МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет
«Харківський авіаційний інститут»

Факультет інтелектуальних систем управління

Кафедра комп'ютерних наук та інформаційних технологій

Пояснювальна записка до кваліфікаційної роботи

(тип кваліфікаційної роботи)

магістра

(освітній ступінь)

на тему «Прогнозування ефективності спортсменів із використанням методів машинного навчання»

ХАІ.302.366.25О.122.203078 ПЗ

Виконав: студент 6 курсу групи № 366
Спеціальність 122 «Комп'ютерні науки»
(код та найменування)
Освітня програма Комп'ютеризація обробки інформації та управління
Радченко Є. П.
(прізвище й ініціали студента)
Керівник: доц. каф. 302, к.т.н.
Лещенко Ю. О.
(прізвище й ініціали)
Рецензент: Плєхова Г.А.
(прізвище й ініціали)

Харків – 2025

Міністерство освіти і науки України
Національний аерокосмічний університет
«Харківський авіаційний інститут»

Факультет інтелектуальних систем управління
Кафедра комп'ютерних наук та інформаційних технологій
Рівень вищої освіти другий (магістерський)
Галузь знань 12 «Інформаційні технології»
(код та найменування)
Спеціальність 122 «Комп'ютерні науки»
Освітня програма Комп'ютеризація обробки інформації та управління

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Олег ФЕДОРОВИЧ
(підпис) (ім'я та прізвище)
«__» _листопада_ 20_25_ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

Радченко Єгору Петровичу
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи __Прогнозування ефективності спортсменів із використанням методів машинного навчання

керівник кваліфікаційної роботи _Лещенко Ю.О., к.т.н., доц. каф. 302,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом Університету №_1765-уч. від «_14» листопада 20_25_ року

2. Термін подання кваліфікаційної роботи ___08.12.2025

3. Вихідні дані до проекту (роботи)__ датасет NHL Draft Hockey Player Data (1963–2022)

4. Зміст пояснювальної записки (перелік завдань, які потрібно розв'язати)_ проаналізувати існуючі підходи до оцінювання ефективності спортсменів та методи прогнозування результативності; розробити набір показників ефективності спортсменів на основі статистичних змінних; побудувати та порівняти моделі машинного навчання для прогнозування ефективності; провести кластеризацію спортсменів для виявлення типових профілів результативності; оцінити точність моделей і визначити найінформативніші ознаки, що впливають на прогноз; зробити висновки та рекомендації щодо практичного застосування результатів у спортивній аналітиці

5. Перелік графічного матеріалу _завдання оцінювання та прогнозування ефективності спортсменів; мета та завдання; поняття ефективності спортсменів

у спортивній аналітиці; класичні методи оцінювання результативності в індивідуальних і командних видах спорту; основи прогнозування в спортивній статистиці; методи машинного навчання, застосовні для спортивних даних; відкриті набори даних і платформи для досліджень; характеристика датасету NHL Draft Hockey Player Data; дослідницький аналіз даних NHL Draft Hockey Player Data; вибір показників ефективності для прогнозування; аналіз кореляцій; попередня обробка даних; побудова моделей прогнозування; оцінювання якості моделей; аналіз важливості ознак; обґрунтування потреби кластеризації; вибір ознак для сегментації спортсменів; застосування методів кластеризації; аналіз отриманих кластерів; зв'язок між кластерами та прогнозними результатами моделей_

6. Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада Консультанта	Підпис, дата	
		завдання видав	завдання прийняв
з основної частини	Лещенко Ю.О. доц. каф. 302		
з спеціального розділу	Лещенко Ю.О. доц. каф. 302		
з економіки та управління	Малєєва Ю. А. доц. каф. 302		

Нормоконтроль _____ Олександр ПРОХОРОВ «___» _____20___ р.
(підпис) (ім'я та прізвище)

7. Дата видачі завдання «_06__» _вересня__2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	ТЕОРЕТИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ СПОРТСМЕНІВ	06.09 – 20.09	
2	АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПІДГОТОВКА ДАНИХ	21.09 – 28.09	
3	ПОБУДОВИ МОДЕЛЕЙ ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ СПОРТСМЕНІВ	29.09 – 20.10	
4	КЛАСТЕРИЗАЦІЯ СПОРТСМЕНІВ ЗА СТИЛЕМ ГРИ ТА ЕФЕКТИВНІСТЮ	21.10 – 20.11	
5	ЕКОНОМІЧНИЙ РОЗДІЛ	20.11 – 25.11	

Здобувач _____ __Єгор РАДЧЕНКО__
(підпис) (ім'я та прізвище)

Керівник кваліфікаційної роботи _____ Юлія ЛЕЩЕНКО
(підпис) (ім'я та прізвище)

УДК 004.8

РЕФЕРАТ

Прогнозування ефективності спортсменів із використанням методів машинного навчання – Радченко Єгор, кваліфікаційна робота магістра, Харків, Національний аерокосмічний університет «ХАІ», кількість сторінок, враховуючи додатки 97, кількість таблиць 27, кількість рисунків 32, кількість джерел літератури 15.

