

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Системотехніки
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)
(рівень вищої освіти)

Дослідження та розробка методу прогнозування параметрів статури людини для системи планування тренувань
(тема)

Виконав: здобувач групи ІТПм-20-1
Петренко А. Ю.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Освітня програма ОПП Інформаційні технології проектування
(повна назва освітньої програми)

Керівник доц. Тітов С. В.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри системотехніки _____
(підпис)

Гребеннік І. В.
(прізвище, ініціали)

2021 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
Кафедра _____ Системотехніки _____
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Освітня програма _____ ОПП Інформаційні технології проектування _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
« ____ » _____ 20 ____ р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

Студентові Петренко Антону Юрійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження та розробка методу прогнозування параметрів статури людини для системи планування тренувань
затверджена наказом по університету від 8 листопада 2021 р. № 1663 Ст
2. Термін подання студентом роботи до екзаменаційної комісії 17 грудня 2021 р.
3. Вихідні дані до роботи Функції системи: надати інформаційну поміч спортсмену у виконанні вправ, врахування кількості виконаних підходів, а також відстеження форми за допомогою врахування та прогнозування ваги атлета.
4. Перелік питань, що потрібно опрацювати в роботі 4.1 Вступ, 4.2 Аналіз предметної області, 4.2.1 Класифікація інформаційних систем та їх архітектура, 4.2.2 Аналіз даних як метод обробки інформації, 4.2.2.1 Важливість аналізу даних для бізнесу, 4.2.2.2 Процес аналізу даних, 4.2.2.3 Важливість аналізу даних у дослідженні, 4.2.2.4 Типи аналізу даних, 4.2.2.5 Техніки аналізу даних, 4.2.3 Використання прогнозування в професійному спорті, 4.3 Метод регресійного аналізу як метод передбачення, 4.3.1 Визначення регресійного аналізу, 4.3.2 Цілі регресійного аналізу, 4.3.3 Припущення регресійного аналізу, 4.3.4 Проста модель регресії, 4.3.5 Модель множинної регресії, 4.3.6 Коефіцієнт детермінації R², 4.3.7 t-критерій Стьюдента, 4.3.8 Постановка задачі, 4.3.9 Використання регресійного аналізу на олімпіадах, 4.3.10 Розробка методу прогнозування параметрів статури людини, 4.3.11 Опис результатів обчислювального експерименту, 4.4 Опис розробленого додатку, 4.4.1 Опис розробки, 4.4.2 Технічні вимоги, 4.4.3 Проектування додатку, 4.4.3.1 Розробка інтерфейсу користувача, 4.4.4 Сторінка прогнозування результатів.
5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) Слайди презентації: титул, мета роботи, опис об'єкту і предмета дослідження, постановка задачі, етапи дослідження, процес схуднення, причини схуднення, важливість фізичних навантажень, план дій, вимоги до додатку, висновки.

6. Консультанти розділів роботи


Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Основна частина	доц. Тітов С.В.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1.	Отримання завдання кваліфікаційної роботи	8.11.2021	
2.	Аналіз завдання та аналогів сервісу розумного пошуку крос-медійного контенту	9.11.2021-12.11.2021	
3.	Аналіз літератури	13.11.2021 – 14.11.2021	
4.	Дослідження впливу факторів на процес схуднення	15.11.2021-18.11.2021	
5.	Дослідження регресійного аналізу та моделі множинної регресії	19.11.2021-20.11.2021	
6.	Розробка методу прогнозування статури людини	21.11.2021-26.11.2021	
7.	Розробка програмного засобу	27.11.2021 – 10.12.2021	
8.	Проведення експериментальних досліджень	11.12.2021	
9.	Оформлення пояснювальної записки та документації	12.12.2021 – 14.12.2021	
10.	Оформлення презентаційних матеріалів	15.12.2021	
11.	Представлення на рецензування	17.12.2021	
	Представлення кваліфікаційної роботи в ДЕК	17.12.2021	

Дата видачі завдання 8 листопада 2021 р.

Студент Петренко А.Ю.
(підпис)

Керівник роботи  доцент Тітов С. В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Записка пояснювальна : 64 с., 18 рис., 5 табл., 2 дод., 25 джерел.

МОБІЛЬНИЙ ДОДАТОК, ТОЧНІСТЬ МЕТОДУ, ПРОГНОЗУВАННЯ, ТРЕНУВАННЯ, РЕГРЕСІЯ, БАГАТОФАКТОРНА РЕГРЕСІЯ, НЕЗАЛЕЖНА ЗМІННА, ЗАЛЕЖНА ЗМІННА, iOS

Об'єктом дослідження є система планування тренувального процесу.

Предмет дослідження є методи прогнозування даних, і зокрема параметрів статури людини в залежності від фізичних навантажень.

Мета роботи – дослідження та розробка методу прогнозування параметрів статури людини для системи планування тренувань.

Методи дослідження – системний аналіз, регресійний аналіз.

В результаті проведених досліджень розроблено метод для прогнозу статури людини на основі даних, зібраних протягом тренувального процесу. Прогнозування організму людини є актуальним оскільки ще не існує методу який би дозволив точно спрогнозувати людський організм хоча б у якомусь аспекті. Отриманий метод може бути використаний атлетами різних рівнів, у спортивній медицині, а також у медицині вцілому для представлення лікарю більш повної картини тренувального процесу пацієнта. Додаток було розроблено на мові програмування Swift і використано бібліотеку UIKit, а також сховище даних Firebase. Покращення методу можливо за допомогою додавання більшої кількості незалежних змінних для підвищення точності методу, а у застосунок можна додати функціонал прогнозування спортивних звершень.

ABSTRACT

Master`s Thesis: 64 pages, 18 images., 5 tables, 2 applications, 25 title.

MOBILE APPLICATION, METHOD ACCURACY, PREDICTION, TRAINING, REGRESSION, MULTIPLE REGRESSION, INDEPENDENT VARIABLE, DEPENDENT VARIABLE, iOS

The object of study is system for planning training process.

Subject of research is method of data predicting, and in particular, athlete`s physique in dependence from exercising.

The aim of the thesis – research and development of the method for prediction of human physique for training planning system.

Research methods – system analysis, regression analysis.

As a result of conducted research was created the method for predicting human physique based on collected data during trainings. Human body prediction is relevant because there is no method that allows to exactly predict human body at least in one aspect. Developed method could be used by athletes of different level, at sports medicine, and also at medicine at all for showing a doctor more detailed process of patient`s training. Application was developed with programming language Swift and framework UIKit and although with storage Firebase. Method could be improved by adding more independent variables for raising method accuracy, as well the application could be filled with a method for predicting sports achievements.

ЗМІСТ

ВСТУП.....	7
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	9
1.1 Класифікація інформаційних систем та їх архітектура.....	9
1.2 Аналіз даних як метод обробки інформації	14
1.2.1 Важливість аналізу даних для бізнесу	14
1.2.2 Процес аналізу даних.....	15
1.2.3 Важливість аналізу даних у дослідженні.....	16
1.2.4 Типи аналізу даних.....	16
1.2.5 Техніки аналізу даних.....	18
1.3 Використання прогнозування в професійному спорті	19
2 МЕТОД РЕГРЕСІЙНОГО АНАЛІЗУ ЯК МЕТОД ПЕРЕДБАЧЕННЯ	21
2.1 Визначення регресійного аналізу	21
2.2. Цілі регресійного аналізу	22
2.3 Припущення регресійного аналізу	23
2.4 Проста модель регресії.....	24
2.5 Модель множинної регресії	25
2.6 Коефіцієнт детермінації R²	26
2.7 t-критерій Стьюдента.....	27
2.8 Постановка задачі	29
2.9 Використання регресійного аналізу на олімпіадах	29
2.10 Розробка методу прогнозування параметрів статури людини	34
2.11 Опис результатів обчислювального експерименту	37
3 ОПИС РОЗРОБЛЕНОГО ДОДАТКУ	40
3.1 Опис розробки	40
3.2 Технічні вимоги	41
3.3 Проектування додатку	41
3.3.1 Розробка інтерфейсу користувача.....	42
3.4 Сторінка прогнозування результатів	48
ВИСНОВКИ	50
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	52
ДОДАТОК А Графічні матеріали	55
ДОДАТОК Б Текст програми.....	59

ВСТУП

У ряді досліджень була зроблена спроба передбачити майбутні олімпійські результати з легкої атлетики та плавання на основі тенденцій, показаних на попередніх олімпійських іграх. Деякі використовували лінійні моделі для побудови графіка та прогнозування змін, тоді як інші використовували методи оцінки кількох кривих на основі обернених, сигмоїдальних, квадратичних, кубічних, складених, логістичних, зростаючих та експоненціальних функцій. Нелінійні моделі відображалися ближче до фактичних даних і використовувалися для прогнозування змін продуктивності через 10, 100 і 1000 років у майбутньому. Деякі моделі передбачали, що в деяких випадках часи та відстані між чоловіками та жінками будуть перетинатися, і жінки в кінцевому підсумку демонструватимуть кращу продуктивність, ніж чоловіки. Прогнози з використанням математичних моделей на основі легкої атлетики до 1996 року та результатів у плаванні до 1998 року оцінювалися на основі того, наскільки точно вони передбачали спринти та стрибки, а також результати плавання у вільному стилі як для чоловіків, так і для жінок на Олімпійських іграх 2000 та 2004 років. Аналіз показав, що прогнози були ближчими для коротких змагань з плавання, де фактичний час на дистанціях 50 м та жінках на 50 та 100 м у чоловіків був майже ідентичним прогнозованим. Як для чоловіків, так і для жінок із збільшенням дистанцій пропливу точність прогнозованої моделі зменшилась, коли прогнозований час був на 4,5-7% швидше, ніж фактичний. Математичні моделі не передбачали реальних тенденцій у деяких змаганнях, які наразі демонструють зниження продуктивності, які передбачали постійні покращення в усіх видах легкої атлетики та плавання, обраних для цього дослідження.

У кваліфікаційній роботі зроблено спробу вийти на новий рівень у

прогнозуванні людського організму і визначити наскільки точними можуть бути передбачення, що до поведінки організму людини під навантаженнями. Викладений матеріал у роботі можливо використовувати у великій кількості різних сфер та напрямків, наприклад, у спортивній медицині. Також викладені методи у роботі допоможуть спортсменам у різних галузях зрозуміти темпи росту свого фізичного стану та проаналізувати або скорегувати подальші дії.

Об'єктом дослідження є система планування тренувального процесу. Предмет дослідження є методи прогнозування даних, і зокрема параметрів статури людини в залежності від фізичних навантажень.

