<a href="https://nlp.stanford.edu/~manning/courses/ling289/logistic.pdf">https://nlp.stanford.edu/~manning/courses/ling289/logistic.pdf</a>	16 (1) 0.22 %
4	<a href="https://repository.tmu.edu.ua/bitstream/handle/123456789/11015/%D0%9C%D0%90%D0%A2%D0%95%D0%A0%D0%86%D0%90%D0%9B%D0%98%20%D0%9A%D0%9E%D0%9D%D0%93%D0%A0%D0%95%D0%A1%D0%A3.pdf?sequence=1&amp;isAllowed=y">https://repository.tmu.edu.ua/bitstream/handle/123456789/11015/%D0%9C%D0%90%D0%A2%D0%95%D0%A0%D0%86%D0%90%D0%9B%D0%98%20%D0%9A%D0%9E%D0%9D%D0%93%D0%A0%D0%95%D0%A1%D0%A3.pdf?sequence=1&amp;isAllowed=y</a>	12 (1) 0.17 %
5	<a href="https://blog.onnasoft.us/es/metricas-de-evaluacion-para-modelos-de-clasificacion-con-valores-discretos/">https://blog.onnasoft.us/es/metricas-de-evaluacion-para-modelos-de-clasificacion-con-valores-discretos/</a>	11 (1) 0.15 %
6	<a href="https://prasanmally.medium.com/logistic-regression-unlocked-practical-guide-with-data-cleaning-feature-engineering-and-more-bd4b725d4f3f">https://prasanmally.medium.com/logistic-regression-unlocked-practical-guide-with-data-cleaning-feature-engineering-and-more-bd4b725d4f3f</a>	8 (1) 0.11 %
7	<a href="https://medium.com/@bhoonika_k_m/logistic-regression-and-regularization-techniques-a-practical-guide-wilf-cyfhon-9e93a472b038">https://medium.com/@bhoonika_k_m/logistic-regression-and-regularization-techniques-a-practical-guide-wilf-cyfhon-9e93a472b038</a>	6 (1) 0.08 %

**Список прийнятих фрагментів (немає прийнятих фрагментів)**

ПОРЯДКОВИЙ НОМЕР	ЗМІСТ	КІЛЬКІСТЬ ОДНОКОВИХ СЛІВ (ФРАГМЕНТІВ)
------------------	-------	---------------------------------------

1

**ВСТУП**

Машинне навчання (ML) стало рушійною силою в сучасній медицині, кардинально змінюючи підходи до прогнозування, діагностики та планування лікування. Зростання обсягів медичних даних у поєднанні з розвитком обчислювальних можливостей створило унікальні можливості для розробки складних моделей прогнозування, які допомагають лікарям ухвалювати точніші та своєчасніші рішення щодо догляду за пацієнтами.

Значення машинного навчання у прогнозуванні захворювань важко переоцінити. Традиційна діагностика в медицині часто базується на симптомах, лабораторних тестах і досвіді лікаря. Однак такий підхід може пропустити тоні закономірності або ранні ознаки, які є критичними для своєчасного втручання. Алгоритми машинного навчання вирізняються здатністю аналізувати великі обсяги даних, виявляючи взаємозв'язки та фактори ризику, які не завжди очевидні за допомогою традиційних методів.

Особливу увагу привертає застосування ML у діагностиці діабету. Це складний метаболічний розлад, що залежить від багатьох чинників, зокрема генетики, способу життя, екології та супутніх захворювань. За оцінками ВООЗ, понад 422 мільйони людей у світі страждають на діабет, і ця цифра постійно зростає. Зважаючи на серйозні ускладнення, які виникають через пізню діагностику, раннє прогнозування стає вкрай важливим для ефективного лікування та покращення здоров'я пацієнтів.

Моделі машинного навчання для прогнозування діабету враховують

ДОДАТОК Б  
Слайди презентації

**Дослідження моделей**  
**прогнозування захворювань**  
**для раннього виявлення**  
**ризиків та покращення**  
**медичної діагностики**

**Виконала:** Потьомкіна  
Катерина Олексіївна,  
група ІПЗм-23-4

**Науковий керівник:** д.т.н.  
проф., Кирило Сергійович  
Смеляков

## Мета роботи

Дослідження та порівняння моделей машинного навчання для прогнозування ризику діабету з метою визначення оптимального алгоритмічного підходу для застосування в клінічній практиці.



