

Міністерство освіти і науки України  
Національний технічний університет  
«Дніпровська політехніка»

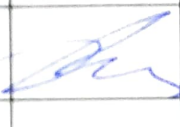
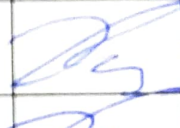
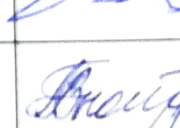


Факультет інформаційних технологій  
(факультет)

Кафедра системного аналізу та управління  
(повна назва)

**ПОЯСНЮВАЛЬНА ЗАПИСКА**  
Кваліфікаційної роботи ступеня магістра  
(назва освітньо-кваліфікаційного рівня)

До захисту  
18.01.2022

Студента Симонця Іллі Юрійовича  
академічної групи 124М-20-1  
спеціальності 124 Системний аналіз  
на тему: «Прогнозування касових зборів фільмів із залученням методів машинного навчання»

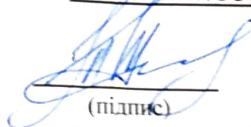
Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	Інституційною	
кваліфікаційної роботи	Д.ф.-м.н., проф. Купенко О.П.	95	Відрінок	
розділів:				
Інформаційно- аналітичний	Д.ф.-м.н., проф. Купенко О.П.	95	Відрінок	
Спеціальний	Д.ф.-м.н., проф. Купенко О.П.	95	Відрінок	
Рецензент	Д.т.н., проф. Гнатушенко В.В	95	Відрінок	
Нормоконтроль	к.ф.-м.н., доц. Хом'як Т.В	90	Відрінок	

fo 80

ЗАТВЕРДЖЕНО:

завідувач кафедри

Системного аналізу та управління  
(повна назва)

  
(підпис)

к.т.н., доц. Т.А.Желдак  
(прізвище, ініціали)

« 10 » 12. 2021 року

**ЗАВДАННЯ**  
на кваліфікаційну роботу  
ступеня магістра

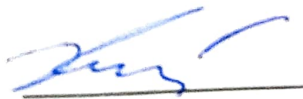
студенту Симонцю І. Ю. академічної групи 124М-20-1  
спеціальності: 124 Системний аналіз  
на тему «Прогнозування касових зборів фільмів із залученням методів  
машинного навчання»

затверджену наказом ректора НТУ «Дніпровська політехніка»

від 10.12.2021 р. №1036 – С

Розділ	Зміст	Терміни виконання
1. Інформаційно-аналітичний розділ	<i>Аналіз методів, алгоритмів та інструментів для прогнозування та попередньої обробки даних.</i>	10.09.2021 – 01.11.2021
2. Спеціальний розділ	<i>Розробка програмних модулів для попередньої обробки, конструювання ознак та прогнозування касових зборів фільмів. Аналіз результатів.</i>	01.11.2021 – 30.12.2021

Завдання видано

  
(підпис)

проф. Купенко О.П.  
(прізвище, ініціали)

Дата видачі: 06.09.2021 р.

Дата подання до екзаменаційної комісії:

12 01 2022

Прийнято до виконання

  
(підпис студента)

Симонець І.Ю.  
(прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка: 74 с., 37 рисунків, 2 таблиці, 4 додатки, 10 джерел.

*Об'єктом дослідження* є дані та метадані про фільми.

*Предметом дослідження* є методи машинного навчання та штучного інтелекту для прогнозування та пошуку цінних знань у структурованих даних.

*Метою дослідження* є розробка методики прогнозування касових зборів фільмів на основі методів штучного інтелекту і машинного навчання.

*Методи дослідження:* в роботі проведено збір даних та метаданих про фільми та виділення факторів, що впливають на касові збори фільмів, для подальшого їх використання з метою прогнозування касових зборів фільмів. Виділення факторів зроблено за допомогою методів попередньої обробки даних та машинного навчання.

В *інформаційно-аналітичному розділі* розглянуто методи, алгоритми та інструменти для прогнозування та попередньої обробки даних.

В *спеціальному розділі* розроблено модулі для попередньої обробки, конструювання ознак та прогнозування касових зборів фільмів.

*Практична цінність* отриманих у роботі результатів полягає у розробці програмного забезпечення, за допомогою якого можна швидко в автоматичному режимі здійснювати прогнозування касових зборів фільмів, на основі даних та метаданих про фільми. Розроблена методика аналізу дозволить виявити взаємозв'язок між параметрами що впливають на касові збори, на основі отриманої інформації зробити прогноз касового збору фільму.

*Ключові слова:* РЕГРЕСІЙНИЙ АНАЛІЗ, ДЕРЕВА РІШЕНЬ, ТРАНСФОРМЕРИ РЕЧЕНЬ, ЛІНІЙНА РЕГРЕСІЯ, МЕТОД k-НАЙБЛИЖЧИХ СУСІДІВ, XGBOOST, КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ, НЕЙРОНІ МЕРЕЖІ, JSON, CSV.

## ABSTRACT

Explanatory note: 74 p., 37 drawings, 2 tables, 4 appendices, 10 sources.

The object of research is data and metadata about films.

The subject of research is the methods of machine learning and artificial intelligence for predicting and searching for valuable knowledge in structured data.

The aim of the study is to develop a methodology for predicting the revenue of films based on methods of artificial intelligence and machine learning.

Research methods: the collection of data and metadata about films and the selection of factors influencing the revenue of films, for their further use to predict the films revenue. The selection of factors is made using the methods of data pre-processing and machine learning. In the information-analytical section methods, algorithms and tools for forecasting and pre-processing of data are considered.

In the information-analytical section methods, algorithms and tools for forecasting and pre-processing of data are considered.

In a special section, modules have been developed for pre-processing, feature design and forecasting films' revenue.

The practical value of the results obtained in the work lies in the development of software that can quickly and automatically predict the cash collection of films, based on data and metadata about the films. The developed method of analysis will allow identifying the relationship between the parameters that affect the revenue, based on the information obtained to make a forecast of the films revenue.

Keywords: REGRESSION ANALYSIS, DECISION TREES, SENTENCE TRANSFORMERS, LINEAR REGRESSION, k-NEAREST NEIGHBORS METHOD, XGBOOST, DETERMINATION COEFFICIENT, NEURAL NETWORKS, JSON, CSV.