Мета роботи – дослідження та розробка методу прогнозування параметрів статури людини для системи планування тренувань.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Класифікація інформаційних систем та їх архітектура

Сьогодні людина, будь-яка спільнота і всесвіт оточені інформаційними системами. За останні 50 років інформаційні системи кардинально змінили напрямок свого росту. Спочатку це було щось дуже дороге і незрозуміле, але вже сьогодні кожна людина носить в кишені доступ до сотні або навіть тисячі інформаційних систем. Ціла бізнес індустрія будується на спеціалістах у цій галузі. Саме тому класифікувати інформаційні системи досить складно через їх широкий попит. Інформаційні системи (ІС) класифікуються за багатьма ознаками [1], наприклад, за способом використання, рівнем автоматизації, сферою застосування, функціональним призначенням тощо. Розглянемо класифікації ІС за суттєвими ознаками:

За сферою застосування:

- а) системи для наукових досліджень;
- б) системи автоматизованого проектування;
- в) системи організаційного управління;
- г) системи управління організаційно технічними процесами;
- д) системи управління виробничими процесами;
- е) системи управління технологічними процесами.

За призначенням:

- а) культорологічні;
- б) владні;
- в) науково-технічні;
- г) соціальні;
- д) фінансово-економічні.

За часом обробки інформації:

- а) реального часу;
- б) квазіреального часу;

в) нереального часу.

Також широкого попиту ІС набули у фітнес індустрії та носимій електроніці. Їх використовують для відстеження та контролю життєвих показників. У професійному спорті використовують різні моніторингові системи датчиків які відстежують життєво важливі параметри активності людини під навантаженням, наприклад серцевий ритм або артеріальний тиск. Завдяки зібраним даним формуються фізіологічний та біомеханічний профіль атлета. Ці профілі допомагають зрозуміти рівень підготовки спортсмена, побудувати оцінку його стану здоров'я, або оптимізувати тренувальні навантаження. Але такі системи не завжди коштують велику кількість грошей або їх неможливо добути звичайному користувачеві. Збір схожого набору даних також є доступним таким пересічним користувачам. Найпростішими прикладами легкодоступних для кожної людини систем, які б збирали дані за добовою активністю та харчуванням є застосунки для підрахунку калорій або Apple Watch. Завдяки такого типу годинникам можливо приблизно підрахувати добове використання/споживання калорій, витрату калорій під час різних типів тренувань та відстежувати серцебиття, а застосунки для підрахування корисних мікро- і макро- елементів допоможуть збалансувати харчування. Завдяки цьому набору даних можливо спланувати схуднення, набір м'язової маси або підтримання поточного стану.

На сьогодні найчастіше зустрічаються дві архітектури комп'ютерних систем: файл-сервер та клієнт-сервер.

Інформаційна система типу файл-сервер складається з трьох компонент: сервер баз даних, клієнт (персональний комп'ютер із клієнтськими застосуваннями і СУБД), мережа і комунікаційне програмне забезпечення (рис. 1.1).

На сервері розташовані СУБД і файли, які необхідні для роботи клієнтських застосувань. Клієнтські застосування і їхні персональні СУБД розташовані та функціонують на окремих робочих станціях і звертаються до

файлового сервера тільки в міру потреби отримання доступу до файлів. Сервер відбирає з бази потрібні файли (а не окремі їх записи), які мережею відправляються клієнтові для опрацювання. Таким чином, файловий сервер функціонує як сумісно використовуваний жорсткий диск. Архітектура з використанням файлового сервера характеризується такими основними недоліками: великий обсяг мережевого графіка; на кожній робочій станції має бути повна копія користувацької СУБД; управління паралельністю, відновленням і цілісністю бази даних ускладнюються, оскільки доступ до одних і тих самих файлів здійснюється одночасно кількома СУБД [2].



Рисунок 1.1 – Архітектура файл-сервер

Клієнт-серверна інформаційна система складається з трьох основних компонент: програмне забезпечення сервера; програмне забезпечення кінцевого користувача; проміжне програмне забезпечення (рис. 1.2).

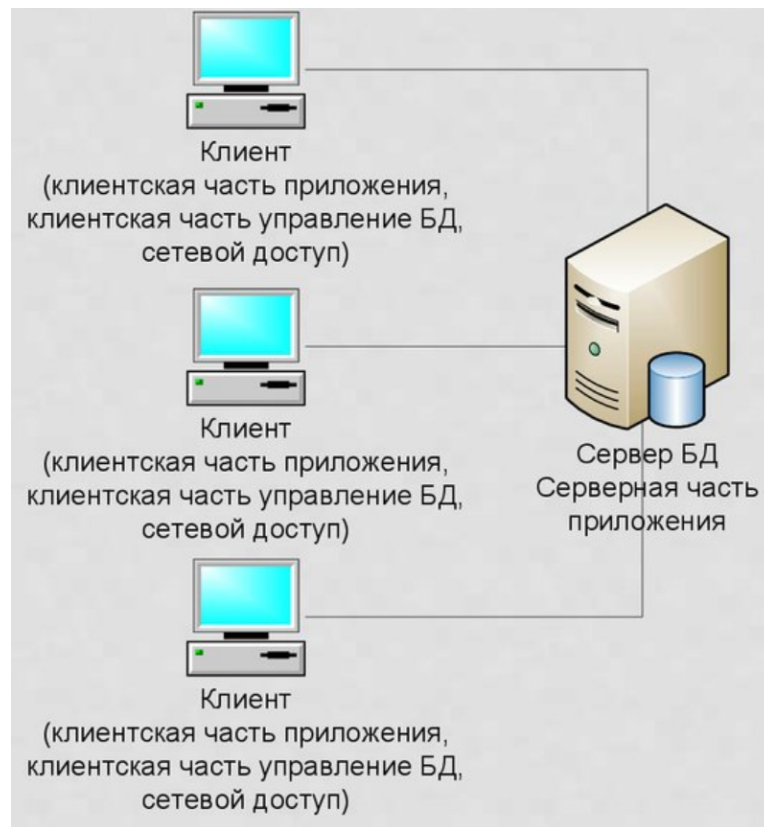


Рисунок 1.2 – Архітектура клієнт-сервер

Програмне забезпечення сервера забезпечує обслуговування клієнтів. Для реалізації архітектури клієнт-сервер зазвичай використовують багатокористувацькі СУБД, наприклад, Oracle або Microsoft SQL Server. У таких СУБД передбачені механізми блокування та елементи управління багатокористувацьким доступом, які забезпечують захист даних від небезпеки паралельного доступу. Крім цього, серверу баз даних доводиться охороняти дані від несанкціонованого доступу, оптимізувати запити до бази даних, забезпечувати цілісність даних і контроль завершення транзакцій. У клієнт-серверній організації клієнти можуть бути досить "тонкими", а сервер має бути "товстим" настільки, щоб задовольняти потреби всіх клієнтів.

До програмного забезпечення кінцевого користувача відносять засоби розробки програм і генератори звітів, у тому числі електронні таблиці і текстові процесори. За допомогою цього програмного забезпечення користувачі встановлюють зв'язок із сервером, формують запити, які

автоматично генеруються в запити мовою SQL і відправляються на сервер. Сервер приймає і опрацьовує запити, а потім передає отримані результати клієнтам.

Проміжне програмне забезпечення - це та частина системи "клієнт-сервер", яка пов'язує програмне забезпечення кінцевого користувача із сервером. Схема клієнт-сервер проста: клієнт направляє серверу запит на потрібні дані; сервер їх приймає, опрацьовує і відправляє клієнтові тільки ті дані, які були замовлені. Дворівнева модель клієнт-сервер оптимальна для підприємств із кількістю користувачів меншою за 100, оскільки операційна система сервера під час обслуговування великої кількості клієнтів надто перевантажується управлінням численними підключеннями до сервера.

Трирівнева модель, на відміну від дворівневої, розв'язує проблеми масштабування. У разі використання трирівневої моделі, окрім клієнта і сервера, є ще й додатковий проміжний ланцюг (сервер застосування)[3], який управляє транзакціями - аналізує запити, організує їх чергу, спрямовує запити на виконання тощо (рис. 1.3).

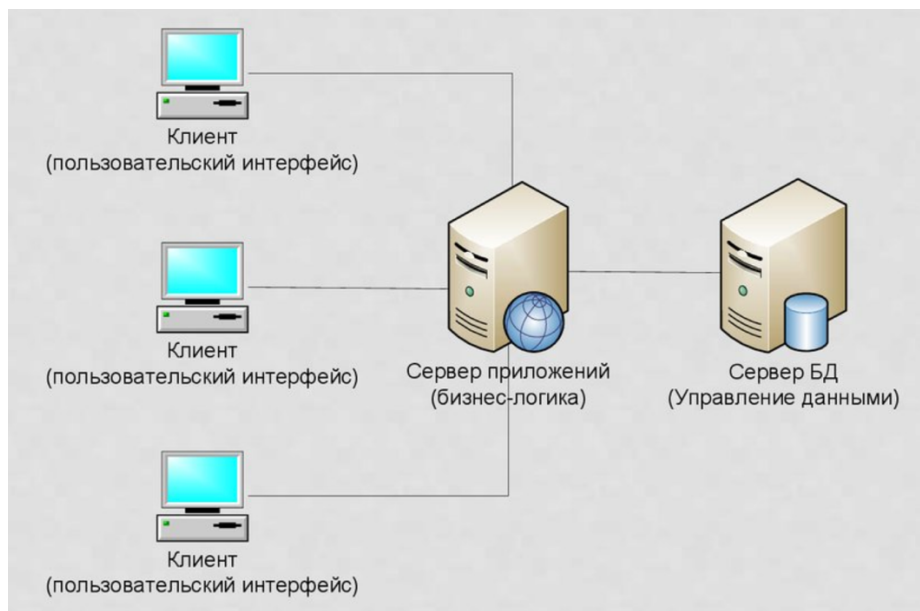


Рисунок 1.3 – Трирівнева архітектура клієнт-сервер

1.2 Аналіз даних як метод обробки інформації

Сьогодні дані для опрацювання в різних інформаційних системах збираються в різних точках процесів і транзакцій, що має величезний потенціал, щоб змінити спосіб роботи на краще [4]. Однак цей аналіз даних може додати цінності для бізнесу лише тоді, коли його аналізують, щоб отримати уявлення про те, як покращити свої продукти та послуги. Аналіз даних дає змогу дізнатися та інтерпретувати інформацію, щоб визначити цінності.