# Актуальність проблеми

## Епідеміологічні показники:

- Понад 537 мільйонів дорослих із діабетом у світі (2021 р.)
- Прогноз зростання до 783 мільйонів до 2045 року
- Діабет є 7-ою провідною причиною смертності у світі

## Ситуація в Україні:

- Понад 1.3 мільйона зареєстрованих хворих (2022 р.)
- Реальна кількість може перевищувати 2 мільйони осіб
- До 50% випадків діабету 2 типу залишаються недіагностованими

## Медико-економічний тягар:

- Ускладнення діабету: ретинопатія, нефропатія, нейропатія
- Збільшення смертності у 2-3 рази
- Скорочення тривалості життя на 8-10 років
- Значні витрати системи охорони здоров'я

# Постановка задачі



## ● Проблема:

Зростає кількість випадків діагнозу діабету — важливо впроваджувати машинне навчання для раннього прогнозування та профілактики.

## ● Мета дослідження:

Дослідити ефективність моделей ML для прогнозування діабету, побудувати 3 моделі та порівняти їх точність.

## ● Задачі:

Побудувати логістичну регресію, дерево рішень і XGBoost  
Навчити моделі на медичних даних  
Оцінити якість прогнозування (точність, AUC, чутливість, специфічність)  
Визначити найбільш ефективну модель

## ● Очікуваний результат:

Визначення найбільш ефективної моделі для медичної практики прогнозування діабету.

# Методологія

- ◆ **Моделі:**

- Логістична регресія
- Дерево рішень
- XGBoost

- ◆ **Набір даних:**

- ◆ **"Diabetes 130-US Hospitals (1999–2008)"**

- 101 766 госпіталізацій з діагнозом діабету
- Витяг із бази Health Facts (Cerner, США)
- 55 характеристик: вік, стать, расові дані, аналізи, діагнози, медикаменти тощо
- Відібрані лише повноцінні звернення (1–14 днів, без летального кінця)



# Методологія

- ◆ **Попередня обробка:**

- Видалено ознаки з великим % пропусків (наприклад, вага — 97%)
- Враховано тільки перше звернення кожного пацієнта
- Видалено дані про летальні випадки та хоспіс
- Результат закодовано як бінарну змінну

- ◆ **Підготовка до навчання:**

- One-hot кодування категоріальних змінних
- Масштабування числових ознак
- Розподіл на тренувальний/тестовий набір (80/20)



# Логістична регресія (Logistic Regression)

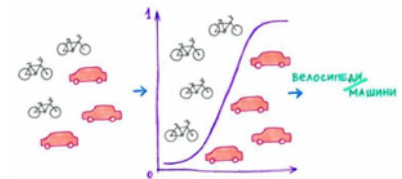
## ✦ Суть моделі:

- Це проста модель для **класифікації**, яка оцінює ймовірність того, що подія відбудеться (наприклад, **повторна госпіталізація** — так або ні).
- Замість прямої лінії, як у звичайній регресії, логістична регресія використовує **S-подібну (сигмоїдну) криву**, щоб перетворити будь-яке значення у **діапазон [0, 1]**.
- Значення  $>0.5 \rightarrow$  клас 1,  $\leq 0.5 \rightarrow$  клас 0.

## ✓ Переваги:

- Швидка у виконанні.
- Добре працює при **лінійній залежності** між ознаками та результатом

## ЛОГІСТИЧНА РЕГРЕСІЯ



# Дерево рішень (Decision Tree)

## ✦ Суть моделі:

- Модель **імітує процес ухвалення рішень**, як людина.
- Розбиває дані на підгрупи, ставлячи питання ("чи більше за", "чи рівне", і т.д.).
- Рішення приймаються шляхом **спуску від кореня до листка** дерева.

## ✓ Переваги:

- **Інтуїтивно зрозуміле** (візуально легко пояснити).
- Працює з **як числовими, так і категоріальними** даними.
- Не потребує масштабування даних.