КЛЮЧОВІ СЛОВА: спортивна аналітика, статистика, прогнозування ефективності спортсменів, моделі машинного навчання, кластеризація.

ЦІЛЬ: розроблення та дослідження моделей машинного навчання для прогнозування ефективності спортсменів на основі історичних статистичних даних, з метою підвищення обґрунтованості прийняття рішень у спортивній аналітиці.

ОБ’ЄКТ ДОСЛІДЖЕННЯ: процес оцінювання та прогнозування ефективності спортсменів на основі статистичних показників їхньої діяльності.

ПРЕДМЕТ ДОСЛІДЖЕННЯ: моделі та методи машинного навчання, що забезпечують прогнозування результативності спортсменів за їхніми історичними статистичними даними.

ЗАДАЧІ:

1. Проаналізувати існуючі підходи до оцінювання ефективності спортсменів та методи прогнозування результативності.
2. Розробити набір показників ефективності спортсменів на основі статистичних змінних.
3. Побудувати та порівняти моделі машинного навчання для прогнозування ефективності.
4. Провести кластеризацію спортсменів для виявлення типових профілів результативності.
5. Оцінити точність моделей і визначити найінформативніші ознаки, що впливають на прогноз.
6. Зробити висновки та рекомендації щодо практичного застосування результатів у спортивній аналітиці.

МЕТОДИ ДОСЛІДЖЕННЯ: методи штучного інтелекту та машинного навчання.

ОТРИМАНІ РЕЗУЛЬТАТИ: розроблено систему аналітики для прогнозування ефективності хокеїстів на основі моделей машинного навчання.

UDC 004.8

ABSTRACT

Predicting athlete performance using machine learning methods – Yehor Radchenko, Master Thesis, Kharkiv, National Aerospace University "KhAI", amount of pages 97, amount of tables 27, amount of figures 32, amount of literature's sources 15.

KEYWORDS: sports analytics, statistics, athlete performance forecasting, machine learning models, clustering.

PURPOSE OF RESEARCH: to develop and investigate machine learning models for forecasting athlete performance based on historical statistical data, in order to improve the validity of decision-making in sports analytics.

OBJECT OF RESEARCH: the process of evaluating and forecasting athlete performance based on statistical indicators of their activity.

SUBJECT OF RESEARCH: machine learning models and methods that provide forecasting of athlete performance using their historical statistical data.

TASKS:

1. To analyze existing approaches to evaluating athlete performance and methods of performance forecasting.
2. To develop a set of athlete performance indicators based on statistical variables.
3. To build and compare machine learning models for performance forecasting.
4. To perform clustering of athletes to identify typical performance profiles.
5. To evaluate the accuracy of the models and determine the most informative features influencing the forecast.
6. To formulate conclusions and recommendations regarding the practical application of the results in sports analytics.

METHODS OF RESEARCH: methods of artificial intelligence and machine learning.

RESULTS: a sports analytics system for forecasting the performance of hockey players based on machine learning models was developed.

ЗМІСТ

ВСТУП.....7
1 ТЕОРЕТИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ СПОРТСМЕНІВ.....9
1.1 Поняття ефективності спортсменів у спортивній аналітиці..... 10
1.2 Класичні методи оцінювання результативності в індивідуальних і командних видах спорту.....14
1.3 Основи прогнозування в спортивній статистиці..... 17
1.4 Методи машинного навчання, застосовні для спортивних даних.....20
1.5 Аналіз сучасних підходів до прогнозування ефективності спортсменів..... 26
1.5.1 Сучасні наукові підходи..... 26
1.5.2 Практичні системи спортивного прогнозування..... 26
1.5.3 Відкриті набори даних і платформи для досліджень.....28
Висновки за розділом.....29
2 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПІДГОТОВКА ДАНИХ.....31
2.1 Характеристика датасету NHL Draft Hockey Player Data..... 31
2.2 Дослідницький аналіз даних та змісту ознак.....33
2.3 Вибір показників ефективності для прогнозування.....38
2.4 Аналіз кореляцій.....46
2.5 Попередня обробка даних.....49
Висновки за розділом.....50
3 ПОБУДОВИ МОДЕЛЕЙ ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ СПОРТСМЕНІВ.....51
3.1 Вибір алгоритмів машинного навчання для регресійних завдань..... 51
3.2 Побудова моделей прогнозування.....52
3.3 Оцінювання якості моделей за допомогою метрик..... 55
3.4 Аналіз важливості ознак..... 56
3.5 Інтерпретація результатів і порівняння моделей..... 58
Висновки за розділом..... 59
4 КЛАСТЕРИЗАЦІЯ СПОРТСМЕНІВ ЗА СТИЛЕМ ГРИ ТА ЕФЕКТИВНІСТЮ. 60
4.1 Обґрунтування потреби кластеризації.....60
4.2 Вибір ознак для сегментації спортсменів.....61
4.3 Застосування методів кластеризації.....62
4.4 Аналіз отриманих кластерів..... 64
4.5 Зв’язок між кластерами та прогнозними результатами моделей..... 66
4.6 Моделювання процесу ухвалення рішень на основі результатів аналітики.. 68