1.2.1 Важливість аналізу даних для бізнесу

Наведемо кілька причин, чому аналіз даних є важливою частиною ведення бізнесу сьогодні [5]:

а) краще націлювання на клієнтів: в жодному разі не потрібно витрачати дорогоцінний час, ресурси та гроші свого бізнесу на створення рекламних кампаній, орієнтованих на демографічні групи, які майже не цікавляться товарами та послугами, які пропонує бізнес. Аналіз даних допомагає зрозуміти, на чому слід зосередити свої рекламні зусилля;

б) краща ознайомленість зі своїми цільовими клієнтами: аналіз даних відстежує, наскільки добре продукти та кампанії працюють у цільовій демографічній групі. Завдяки аналізу даних бізнес може отримати краще уявлення про звички цільової аудиторії щодо витрат, наявний дохід та найімовірніші сфери інтересів. Ці дані допомагають підприємствам встановлювати ціни, визначати тривалість рекламних кампаній і навіть спрогнозувати необхідну кількість товарів;

в) зменшення експлуатаційних витрат: аналіз даних показує, які галузі бізнесу потребують більше ресурсів і грошей, а які не виробляють, тому їх слід скоротити або повністю виключити;

г) кращі методи вирішення проблем: усвідомлені рішення, швидше за все, будуть успішними. Дані надають підприємствам інформацію. З'являється можливість побачити, куди веде цей прогрес. Аналіз даних допомагає підприємствам зробити правильний вибір і уникнути дорогих підводних каменів;

д) більш точні дані: дані, про які йде мова, повинні бути точними. Аналіз даних допомагає підприємствам отримувати релевантну, точну інформацію, придатну для розробки майбутніх маркетингових стратегій, бізнес-планів та перебудови бачення чи місії компанії.

1.2.2 Процес аналізу даних

Процес аналізу даних або етапи аналізу даних включає збір всієї інформації, її обробку, вивчення даних і використання їх для пошуку закономірностей та інших ідей [6], а саме:

а) збір вимог до даних: необхідно відповісти на запитання, чому робиться аналіз, який тип аналізу даних буде використовуватись і які дані ви плануєте аналізувати;

б) збір даних: Керуючись визначеними вимогами, настав час зібрати дані з джерел. Джерела включають тематичні дослідження, опитування, інтерв'ю, анкети, пряме спостереження та фокус-групи. Обов'язково потрібно впорядкувати зібрані дані для аналізу;

в) очищення даних: не всі зібрані дані, будуть корисними, тому дані очищують. У цьому процесі виділяються пробіли, повторювані записи та основні помилки. Очищення даних є обов'язковим перед відправленням інформації на аналіз;

г) аналіз даних: Тут використовується програмне забезпечення для аналізу даних та інші інструменти, щоб допомогти інтерпретувати, зрозуміти

дані та зробити висновки. Інструменти аналізу даних включають Excel, Python, R, Looker, Rapid Miner, Chartio, Metabase, Redash і Microsoft Power BI;

д) інтерпретація даних: тепер, коли є результат, потрібно його інтерпретувати та виявити найкращі дії на основі висновків;

е) візуалізація даних: це відмінний спосіб показати інформацію графічно так, щоб люди могли її прочитати та зрозуміти. Для цього використовуються діаграми, графіки, карти, маркери або багато інших методів. Візуалізація дозволяє отримати цінну інформацію, допомагаючи порівнювати набори даних і спостерігати за взаємозв'язками.

1.2.3 Важливість аналізу даних у дослідженні

Величезна частина роботи — просіяти дані. Це буквально визначення «дослідження». Однак сьогодення інформаційна ера регулярно породжує хвилю даних, достатньої для того, щоб пригнічувати навіть найвідданішого дослідника. Таким чином, аналіз даних відіграє ключову роль у переробці цієї інформації до більш точної та відповідної форми, що полегшує дослідникам роботу. Аналіз даних також надає дослідникам широкий вибір різних інструментів, таких як описова статистика, аналіз висновків та кількісний аналіз. Отже, підводячи підсумок, аналіз даних пропонує дослідникам кращі дані та кращі способи аналізу та вивчення цих даних.

1.2.4 Типи аналізу даних

Сьогодні існує півдюжини популярних типів аналізу даних, які зазвичай використовуються у світі технологій та бізнесу [7]. Вони є:

а) діагностичний аналіз: діагностичний аналіз відповідає на питання «Чому це сталося?» Використовуючи статистику, отриману в результаті статистичного аналізу, аналітики використовують діагностичний аналіз, щоб

визначити закономірності в даних. В ідеалі аналітики знаходять подібні закономірності, які існували в минулому, і використовують ці рішення для вирішення сучасних проблем;

б) прогнозний аналіз: прогнозний аналіз відповідає на питання: «Що найімовірніше станеться?» Використовуючи закономірності, знайдені в старих даних, а також у поточних подіях, аналітики прогнозують майбутні події. Хоча на 100 відсотків точного прогнозування не існує, шанси покращуються, якщо аналітики мають достатньо детальної інформації та дисципліну для її ретельного дослідження;

в) рекомендаційний аналіз: потрібно змішати всі ідеї, отримані від інших типів аналізу даних, і отримаємо директивний аналіз. Іноді проблему не можна вирішити лише за допомогою одного типу аналізу, і замість цього потрібно кілька аналізів;

г) статистичний аналіз: Статистичний аналіз відповідає на питання «Що сталося?» Цей аналіз охоплює збір даних, аналіз, моделювання, інтерпретацію та представлення за допомогою інформаційних панелей. Статистичний аналіз розбивається на дві підкатегорії:

– описовий: описовий аналіз працює з повними чи виділеними зведеними числовими даними. Він ілюструє середні значення та відхилення в безперервних даних, а також відсотки та частоти в категорійних даних;

– висновок: Висновок працює із зразками, отриманими з повних даних. Аналітик може прийти до різних висновків з одного і того ж вичерпного набору даних, просто вибравши різні вибірки;

д) аналіз тексту: аналіз тексту також званий «інтелектуальним аналізом даних» використовує бази даних та інструменти аналізу даних для виявлення закономірностей, що містяться у великих наборах даних. Він перетворює вихідні дані в корисну бізнес-інформацію. Аналіз тексту, мабуть, є найпростішим і найбільш прямим методом аналізу даних.

1.2.5 Техніки аналізу даних

Штучний інтелект знаходиться на підйомі і зарекомендував себе як цінний інструмент у світі аналізу даних. Супутні методи аналізу включають [8]:

- штучні нейронні мережі;
- дерева рішень;
- еволюційне програмування;
- нечітку логіку.

Математичні та статистичні техніки, що подрібнюють цифри:

- описовий аналіз;
- дисперсійний аналіз;
- дискримінантний аналіз;
- факторний аналіз;
- регресійний аналіз;
- аналіз часових рядів.

Людина є візуально орієнтованою істотою. Зображення привертають нашу увагу та довше залишаються в пам'яті. Техніки графіки та візуалізації включають:

- діаграми;
- рамні діаграми;
- діаграми прямокутного дерева;
- карти;
- діаграми розкиду.

Класифікація прогнозів здійснюється, як правило, по двох ознаках: часовій та функціональній. По часовій ознаці розрізняють прогнози: коротко,

середньо, довгострокові і наддовгострокові. Функціональна класифікація прогнозів припускає їх розподіл на дослідницькі, програмні і ресурсні. Прогнозна модель - модель об'єкту прогнозування, дослідження якої дозволяє одержати інформацію про можливі стани об'єктах в майбутньому і (або) шляхах і термінах їх здійснення.

1.3 Використання прогнозування в професійному спорті

Прогнозування спортивних подій було актуальним ще в кінці ХХ сторіччя. Прикладом цього є спроба побудувати моделі для прогнозування легкоатлетів та плавців ще у 1996 році та 1998 році для плавання. Було розраховано середнє значення фіналістів у кожній події за кожен рік дослідження. Використовувалося середнє, оскільки воно є показником, який репрезентує всі бали в кожній групі. Використання середнього значення фіналістів у цьому дослідженні може бути більш репрезентативним для змін у людській продуктивності, які зафіксовані світовими рекордами. Світовий рекордсмен може бути набагато випереджаючим, ніж будь-який інший учасник, і не відображати загальну продуктивність у змаганні. Наприклад, жіночий світовий рекорд на дистанції 400 метрів вільним стилем, встановлений Трейсі Вікхем у 1978 році, не був покращений до 1988 року на Олімпійських іграх у Сеулі. Середні були потім включені як набір даних для кожної події для кожної Олімпійської ігри для аналізу з використанням програми статистичного пакета для соціальних наук (SPSS) версії 6.1 щоб отримати ряд можливих рівнянь регресії. Для оцінки відповідності кожної похідної функції для кожної окремої події було використано ряд критеріїв.

Щоб дослідити гіпотези відповідності моделі та прогнозування, одинадцять регресійних моделей були індивідуально застосовані до кожного з легкоатлетичних змагань та змагань із плавання. Рівняння регресії, яке найкраще підходить для кожної події, тобто дало найвищий коефіцієнт

детермінації (скорочено R^2), було потім визначено з цих одинадцяти рівнянь. Конкретними критеріями для вибору рівняння регресії з найкращого рівня були величина коефіцієнту детермінації R^2 , значущість аналізу дисперсії α або p -значення та залишки.

2 МЕТОД РЕГРЕСІЙНОГО АНАЛІЗУ ЯК МЕТОД ПЕРЕДБАЧЕННЯ

2.1 Визначення регресійного аналізу

Регресійний аналіз дозволяє передбачити значення однієї змінної на основі іншої з використанням прямої лінії, яка характеризує взаємозв'язок між цими двома змінними.

Регресійний аналіз є одним із найбільш часто використовуваних статистичних методів у соціальних та поведінкових науках, а також у фізичних науках, який передбачає виявлення та оцінку зв'язку між залежною змінною та однією чи кількома незалежними змінними, які також називаються провісниками чи пояснювальними змінними [9]. Це особливо корисно для оцінки та коригування для збентеження. Висувається гіпотеза моделі зв'язку, а оцінки значень параметрів використовуються для розробки оціненого рівняння регресії. Потім використовуються різні тести, щоб визначити, чи є модель задовільною. Якщо модель вважається задовільною, оцінене рівняння регресії можна використовувати для прогнозування значення залежної змінної, заданих значень для незалежних змінних.

Лінійна регресія досліджує залежності, які можна легко описати прямими лініями або їх узагальненням на множині вимірів. Напрочуд велику кількість проблем можна вирішити за допомогою лінійної регресії, а ще більше за допомогою перетворення вихідних змінних, що призводить до лінійних зв'язків між перетвореними змінними.

Коли існує одна неперервна залежна змінна і одна незалежна змінна, аналіз називається простим аналізом лінійної регресії. Цей аналіз передбачає, що між двома змінними існує лінійний зв'язок. Множина регресія полягає в тому, щоб дізнатися більше про зв'язок між кількома незалежними або провісними змінними та залежною чи критеріальною змінною.