## Decision Tree





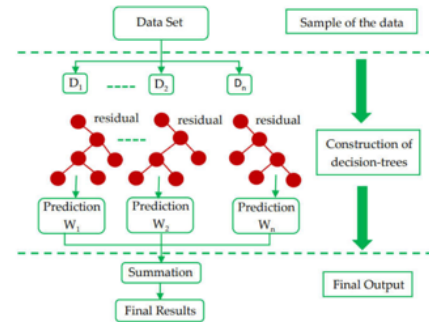
# XGBoost (Extreme Gradient Boosting)

## ✦ Суть моделі:

- Це **потужна ансамблева модель**, яка створює **кілька дерев**, де кожне наступне дерево **виправляє помилки попередніх**.
- Заснована на **градієнтному бустингу** — підхід, що комбінює багато **слабких моделей у сильну**.

## ✓ Переваги:

- Дуже **висока точність** на практиці.
- Добре працює з **нелінійними залежностями**.
- Має **вбудований механізм боротьби з перенавчанням**.



## Опис програмного забезпечення, що було використано у дослідженні

✦ Розробка моделі виконувалась за допомогою мови програмування Python — через її зручність і популярність у задачах машинного навчання.

✦ Для побудови моделей використано бібліотеки:

- scikit-learn — для логістичної регресії та дерева рішень,
- XGBoost — для побудови моделі градієнтного бустингу.

✦ Візуалізація результатів здійснювалась із застосуванням бібліотеки Matplotlib.

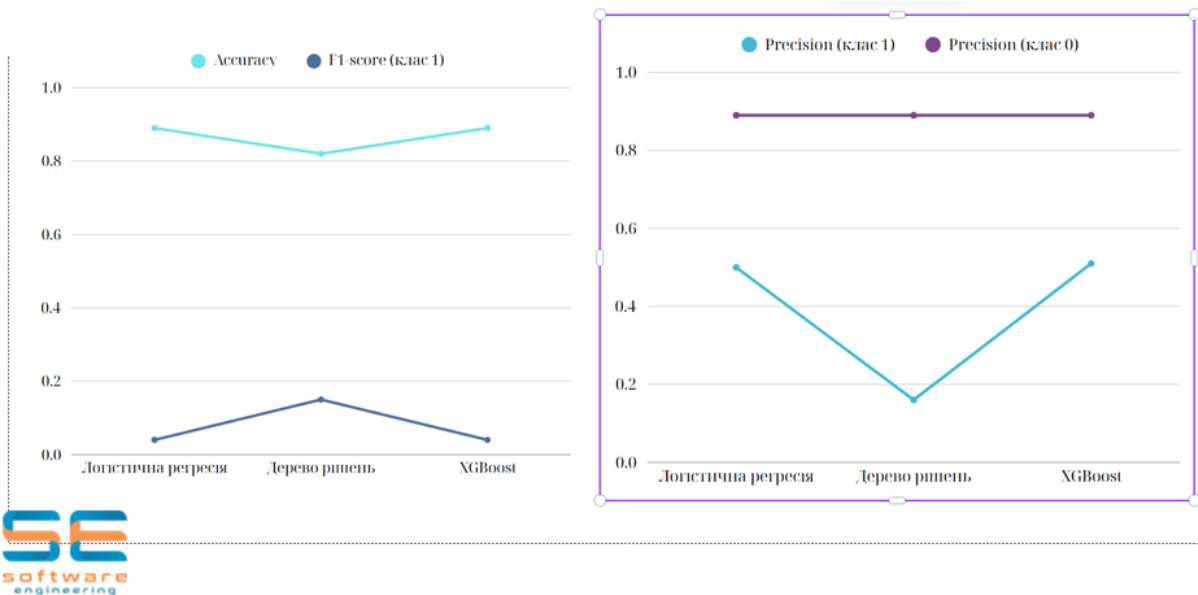
✦ Процес включав етапи: підготовка даних, тренування моделей, оцінка якості та порівняння результатів.



# Зміст проведеного експерименту

- ✦ Методи: Логістична регресія, дерево рішень, XGBoost.
- ✦ Вхідні дані: Набір медичних даних пацієнтів із позначками повторної госпіталізації.
- ✦ Критерії оцінки: Accuracy, Precision, Recall, F1-score, ROC AUC.
- ✦ Послідовність:
  - Розбиття даних на тренувальну і тестову вибірки;
  - Навчання моделей;
  - Оцінка якості на тестових даних;
  - Побудова ROC-кривих і таблиць результатів.
- ✦ Вимірювання: Відстеження метрик точності і здатності моделей відрізнити позитивний клас.