Висновки за розділом..... 69
5 ЕКОНОМІЧНИЙ РОЗДІЛ..... 71
5.1 Мета розділу..... 71
5.2 Визначення тривалості та розрахунок собівартості виконання проєкту (перший варіант)..... 71
5.3 Визначення тривалості та розрахунок собівартості виконання проєкту (другий варіант)..... 77
5.4 Висновки за розділом..... 82
ВИСНОВКИ..... 83
ПЕРЕЛІК ПОСИЛАНЬ..... 84
ДОДАТОК А. Графічні матеріали..... 86

ВСТУП

Актуальність теми дослідження.

У сучасному спорті прийняття рішень дедалі більше спирається на аналіз великих обсягів даних. Статистичні показники, що характеризують виступи спортсменів, команд та окремі ігрові епізоди, утворюють інформаційну базу для побудови систем спортивної аналітики. Водночас традиційні методи оцінювання ефективності спортсменів часто є суб'єктивними та не враховують складних нелінійних залежностей між різними характеристиками гри. Це зумовлює потребу у використанні сучасних інтелектуальних технологій, здатних виявляти приховані закономірності у даних і забезпечувати точні прогнози результативності.

Методи машинного навчання відкривають нові можливості для спортивної аналітики – від оцінки потенціалу молодих гравців і прогнозування їхньої кар'єрної успішності до виявлення оптимальних комбінацій у командній грі. Завдяки здатності обробляти багатовимірні дані та навчатися на прикладах, моделі машинного навчання дозволяють формувати об'єктивні, статистично обґрунтовані показники ефективності. Це сприяє підвищенню якості тренерських рішень, удосконаленню процесів скаутингу, підготовки спортсменів і управління спортивними ресурсами.

В умовах швидкого розвитку спортивних технологій та цифровізації галузі актуальним стає створення моделей, що дозволяють не лише аналізувати минулі результати, а й прогнозувати майбутню ефективність спортсменів. Такі системи мають практичне значення для клубів, федерацій, аналітичних центрів і спортивних менеджерів, адже дають змогу оптимізувати відбір, планування тренувального процесу та оцінку інвестицій у спортсменів.

Таким чином, дослідження методів машинного навчання для прогнозування ефективності спортсменів є актуальним. Воно поєднує сучасні підходи до аналізу даних із практичними завданнями спортивного менеджменту, сприяє розвитку цифрової спортивної аналітики та формує передумови для більш науково обґрунтованого управління спортивною діяльністю.

Мета дослідження – розроблення та дослідження моделей машинного навчання для прогнозування ефективності спортсменів на основі історичних статистичних даних, з метою підвищення обґрунтованості прийняття рішень у спортивній аналітиці.

Для досягнення мети необхідно виконати наступні **завдання**:

1. Проаналізувати існуючі підходи до оцінювання ефективності спортсменів та методи прогнозування результативності.

2. Розробити набір показників ефективності спортсменів на основі статистичних змінних.

3. Побудувати та порівняти моделі машинного навчання для прогнозування ефективності.

4. Провести кластеризацію спортсменів для виявлення типових профілів результативності.

5. Оцінити точність моделей і визначити найінформативніші ознаки, що впливають на прогноз.

6. Зробити висновки та рекомендації щодо практичного застосування результатів у спортивній аналітиці.

Об’єкт дослідження – процес оцінювання та прогнозування ефективності спортсменів на основі статистичних показників їхньої діяльності.

Предмет дослідження – моделі та методи машинного навчання, що забезпечують прогнозування результативності спортсменів за їхніми історичними статистичними даними.

Практичне значення отриманих результатів. Розроблено систему аналітики для прогнозування ефективності хокеїстів на основі моделей машинного навчання.

Структура та обсяг кваліфікаційної роботи магістра. Кваліфікаційна робота магістра складається зі вступу, 5 розділів, висновків, додатків та переліку посилань. Повний обсяг кваліфікаційної роботи магістра становить 97 сторінок, в тому числі: 32 рисунки; 27 таблиць; 1 додаток на 12 сторінках; перелік посилань з 15 найменувань на 2 сторінках.