Незалежні змінні - це характеристики, які можна виміряти безпосередньо; ці змінні також називаються провісниками або пояснювальними змінними, які використовуються для прогнозування або пояснення поведінки залежної змінної.

Залежна змінна — характеристика, значення якої залежить від значень незалежних змінних.

Ознаки надійності і валідності регресійної моделі [10]:

а) модель має інтуїтивно зрозумілий сенс. Легко зрозуміти та інтерпретувати модель;

б) усі коефіцієнти мають бути статистично значущі (р-значення менше ніж 0,05);

в) ознаки пов'язані з коефіцієнтами належним чином;

г) чи передбачає модель значення, які є достатньо близькими до фактичних значень;

д) достатня надійність моделі (Високий R-квадрат, низька стандартна помилка тощо).

2.2. Цілі регресійного аналізу

Регресійний аналіз використовується для пояснення змінення залежної змінної за допомогою однієї або кількох незалежних або контрольних змінних [11] і для аналізу зв'язків між змінними для відповіді на питання про те, наскільки залежна змінна змінюється зі змінами кожної із змінних незалежної, а також спрогнозувати або передбачити значення залежної змінної на основі значень змінних незалежної.

Основна мета регресії полягає в тому, щоб розробити лінійний зв'язок між змінною відповіді та пояснювальними змінними для цілей прогнозування, передбачає, що існує функціональна лінійна залежність, а альтернативні підходи (функціональна регресія) є кращими.

2.3 Припущення регресійного аналізу

Регресійна модель засновується на наступних припущеннях [12]:

- а) зв'язок між незалежною змінною і залежною є лінійною;
- б) очікуване значення терміну помилки дорівнює нулю;
- в) дисперсія члена помилки є постійною для всіх значень незалежної змінної, припущення гомоскедастичності;
- г) автокореляції немає;
- д) незалежна змінна не корелює з членом помилки;
- е) термін помилки розподіляється нормально;
- ж) середня різниця між спостережуваним значенням (y_i) і прогнозованим значенням (\hat{y}_i) дорівнює нулю;
- з) у середньому оцінені значення помилок і значення незалежних змінних не пов'язані між собою;
- и) квадрат різниці між спостережуваним значенням і прогнозованим значенням подібний;
- к) є деякі варіації незалежної змінної. Якщо є більше однієї змінної в рівнянні, то дві змінні не повинні бути повністю корельовані.

Точка перетину з віссю ординат – це точка, в якій регресія перетинає вісь ординат. Точка забезпечує вимірювання середнього значення залежної змінної, коли нахил є нульовим.

Коефіцієнт регресії – значення змінної залежить від точки перетину з віссю ординат. Нульовий коефіцієнт регресії означає, що незалежна змінна не має впливу на залежну змінну. Для лінійної моделі нахил не дорівнює пружності.

2.4 Проста модель регресії

Проста лінійна регресія – це статистичний метод, який дозволяє узагальнити та вивчити зв'язки між двома безперервними (кількісними) змінними [13]. У причинно-наслідковому зв'язку незалежна змінна є причиною, а залежна змінна — наслідком. Найменші квадрати лінійної регресії — це метод прогнозування значення залежної змінної y на основі значення незалежної змінної x .

Одна змінна, позначена (x), розглядається як предиктор, пояснювальна або незалежна змінна, інша змінна, позначена (y), розглядається як відповідь, результат або залежна змінна.

Математично модель регресії представлена таким рівнянням:

$$y = \beta_0 \pm \beta_1 x_1 \pm \varepsilon_1$$

$$\beta_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\bar{x} = \frac{\sum x}{n}; \bar{y} = \frac{\sum y}{n};$$

де: x – незалежна змінна;

y – залежна змінна;

β_1 – нахил лінії регресії;

β_0 – точка перетину лінії регресії та осі y ;

n – кількість випадків або осіб;

$\sum xy$ – сума добутку залежних і незалежних змінних;

$\sum x$ – сума залежних змінних;

$\sum y$ – сума незалежних змінних;

$\sum x^2$ – сума квадратів залежної змінної.

2.5 Модель множинної регресії

Множинна регресія є розширенням простої лінійної регресії. Вона використовується, коли ми хочемо передбачити значення залежної змінної (цільової або критеріальної) на основі значення двох або більше незалежних змінних (провісників або пояснювальних змінних) [14]. Множинна регресія дозволяє визначити загальну відповідність (пояснена дисперсія) моделі та відносний внесок кожного з предикторів у загальну пояснену дисперсію. Наприклад, ви можете знати, наскільки різницю в результативності іспиту можна пояснити часом перегляду та відвідуванням лекції «в цілому», а також «відносним внеском» кожної незалежної змінної в пояснення дисперсії.

Математично модель множинної регресії представлена таким рівнянням:

$$Y = \beta_0 \pm \beta_1 X_1 \dots \pm \beta_n X_n \pm u$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}_1 - \beta_2 \bar{x}_2$$

$$\beta_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\beta_2 = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

де: X_i до X_n - представляє незалежні змінні;

Y - залежна змінна;

β_1 - коефіцієнт регресії змінної x_1 ;

β_2 - коефіцієнт регресії змінної x_2 ;

β_0 - точка перетину лінії регресії та осі y ;

\bar{x}_2 – середнє значення незалежної змінної x_2 .

За допомогою методу відхилення:

$$\sum y = \sum(Y - \bar{Y}) ,$$

$$X_1 \sum x_1 = \sum(X_1 - \bar{X}_1),$$

де \bar{Y} середнє значення залежних змінних;

\bar{X}_1 середнє значення незалежних змінних.

2.6 Коефіцієнт детермінації R^2

Коефіцієнт детермінації демонструє в якій мірі змінення змінної Y пояснюється поведінкою X [15]. Для обчислення коефіцієнту детермінації потрібно звести у квадрат коефіцієнт кореляції. Таким чином, доля варіації Y визначається виразом $1 - R^2$ [16].

$$R^2 = \frac{\text{Regression Sum of Square}(SSR)}{\text{Total Sum of Square}(TSS)}$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$TSS = \sum_{i=1}^n [(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)]^2 = \sum_{i=1}^n (y_i - \bar{y})^2$$

де \hat{y}_i – спргнозоване значення y для i ітерації;

y_i – реальне значення y для i ітерації.

2.7 t-критерій Стьюдента

t-критерій Стьюдента — загальна назва для класу методів статистичної перевірки гіпотез (статистичних критеріїв), заснованих на порівнянні з розподілом Стьюдента [17]. Найчастіші випадки застосування t-критерію пов'язані з перевіркою рівності середніх значень у двох вибірках. Цей параметричний критерій має дві основні форми: незв'язаний (непарний) t – критерій застосовується для порівняння показників двох груп, між елементами яких не існує зв'язку, зв'язаний (парний) t – критерій застосовується для порівняння показників двох груп, між елементами яких існує специфічний зв'язок.

Для застосування даного критерію необхідно аби початкові дані мали нормальний розподіл. У разі застосування двохвибіркового критерію для незалежних виборок також необхідне дотримання умови рівності дисперсій. Існують, проте, альтернативи критерію Стьюдента для ситуації з нерівними дисперсіями. У випадку з розміром вибірки, що трохи відрізняється, застосовується спрощена формула наближених розрахунків (2.1), але якщо розмір вибірки відрізняється значно, застосовується точніша формула (2.2). Кількість ступенів свободи розраховується за формулою (2.3):

$$t = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}, \quad (2.1)$$

$$t = \frac{|M_1 - M_2|}{\sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}, \quad (2.2)$$

$$df = N_1 + N_2 - 2, \quad (2.3)$$

де: M_1, M_2 – середнє арифметичне першої та другої сукупності;

σ_1, σ_2 – стандартне відхилення;

N_1, N_2 – розміри вибірок.

Для обчислення емпіричного значення t-критерію в ситуації перевірки гіпотези про відмінності між двома залежними вибірками (наприклад, двома пробами одного і того ж тесту з часовим інтервалом) застосовується формула (2.4), а кількість ступеней свободи розраховуються за формулою (2.5). Аналогом двостороннього критерію для незалежних вибірок є U-критерій Манна-Уїтні. Для ситуації із залежними вибірками аналогами є критерій знаків і T-критерій Вілкоксона.

$$t = \frac{|M_d|}{\sigma_d/\sqrt{N}} \quad (2.4),$$

$$df = N - 1 \quad (2.5),$$

де: M_d – середня різниця значень;

σ_d – стандартне відхилення різниць.

Після розрахунку визначається критичне значення t-критерію Стьюдента для необхідного рівня значущості (наприклад, $p = 0,05$) і при даному числі ступенів свободи f за таблицею критичних значень Стьюдента. Після чого порівнюється критичне і розраховане значення критерію. Якщо розраховане значення t-критерію Стьюдента дорівнює або більше критичного, знайденого за таблицею, то, робиться висновок про статистичну значущість відмінностей між порівнюваними величинами. Якщо значення розрахованого t-критерію Стьюдента менше табличного, отже відмінності порівнюваних величин статистично не значимі.

2.8 Постановка задачі

Оскільки спорт є невід'ємною частиною нашого життя, а прогнозування параметрів людської статури ще не було вивчено у повному обсязі, то розробка та використання методу в інформаційних системах моніторингу та планування фізичних параметрів людини є актуальною задачею.

Отже, задача дослідження – дослідити та розробити метод прогнозування параметрів статури людини для системи планування тренувань.

При розробці методу потрібно взяти до уваги:

- а) вхідною інформацією є послідовність замірів певного параметру;
- б) прогноз потрібно будувати, на основі попередньо отриманих даних;
- в) прогноз має розраховуватись на основі рівняння регресії;

Для досягнення зазначеної цілі необхідно:

- а) проаналізувати існуючі методи прогнозування;
- б) виконати підбір кращого методу прогнозування під конкретний тип навантаження;
- в) інтегрувати механізм до застосунку;
- г) провести експериментальні дослідження, що до життєздатності існуючого підходу та методів.

2.9 Використання регресійного аналізу на олімпіадах

На сьогодні було зроблено багато спроб у прогнозуванні поведінки людини як такої. В контексті цієї роботи найяскравішим прикладом будуть регресійні моделі, що створювались задля прогнозування результатів олімпійських ігор [19].