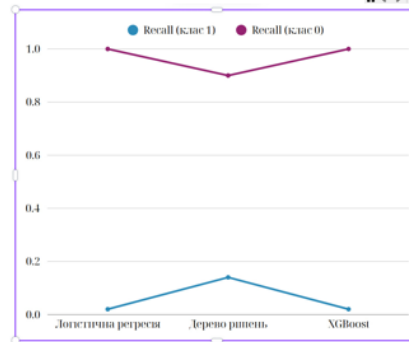
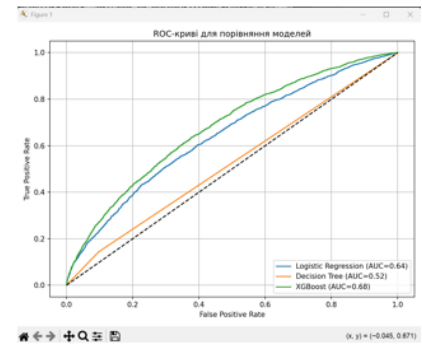
# Результати експерименту



# Результати експерименту

ROC-криві візуалізують здатність моделей розрізняти класи при різних порогах класифікації.

Діагональна лінія – випадкове прогнозування (AUC = 0.5), а ідеальний класифікатор має AUC = 1.0.



# Результати експерименту

✦ Логістична регресія і XGBoost мають високу загальну точність (89%). XGBoost має найкращий ROC AUC (0.68), тобто найкращу здатність ранжувати ризики.

✦ Дерево рішень має нижчі показники, але кращий баланс між точністю та повнотою.

✦ Моделі демонструють низьку чутливість (Recall) до позитивного класу, що є проблемою для практичного застосування.

✦ Результати вказують на необхідність подальшої оптимізації моделей.

# Аналіз отриманих результатів

- ✓ Моделі відповідають меті дослідження — оцінити можливість прогнозування повторної госпіталізації.
- ✓ XGBoost показав найкращі результати за ROC AUC, підтверджуючи ефективність ансамблевих методів.
- ✓ Низька чутливість сигналізує про потребу у доопрацюванні або балансуванні даних.
- ✓ Отримані результати узгоджуються з відомими особливостями моделей машинного навчання у медичних задачах.

## Публікація результатів





## Decision Trees for Cervical Cancer Prediction

Karol Szlachetka\* and Ekaterina Potomkina\*  
 \* KRAMSK, Study Number 14, 81206, Ukraine

**Abstract**  
 This study evaluates the impact of including human papillomavirus (HPV) diagnosis information on the effectiveness of decision tree models for cervical cancer prediction. Using the "UCI Cervical Cancer Risk Factors" dataset, we compare two models: one without considering HPV diagnosis and another including this feature. Results demonstrate that the model incorporating HPV status performs better across all metrics, particularly in sensitivity (increase from 0.64 to 0.73) and area under the ROC curve (increase from 0.80 to 0.85). These findings confirm the importance of considering HPV-associated information when developing risk assessment and screening tools for cervical cancer.

**Keywords**  
 Cervical cancer, HPV, machine learning, decision tree, risk prediction

### 1. Introduction

Cervical cancer remains a serious threat to women's health worldwide, ranking fourth among the most common oncological diseases among women globally [1]. This problem is particularly acute in low- and middle-income countries due to limited access to effective prevention and screening programs [2].

Persistent infection caused by high-risk types of human papillomavirus (HPV) has been proven to be the primary cause of cervical cancer development [3]. In particular, HPV genotypes 16 and 18 are responsible for a significant number of disease cases [4]. Recognition of HPV's key role has led to the development of vaccines targeting these genotypes, which have demonstrated significant potential in preventing infection and, consequently, reducing the risk of malignant transformation [5].

Early detection of cervical cancer through screening methods such as cytological examination (Pap test) and HPV testing is critical for improving treatment outcomes and patient survival [6]. However, the effectiveness and coverage of these programs may be limited due to various factors, including healthcare accessibility, women's adherence to screening, and the sensitivity of the diagnostic methods themselves [7].