Прогнозування спортивних подій було актуальним ще в кінці ХХ сторіччя. Прикладом цього є спроба Лакі і Хіслвуда побудувати моделі для

прогнозування легкоатлетів та плавців ще у 1996 році та 1998 році. Для запису часу використовувався стандарт Міжнародної федерації нації любителів. Потім ці часи були перетворені з формату хвилин і секунд у формат лише секунд, щоб полегшити обчислення під час застосування методів регресії. Потім було розраховано середнє значення фіналістів у кожній події за кожен рік дослідження. Використовувалося середнє, оскільки воно є показником, який репрезентує всі бали в кожній групі. Використання середнього значення фіналістів у цьому дослідженні може бути більш репрезентативним для змін у людській продуктивності, які зафіксовані світовими рекордами. Світовий рекордсмен може бути набагато випереджаючим, ніж будь-який інший учасник, і не відображати загальну продуктивність у змаганні. Наприклад, жіночий світовий рекорд на дистанції 400 метрів вільним стилем, встановлений Трейсі Вікхем у 1978 році, не був покращений до 1988 року на Олімпійських іграх у Сеулі. Середні були потім включені як набір даних для кожної події для кожної Олімпійської ігри для аналізу з використанням програми статистичного пакета для соціальних наук (SPSS) щоб отримати ряд можливих рівнянь регресії. Для оцінки відповідності кожної похідної функції для кожної окремої події було використано низку критеріїв.

Щоб дослідити гіпотези відповідності моделі та прогнозування, одинадцять регресійних моделей були індивідуально застосовані до кожного з легкоатлетичних змагань та змагань із плавання. Рівняння регресії, яке найкраще підходить для кожної події, тобто дало найвищий коефіцієнт детермінації (скорочено R^2), було потім визначено з цих одинадцяти рівнянь. Конкретними критеріями для вибору рівняння регресії з найкращого рівня були величина R^2 , значущість аналізу дисперсії альфа або р-значення та залишки.

Коефіцієнт детермінації (R^2) є мірою точності використаної моделі. Коефіцієнт детермінації 1,00 вказує на ідеальну модель, де передбачені значення відповідають фактичним значенням для кожної незалежної змінної.

Якщо завдяки рівному R^2 можна було вибрати більше однієї моделі, використовувалася найпростіша модель за принципом економії, тобто уникнення марнотратства та дотримання найпростішої пояснювальної моделі.

Залишки є різницею між фактичним значенням і прогнозованим значенням для кожного випадку, використовуючи рівняння регресії і чим менше залишок, тим краще підходить модель. Для кожної моделі залишки були згенеровані програмою SPSS. Велика кількість позитивних залишків вказує на те, що прогноз є завищеною оцінкою (швидше, ніж фактична продуктивність), а велика кількість негативних залишків вказує на недооцінку (більш повільний час, ніж фактична продуктивність).

Рівень значущості, або значення p , є представленням зв'язку між моделлю та даними. Чим менше значення p , тим вищий рівень значущості та більший зв'язок, коли мале значення p вказує на малу ймовірність того, що близькість передбачуваних значень до фактичних значень через випадковість мала.

Також бралася до уваги здатність моделі генерувати екстраполяції, які здаються розумними в порівнянні з попередніми засобами. Коли модель генерувала екстраполяції, які, здається, не узгоджуються з фактичними результатами, цю модель відкидали і вибирали модель з наступним найвищим коефіцієнтом детермінації.

Після вибору моделі, яка буде використовуватися, відповідно до вищевказаних критеріїв, було визначено рівняння, що найкраще підходить, шляхом застосування отриманих констант і коефіцієнтів до загальної формули для цієї моделі. Використовуючи це рівняння, було розраховано прогноз середнього результату заходу на кожній олімпіаді. На цьому етапі на додаток до прогнозованих середніх за допомогою відповідного рівняння регресії також були створені графіки, що представляють середні показники минулих і майбутніх результатів для кожної події в кожній олімпіаді.

Щоб передбачити рівень продуктивності у 2000 та 2004 роках, був обраний набір даних, який забезпечував найбільшу точність, а дані 1996 року повторно включені до набору даних, якщо це доречно. Була зроблена серія регресій з використанням найкращої моделі та набору даних для кожної події. Використовуючи константи та коефіцієнти, згенеровані регресійними моделями, потім були розраховані майбутні прогнози. Важливо зазначити, що в деяких змаганнях середній показник для всього поля учасників не завжди був можливим через дискваліфікацію або травму. У разі травми учасник не завершив змагання. Така ситуація мала місце лише в кількох подіях.

Дані для прогнозованих значень і фактичних значень надаються в таблиці для кожної події. У таблиці 1 наведено події, математичні функції, рівняння та значення R^2 , отримані з легкоатлетичних змагань 1996 року на 100 м, 400 м, стрибках у довжину та у висоту серед чоловіків та жінок [20].

Таблиця 2.1 – Дані, отримані з легкоатлетичних змагань 1996.

Змагання	Тип регресії	Рівняння	R^2
Чоловіки 100 м.	Інверсія	$Y = b_0 + (b_1/t)$.659
Жінки 100 м.	Кубічна	$Y = b_0 + b_3t^3$.902
Чоловіки 400 м.	S	$Y = e^{(b_0 + b_1/t)}$.907
Жінки 400 м.	S	$Y = e^{(b_0 + b_1/t)}$.843
Чоловічий стрибок у довжину	Кубічна	$Y = b_0 + b_3t^3$.777
Жіночий стрибок у довжину	Інверсія	$Y = b_0 + (b_1/t)$.894
Чоловічий стрибок у висоту	Комплексна	$Y = b_0(b_1)^t$.944
	Логістична		.944
	Експоненційна	$Y = b_0e^{b_1t}$.944

	Зростаюча	$Y = e^{b_0 b_1 t}$.944
Жіночий стрибок у висоту	Комплексна	$Y = b_0 (b_1)^t$.947
	Логістична		.947
	Експоненційна	$Y = b_0 e^{b_1 t}$.947
	Зростаюча	$Y = e^{b_0 b_1 t}$.947

Усі значення R^2 значущі при $p < 0.05$, де b_0 – константа, b_1, b_3 – коефіцієнти регресії, t – рік, Y – результат для кожного змагання.

Тенденції математичних функцій вказують на 400 м для чоловіків і жінок і стрибки у висоту показують ідентичні тенденції змін продуктивності з часом, де 400 м був сигмоподібним, а стрибок у висоту був складним, логістичним, експоненційним і зростаючим. У більшості випадків пояснена дисперсія або значення R^2 були статистично дуже значущими ($p < 0,01$). Цікаво відзначити, що значення R^2 були найвищими для стрибків у висоту у чоловіків та жінок (0,94) і найнижчими для стрибків на 100 м (0,66) та чоловіків у довжину (0,78).

У таблиці 2.2 наведено прогнозовані результати та фактичні результати, досягнуті на Олімпійських іграх 2000 та 2004 років. Можна помітити, що обидва 100-метрові рази для чоловіків перевищили прогноз, тоді як у жінок на 100 метрів були значно нижчі за передбачені значення. На дистанції 400 м, стрибках у довжину та у висоту спортсмени як чоловіки, так і жінки були нижче передбаченого часу та дистанцій. Однак фактичні результати 400-метрових стрибків у чоловіків та дистанції стрибків у довжину покращилися з 2000 по 2004 рік [21]. У бігу на 400 м серед жінок спостерігалось зниження продуктивності, а у стрибках у висоту серед жінок та чоловіків показники залишалися відносно статичними з 2000 по 2004 рік.

Таблиця 2.2 – Актуальні та передбачувані результати для чоловіків та жінок на біг 100 та 400 метрів, стрибки у висоту та стрибки у довжину для Олімпіади 2000 та 2004 років.

Передбачення	100Ч	100Ж	400Ч	400Ж	ЧСД	ЖС	ЧСВ	ЖС
						Д		В
2000	10.07	10.84	43.84	48.29	8.36	7.27	2.42	2.05
	с	с	с	с	м	м	м	м
2004	10.03	10.76	43.63	47.78	8.43	7.39	2.45	2.09
	с	с	с	с	м	м	м	м
Результат	100Ч	100Ж	400Ч	400Ж	ЧСД	ЖС	ЧСВ	ЖС
						Д		В
2000	10.05	11.15	44.92	49.92	8.26	6.8м	2.32	1.98
	с	с	с	с	м		м	м
2004	9.93с	11.04	44.67	50с	8.33	6.92	2.31	1.98
		с	с		м	м	м	м

У таблиці 2.2 100Ч – результат бігу на 100 метрів серед чоловіків, 100Ж – результат бігу на 100 метрів серед жінок, 400Ч – результат бігу на 400 метрів серед чоловіків, 400Ж – результат бігу на 400 метрів серед жінок, ЧСД – результати стрибків чоловіків у довжину, ЖСД – результат стрибків жінок у довжину, ЧСВ – результати стрибків чоловіків у висоту, ЖСВ – результат стрибків жінок у висоту.

2.10 Розробка методу прогнозування параметрів статури людини

Під час створення моделі множинної регресії для прогнозування ваги атлета було очевидно, що деякі комплекси вправ мають більшу ефективність. Ми маємо вихідні дані лише за місяць, якщо брати більший відрізок часу тоді важливість врахування такої особливості може значно наблизити модель до

реальних одиниць. За місяць спостережень було вирішено побудувати графік, який би демонстрував скільки кілограмів було витрачено саме завдяки конкретній групі вправ. Результат побудови на рисунку 2.5 [22].

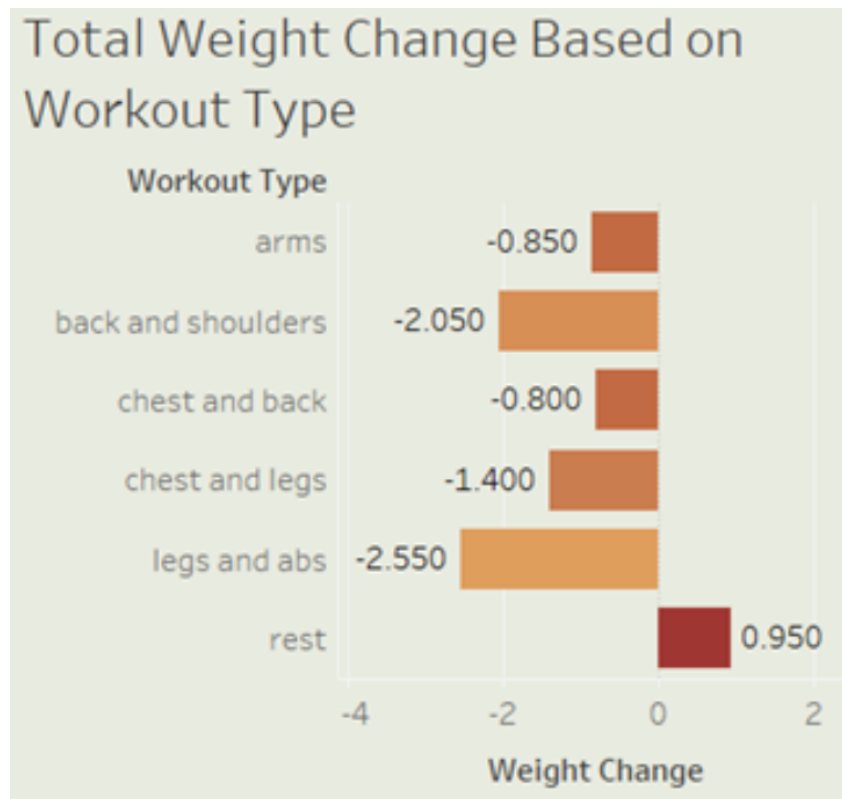


Рисунок 2.5 – Графік сумарної втрати ваги за кожною групою вправ.

Серед усіх видів можна виділити фаворита - і це тренування ніг та м'язів пресу. І саме в цьому полягає основний недолік звичайної множинної регресії. Завдяки XGBoost – імплементація дерева рішень із посиленням градієнту [23], ми отримуємо результати аналізу важливості коефіцієнтів на рисунку 2.6.

Саме на графіку можна чітко побачити, що найбільш важливим фактором під час схуднення була кількість калорій, що перевищують розраховану норму, а серед тренувань найбільш ефективним було тренування ніг та пресу. Досі багатокритеріальна регресія не навчилася враховувати важливість критеріїв, саме тому нам потрібно додати вагові коефіцієнти до існуючого рівняння регресії задля врахування цієї особливості. Але нам

потрібно зберегти певний баланс між впливом змінних, тому потрібно ввести обмеження, що $\sum_{i=1}^n k_i = 1$.

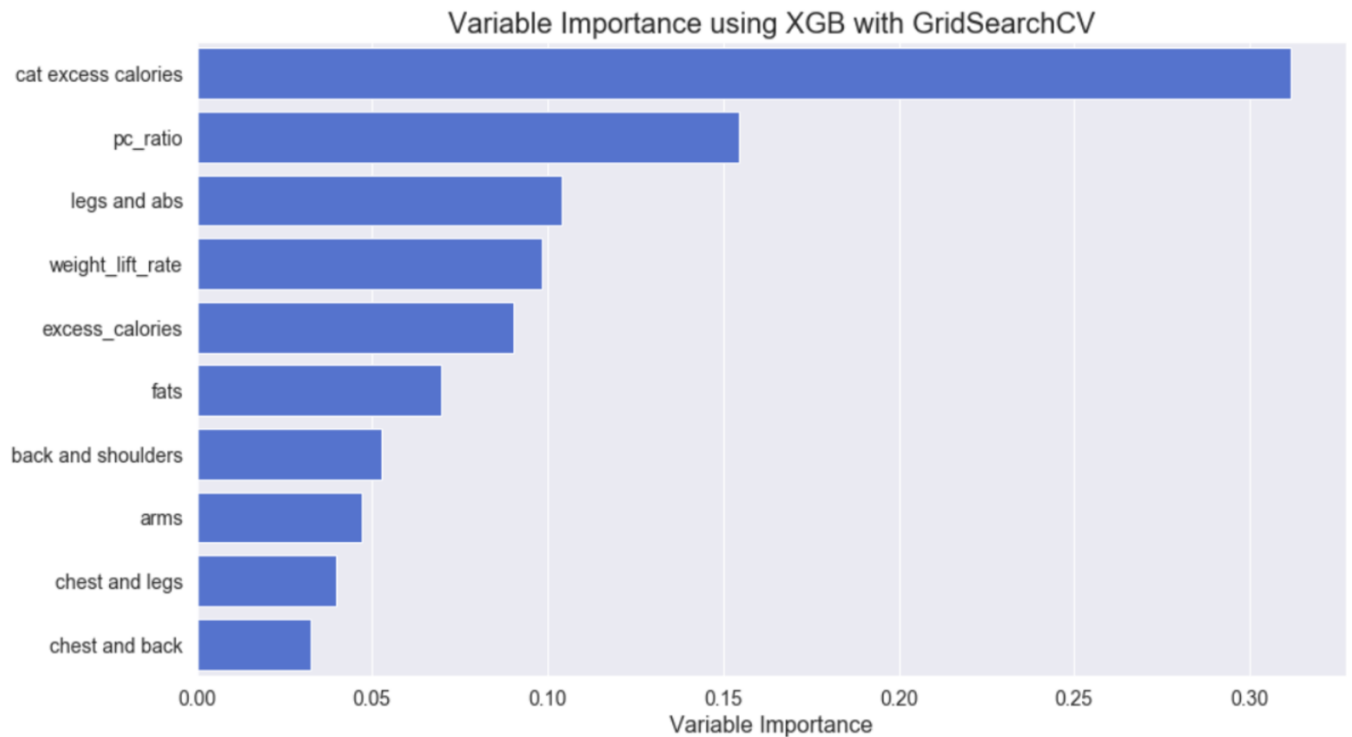


Рисунок 2.6 – Результат аналізу важливості критеріїв.

Рівняння регресії з ваговими коефіцієнтами виглядає наступним чином:

$$Y = \beta_0 \pm \beta_1 k_1 X_1 \dots \pm \beta_n k_n X_n \pm u$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}_1 - \beta_2 \bar{x}_2$$

$$\beta_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\beta_2 = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

де: X_i до X_n - представляє незалежні змінні;

k_i - ваговий коефіцієнт змінної X_i ;

Y - залежна змінна;

β_1 - коефіцієнт регресії змінної x_1 ;

β_2 - коефіцієнт регресії змінної x_2 ;

β_0 - точка перетину лінії регресії та осі y .

2.11 Опис результатів обчислювального експерименту

Використаємо розроблений алгоритм на зібраних даних за тиждень схуднення. Таблиця з початковими даними виглядає наступним чином:

Таблиця 2.3 – Початкові дані.

Вага поточна	Вага у наступний день	Залишок калорій	Тип тренування
72.5	72	-56.31	Груди та ноги
72	72.4	-60.41	Спина та плечі
72.4	72	-68.61	Ноги та прес
72	72.2	-40.4	Руки
72.2	72	-37.08	Груди та спина
72	72.3	46.71	Відпочинок

Під залишком калорій мається на увазі чи вклався спортсмен у розраховані рамки по споживанню їжі. Відповідно до цього, якщо значення негативне, тоді вклався і навіть з'їв менше, а якщо позитивне, тоді переїв. Спочатку треба розрахувати вагові коефіцієнти для кожного з типів тренувань. Відповідно до рисунку 2.5 робимо висновок, що найефективнішим тренуванням по втраті ваги є комбінація м'язів ніг та пресу. На рисунку також зображено скільки вдалося скинути ваги за допомогою конкретної вправи. Візьмемо критерій втрати ваги за показник ефективності і розрахуємо

ефективність вправ за шкалою від 0 до 1, де 2.55 це найкращий показник, тобто 1. Відповідні результати розрахунку ефективності є у таблиці 2.4.

Таблиця 2.4 – Таблиця ефективності тренувань.

Руки	0.33
Спина та плечі	0.804
Грудь та спина	0.3137
Грудь та ноги	0.549
Ноги та прес	1
Відпочинок	0.3725

Отже значення коефіцієнту k будемо брати з таблиці вище в залежності від типу тренувань. Далі були виконані підготовчі розрахунки для змінних, що беруть участь у розрахунках передбачення, серед яких: середнє значення залежної і незалежних змінних, коефіцієнти регресії, точка перетину лінії регресії з віссю y та ін.:

$$n = 6$$

$$\bar{x}_1 = 72,184$$

$$\bar{x}_2 = -24,956$$

$$\bar{y} = 72,15$$

$$\sum x_1^2 = 31262,85;$$

$$\sum x_2^2 = 18348,059;$$

$$\sum x_1 y = 31248;$$

$$\sum x_2 y = -15989,646;$$

$$\sum x_1 x_2 = -15960,635;$$

$$\beta_0 = 0,044;$$

$$\beta_1 = 0,998;$$

$$\beta_2 = -0,003;$$

Тепер коли були розраховані усі необхідні значення для прогнозу. У таблиці 2.5 наведені результати звичайної багатокритеріальної регресії та результати розрахунків з додатковим коефіцієнтом.

Таблиця 2.5 – Порівняння результатів регресій.

Прогнозовані залежні змінні	Регресія з ваговим коефіцієнтом	Звичайна багатofакторна регресія
\widehat{y}_1	72,485	72,578
\widehat{y}_2	72,05	72,093
\widehat{y}_3	72,13	72,13
\widehat{y}_4	71,959	72,124
\widehat{y}_5	72,122	72,222
\widehat{y}_6	71,927	72,011

Тепер можемо розрахувати коефіцієнт кореляції для результатів нової моделі регресії.

$$R^2 = \frac{0,21}{0,155} = 1,356;$$

$$1 - R^2 = 0,356 = 35,6\%;$$

Коефіцієнт детермінації демонструє наскільки результат обчислення залежної змінної пояснюється незалежними змінними. В нашому випадку це 35.6%. Якщо врахувати кількість факторів, що впливають на кінцевий результат в реальному житті, тоді це непоганий результат.

3 ОПИС РОЗРОБЛЕНОГО ДОДАТКУ

3.1 Опис розробки

Користувач - це людина, віком від 14 до 40 років, яка збирається на протязі якогось часу займатися в спортивному залі або виконувати певний набір вправ. Що найголовніше, вона потребує постійного відстежування свого тренувального прогресу та підказок у техніці виконання. Дуже важливим в цьому процесі є концентрація на важливих деталях виконання та техніці бо неправильна техніка може призвести до травмування під час виконання вправи. Програмний продукт допомагає відстежувати тренувальний прогрес за кожною окремою вправою, а також отримати довідку в якій міститиметься: відео пояснення вправи від професіонала, текстовий опис груп м'язів, які використовуються переважно та другорядно, а також текстовий опис виконання вправи.

Найбільш важливими завданнями є запис кількості виконаних підходів і разів у підході, а також отримання швидкої підказки, щодо виконання вправи і опису техніки.

Для вирішення задачі користувач має завести акаунт або увійти у вже існуючий, після чого має зібрати своє тренування і може перейти до його виконання перед тим визначив кількість підходів для вправи. Також користувач може фотографувати стан своєї фігури і зберігати ці фотографії у додатку. Наприклад, цілий рік користувач фотографував з періодичністю через місяць і побачити свій результат.

Однією з основних цілей користувача є не покалічити себе і вести щоденник тренувань, що допоможе йому при майбутньому аналізі своїх результатів та корегуванні тренувань.

Для виконання завдань треба мати логін та пароль від дійсного акаунту, а також знати яку групу м'язів хоче тренувати користувач і в якому напрямку.

Для вирішення завдання користувачу потрібен лише його телефон, але наявність в ньому інтернета допоможе користувачеві оновити список вправ і користуватися вже найбільш актуальною версією.