In this context, machine learning methods, particularly decision tree models, open new perspectives for developing supportive prognostic tools in healthcare, especially in cervical cancer risk assessment. One of the key advantages of decision tree models is their interpretability, which allows healthcare providers to understand better the factors influencing prognosis [8].

This study aims to evaluate the effectiveness of decision tree models in predicting cervical cancer risk, with particular emphasis on the impact of including information about HPV diagnosis on model accuracy and performance. Previous studies have already demonstrated the successful application of various machine learning algorithms, including decision trees, for cervical cancer prediction based on diverse clinical and demographic data [9-11].

Key predictors used in such models included age, sexual behavior, contraceptive use, and HPV status [12]. Numerous epidemiological studies have also confirmed the important role of HPV in cervical cancer development [13]. However, the question of how significant the inclusion of direct information about HPV diagnosis in decision tree models is for improving predictive ability remains important for further study.

In this study, we use the publicly available "UCI Cervical Cancer Risk Factors" dataset [14] to build and compare two decision tree models: one without considering HPV diagnosis and another including

HPV status. The goal is to quantitatively assess the impact of HPV information on model performance and substantiate its potential value for clinical practice.

The results presented in Table 1 demonstrate that including the "Dn" feature in the decision tree model improves all considered performance metrics. In particular, there is a significant increase in model sensitivity (from 0.64 to 0.73), which is critically important for medical applications as it reflects the model's ability to identify patients with cervical cancer correctly. There is also a slight improvement in the model's accuracy and specificity and an increase in AUC value, indicating a better overall ability of the model to distinguish between classes.

ROC curves were constructed to visualize the classification ability of both models (Figure 1). The area under the ROC curve (AUC) for the model with Dn is higher (0.85) compared to the model without Dn (0.80), confirming the better discriminatory capability of the model that takes HPV status into account.

**3. Conclusion**  
 The results of this study demonstrate that including information about HPV diagnosis significantly improves the performance of decision tree models in predicting cervical cancer. The model that takes HPV status into account showed higher sensitivity, meaning better ability to correctly identify women with cervical cancer, which is especially important in the clinical context of early diagnosis. Improvements in other metrics, such as accuracy, specificity, and AUC, also confirm the advantage of the model with the included "Dn" feature.

The obtained results are consistent with existing knowledge about the key role of HPV in cervical cancer development and emphasize the importance of considering HPV-associated information when developing risk assessment and screening tools. Even though the dataset used did not contain direct information about HPV vaccination status, the results confirm the significance of information about the presence of HPV infection for improving predictive models.

Further research could focus on using more detailed data about HPV, including HPV types and vaccination history, as well as exploring the application of other machine learning algorithms for cervical cancer prediction. Considering a broader spectrum of risk factors, including socioeconomic determinants and lifestyle, may also contribute to creating more accurate and comprehensive predictive models.

Overall, this study highlights the potential of using machine learning models, particularly decision trees, to improve screening strategies and individualized risk assessment for cervical cancer, especially when considering information about HPV status.

**4. References**

- [1] Global cancer statistics, World Cancer Research Fund. Available at: <https://www.wcrf.org/preventing-cancer/cancer-statistics/cervical-cancer/>
- [2] UNICEF. The consequences of HPV and cervical cancer in developing nations. *Communications of the Global Health* 48 (2023) 36-44. doi:10.1145/2348913.2348913.
- [3] WHO Study Group. Understanding HPV: pathogenesis and prevention strategies. *J. World Health Organization* 54 (2022). doi:10.1145/2219992.2219993.
- [4] M. Rodriguez (Ed.), Special issue: Global Cervical Cancer Elimination, volume 39, 2022.
- [5] D. Harper. *Understanding HPV and Cervical Screening*, 2nd ed., Oxford University Press, Oxford, 2021.
- [6] K. Mawor. *Cervical Cancer Prevention*, volume 68 of *Advances in Oncology*, Springer-Verlag, New York, NY, 2022. doi:10.1007/978-1-4939-237-4.
- [7] J. Smith (Ed.), *Comprehensive Guide to Gynecologic Oncology*, volume 9 of *Cancer Prevention and Treatment*, 1st ed., University of Chicago Press, Chicago, 2022. doi:10.1007/978-0-495-9217-4.
- [8] S. Johnson (Ed.), *Advanced Screening Technologies for HPV*, American Cancer Society Press, Atlanta, 2021. doi:10.1007/978-0-495-9217-4.