3.2 Технічні вимоги

Цей програмний продукт може бути використаний користувачами мобільних телефонів Apple. Одніє з вимог є побудова архітектури MVC у застосунку [24]. Цей додаток доступни від моделі iPhone 5 і вище. Рекомендовані технічні характеристики:

- оперативна пам'ять: від 2 Гб;
- процесор: Apple A10 Fusion [25];
- екран:4,7;

3.3 Проектування додатку

Процес розробки під iOS виконується у середовищі Xcode 13. Також існує AppCode, але в ньому немає Interface Builder, який полегшує побудову елементів UI. Для збереження клієнтських даних у додатку використовується реляційна база даних Firebase. Це облачне сховище даних до якого отримує доступ додаток через інтернет. У БД лежать відомості про зареєстрованих користувачів, вага та їх процес тренування (Рис. 3.1).

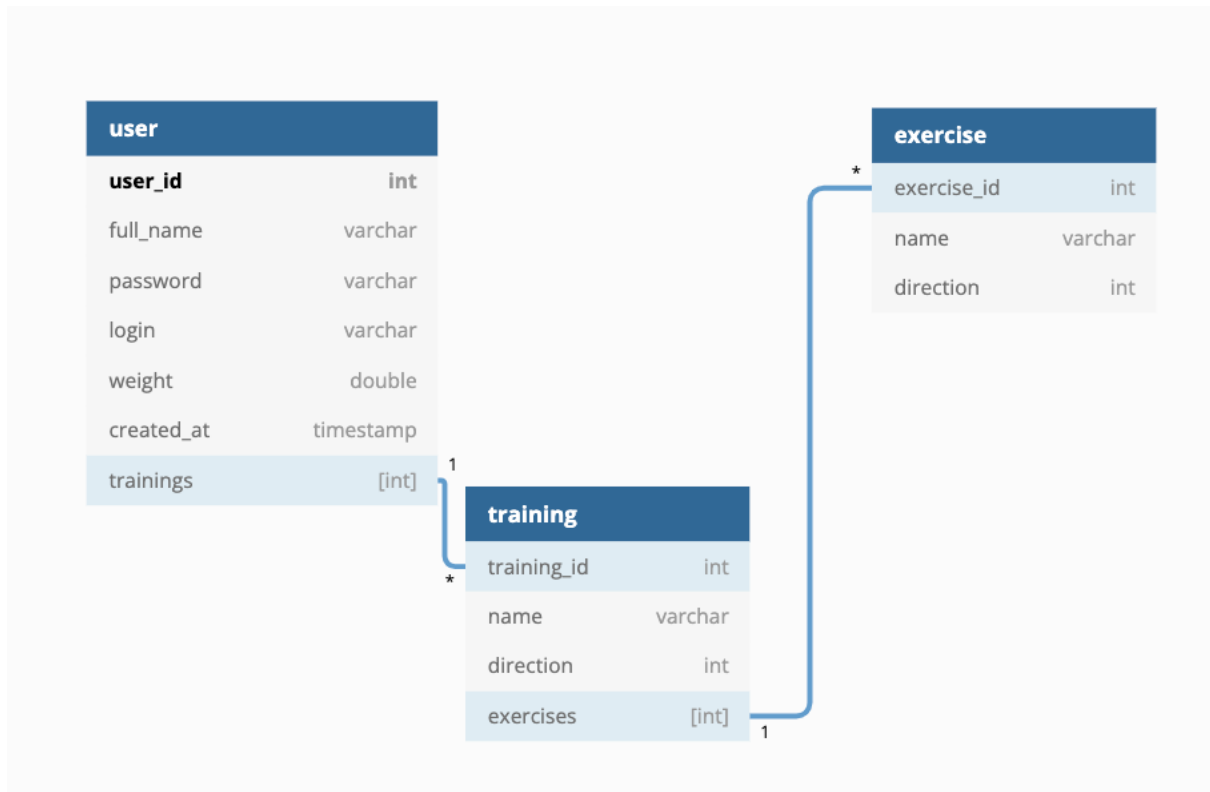


Рисунок 3.1 – Схема бази даних.

Для логіну також використовується сервіс від Firebase, який зберігає на своєму боці дані про користувачів.

3.3.1 Розробка інтерфейсу користувача

З самого початку система пропонує користувачеві увійти під дійсним обліковим записом до системи, але якщо користувач не має облікового запису, тоді він має натиснути кнопку зареєструватись, де користувач має узупевнити поля з особистою інформацією і все це можна побачити вгорі екрану реєстрації (Рис. 3.2). Після введення користувачем потрібних даних уся інформація з полів буде проходити валідацію за загальною наявністю тексту чи заповнення полів символами пробілу.

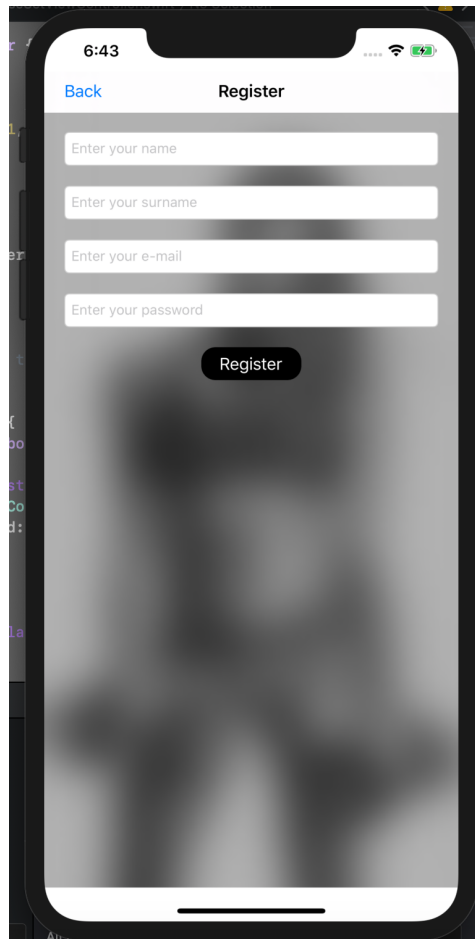


Рисунок 3.2 - Сторінка реєстрації.

Після отримання повідомлення о успішній реєстрації користувач потрапить на сторінку логіну, де має ввести дані від вже існуючого запису, а саме адрес електронної пошти та пароль (Рис. 3.3).



Рисунок 3.3 - Сторінка логіну.

Після входу до вже існуючого облікового запису користувач потрапляє на сторінку довідки, де може отримати детальну інструкцію за конкретною вправою. Для цього він може скористатися пошуком вгорі інтерфейсу або подивитися увесь список запропонованих вправ і обрати потрібну(Рис. 3.4). Для того, щоб додати вправу до списку тренувань потрібно натиснути кнопку "Add", яка знаходиться внизу кожного пункту з вправами. Щоб користувач не губився у інтерфейсі програмного продукту в самому низу кожен екран підписаний і додатково змінює колір, якщо він є активний.

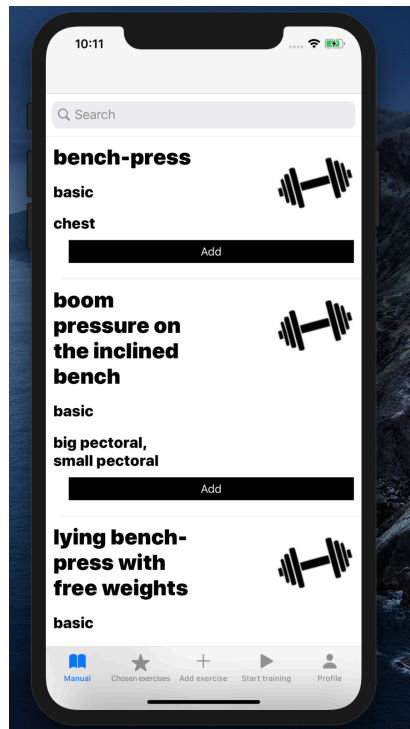


Рисунок 3.4 - Екран довідки з вправ.

Далі користувач може розпочати тренування за обраними вправами і для цього користувач має особливий екран (Рис. 3.5). На якому посередині екрану є лише одна кнопка, яка розпочинає тренування.

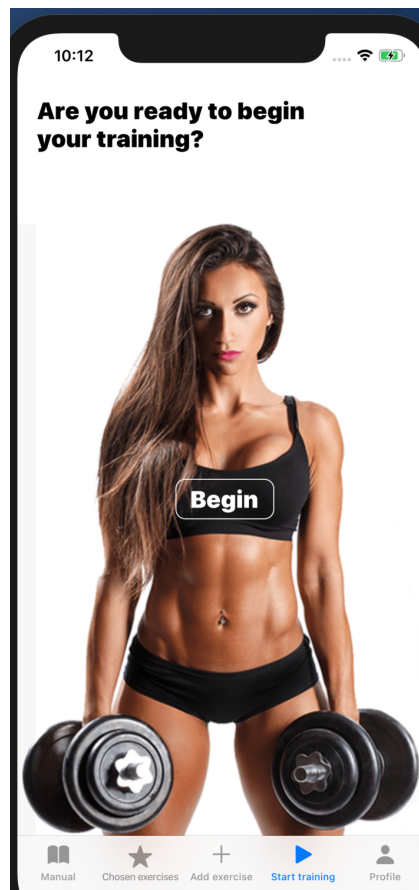


Рисунок 3.5 - Екран початку тренування.

Екран для обирання кількості підходів для вправ має посередині кільце прокрути, а нижче кнопку для підтвердження обраної кількості разів виконання вправи (Рис. 3.6).

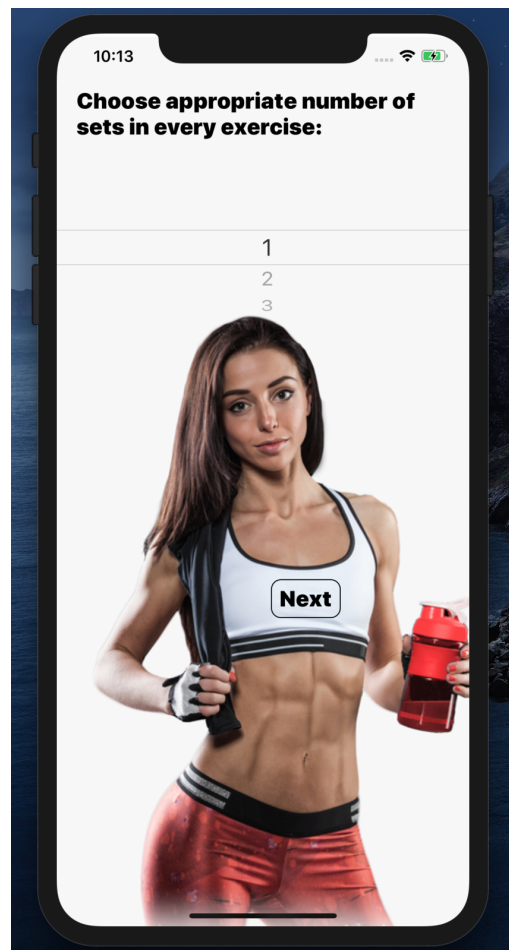


Рисунок 3.6 - Екран для обрання кількості підходів.