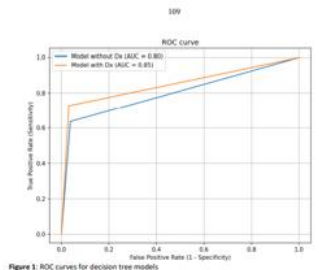


Figure 1. ROC curves for decision tree models

Further research could focus on using more detailed data about HPV, including HPV types and vaccination history, as well as exploring the application of other machine learning algorithms for cervical cancer prediction. Considering a broader spectrum of risk factors, including socioeconomic determinants and lifestyle, may also contribute to creating more accurate and comprehensive predictive models.

Overall, this study highlights the potential of using machine learning models, particularly decision trees, to improve screening strategies and individualized risk assessment for cervical cancer, especially when considering information about HPV status.

**4. References**

- [1] Global cancer statistics, World Cancer Research Fund. Available at: <https://www.wcrf.org/preventing-cancer/cancer-statistics/cervical-cancer/>
- [2] UNICEF. The consequences of HPV and cervical cancer in developing nations. *Communications of the Global Health* 48 (2023) 36-44. doi:10.1145/2348913.2348913.
- [3] WHO Study Group. Understanding HPV: pathogenesis and prevention strategies. *J. World Health Organization* 54 (2022). doi:10.1145/2219992.2219993.
- [4] M. Rodriguez (Ed.), Special issue: Global Cervical Cancer Elimination, volume 39, 2022.
- [5] D. Harper. *Understanding HPV and Cervical Screening*, 2nd ed., Oxford University Press, Oxford, 2021.
- [6] K. Mawor. *Cervical Cancer Prevention*, volume 68 of *Advances in Oncology*, Springer-Verlag, New York, NY, 2022. doi:10.1007/978-1-4939-237-4.
- [7] J. Smith (Ed.), *Comprehensive Guide to Gynecologic Oncology*, volume 9 of *Cancer Prevention and Treatment*, 1st ed., University of Chicago Press, Chicago, 2022. doi:10.1007/978-0-495-9217-4.
- [8] S. Johnson (Ed.), *Advanced Screening Technologies for HPV*, American Cancer Society Press, Atlanta, 2021. doi:10.1007/978-0-495-9217-4.



## Підсумки

- ✦ Отримані результати є реалістичними та демонструють потенціал використання ML-моделей для прогнозування ризику повторної госпіталізації.
- ✦ Модель XGBoost виявилась найбільш перспективною, але загальна низька чутливість вказує на необхідність покращень.
- ✦ Подальший розвиток: використання додаткових ознак, балансування класів, тонке налаштування моделей.
- ✦ Важливо продовжувати дослідження для підвищення клінічної значущості моделей та впровадження їх у медичну практику.

Дякую за увагу!

ДОДАТОК В

Апробація результатів роботи



# ДОДАТОК Г

## Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент  
(посада)

програмної інженерії  
(кафедра)

ІПЗм-23-4  
(група)

Потьомкіна Катерина Олексіївна

(прізвище, ім'я, по батькові)

### Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	<b>7.1 Загальні положення</b>	
	<b>7.3 Нумерація сторінок звіту</b>	
	<b>7.4 Нумерація розділів, підрозділів, пунктів, підпунктів</b>	
	<b>7.5 Рисунки</b>	
	<b>7.6 Таблиці</b>	
	<b>7.7 Переліки</b>	
	<b>7.8 Примітки</b>	
	<b>7.9 Виноски</b>	
	<b>7.10 Формули та рівняння</b>	
	<b>7.11 Посилання</b>	
	<b>7.13 Список авторів</b>	
	<b>7.14 Скорочення та умовні позначки</b>	
	<b>7.15 Додатки</b>	

зауважень немає

Експерт

\_\_\_\_\_  
(підпис)

Олена ОЛІЙНИК

(прізвище, ініціали)

18.06.2025