Після цього користувачеві буде запропоновано пройти серію екранів, на яких користувач має заповнити кількість разів зроблених у конкретному підході, де буде також назва вправи і номер підходу (Рис. 3.7). Для обрання кількості разів у підході користувач так само має обрати кільцем прокрутки необхідну йому кількість.

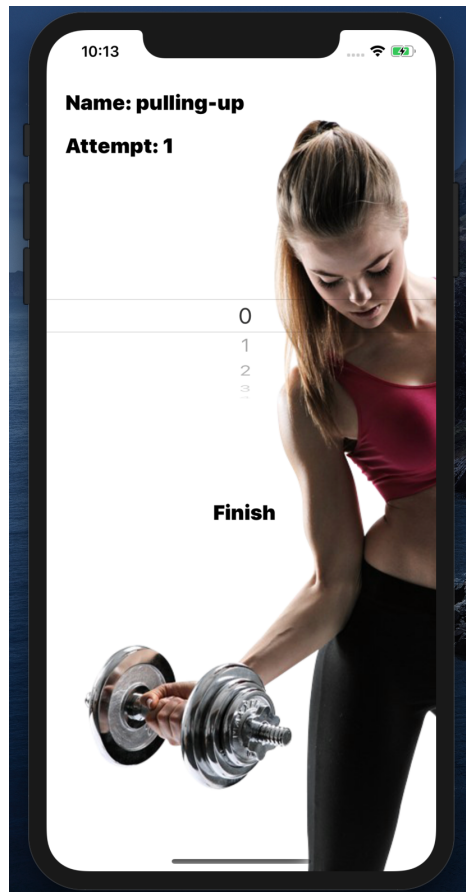


Рисунок 3.7 - Экран для заповнення кількості разів у підході

3.4 Сторінка прогнозування результатів

Існує можливість переглянути прогноз у розгорнутому стані та за різний часовий інтервал(Рис. 3.9).

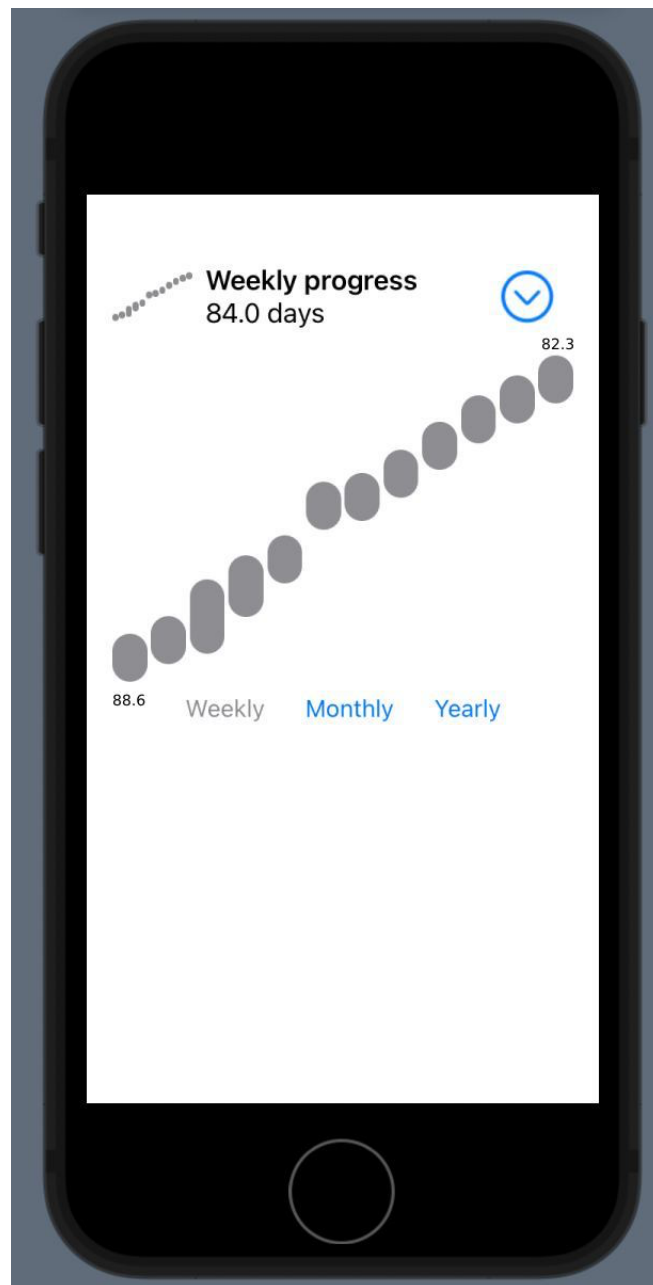


Рисунок 3.9 – Графік прогнозу

ВИСНОВКИ

В результаті виконання кваліфікаційної роботи було виконано дослідження та розробка методу прогнозування параметрів статури людини для системи планування тренувань, а також створено мобільний додаток для збору даних та створення прогнозу за розробленим методом.

В роботі на основі аналізу результатів фактичного змінення параметрів статури людини в залежності від видів тренувальних вправ виявлено вплив окремих видів тренувань на змінення параметрів статури і запропоновано ввести певні вагові коефіцієнти у модель регресії. Проведені обчислювальні експерименти показали достовірність отриманих результатів прогнозу.

Кінцева ціль розробки була досягнена, але не у повному її обсязі, адже на даний момент частина функціоналу знаходиться в стадії розробки. У даному методі є врахування впливу конкретного типу справ на загальний результат схуднення, а мобільний додаток вміє збирати дані за конкретними вправами, робити прогноз за розробленим методом, створювати програму тренувань та отримати довідку за кожною вправою окремо. Також існує власний обліковий запис, да саме й зберігаються усі дані користувача. Також додаток не буде питати користувача о постійному введенні логіну та паролю, бо після першого входу на пристрої логін і пароль зберігаються у захищених користувальницьких даних.

Для підвищення точності розроблений метод потребує доопрацювання у виді введення додаткових параметрів для підвищення його точності. У мобільному додатку велика кількість корисного функціоналу знаходиться ще на стадії розробки, а саме такі як: додавання користувачем особистих фотографій, побудова графіків прогнозу за спортивними зверненнями атлета і ще багато іншого.

До недоліків розробленого методу можливо віднести його точність на поточному етапі. На даний момент розроблений прототип працює на Apple iPhone 11 Pro Max, але адаптованість інтерфейсу вже знаходиться в розробці.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Види та Структура Інформаційних Систем Підприємств. URL: <https://pidru4niki.com/> (дата звернення: 9.11.2021).
2. Класифікація інформаційних систем. URL: https://pidru4niki.com/klasifikatsiya_informatsiynih_sistem (дата звернення: 11.11.2021).
3. Архитектура информационных систем. URL: https://studbooks.net/informatika/arhitektura_informatsionnyh_sistem (дата звернення: 11.11.2021).
4. Чому візуалізація даних – обов’язкова складова бізнес-аналітики. URL: <https://www.microsoft.com/> (дата звернення: 12.11.2021).
5. АНАЛІЗ ДАНИХ. URL: <https://ua.nesrakonk.ru/data-analytics/> (дата звернення: 12.11.2021).
6. Видобуток даних: процес, методи та основні проблеми аналізу даних. URL: <https://uk.myservername.com/data-mining-process> (дата звернення: 9.11.2021).
7. Види аналізу даних. URL: <https://uk.education-wiki.com/> (дата звернення: 12.11.2021).
8. Багаторівневе Інтелектуальне Керування. URL: <https://salessector.ru/> (дата звернення: 19.11.2021).
9. Що таке регресійний аналіз? URL: <https://uk.education-wiki.com/what-is-regression-analysis> (дата звернення: 19.11.2021).
10. Model performance metrics. URL: <http://www.sthda.com> (дата звернення: 20.11.2021).
11. Регресійний аналіз даних приклад. Методи регресійного аналізу. URL: <https://dprvrn.ru/uk/regressionnyi-analiz-dannyh-primer-metody-regressionnogo-analiza-predpolozheniya-lineinoi-regressii/> (дата звернення: 20.11.2021).

12. Однофакторний дисперсійний, кореляційний, регресійний аналіз. URL: <https://tdmuv.com/kafedra/internal/informatika/> (дата звернення: 19.11.2021).
13. Correlation and Regression Analysis. Dr. Mohamed Ahmed Zaid. SESRIC, 2015. с. 13 -20.
14. Understanding Multiple Regression. URL: <https://towardsdatascience.com/understanding-multiple-regression> (дата звернення: 21.11.2021).
15. Ендрю Ф. Сигел. Практическая бизнес-статистика. 2002. 4 издание. стр. 558.
16. Statistics Dictionary. URL: <https://stattrek.com/statistics/dictionary.coefficient-of-multiple-determination> (дата звернення: 22.11.2021).
17. t-критерій Стьюдента. URL: <http://fpo.bsmu.edu.ua/static/t-kryteriy-styudenta> (дата звернення: 21.11.2021).
18. t-критерий Стьюдента для Независимых Совокупностей. URL: <https://medstatistic.ru/> (дата звернення: 22.11.2021).
19. Prediction Versus Reality: The Use of Mathematical Models to Predict Elite Performance in Swimming and Athletics at the Olympic Games. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861753/> (дата звернення: 23.11.2021).
20. 15 TYPES OF REGRESSION IN DATA SCIENCE. URL: <https://www.listendata.com/regression-analysis> (дата звернення:).
21. Результаты Олимпийских Игр. URL: <https://olympteka.ru/> (дата звернення: 24.11.2021).
22. Do you even lift? Predicting weight loss with workout and nutrition. URL: <https://towardsdatascience.com/do-you-even-lift-predicting-weight-change-with-workouts-and-nutrition> (дата звернення: 25.11.2021).

23. A Gentle Introduction to XGBoost for Applied Machine Learning. URL: <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/> (дата звернення: 24.11.2021).
24. Архітектури мобільних додатків. URL: <https://habr.com/ru> (дата звернення: 27.11.2021).
25. List of Apple Processors. URL: https://apple.fandom.com/wiki/List_of_Apple_processors (дата звернення: 27.11.2021).