

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ ЕЛЕКТРОНІКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА КОМП'ЮТЕРНИХ НАУК
СЕКЦІЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ПРОЕКТУВАННЯ

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему: «Інформаційна система прогнозування результатів тенісних матчів»

за спеціальністю 122 «Комп'ютерні науки»,
освітньо-професійна програма «Інформаційні технології проектування»

Виконавець роботи: студент групи ІТ-53-7 Гагін Дмитро Олегович

**Кваліфікаційна робота бакалавра
захищена на засіданні ЕК
з оцінкою**

_____ «___» _____ 2019 р.

Науковий керівник

(підпис)

к.т.н., доц., Парфененко Ю.В.
(науковий ступінь, вчене звання, прізвище та ініціали)

Голова комісії

(підпис)

Шифрін Д. М.
(науковий ступінь, вчене звання, прізвище та ініціали)

Засвідчую, що у цій дипломній роботі немає
запозичень з праць інших авторів
без відповідних посилань.

Студент _____
(підпис)

Суми-2019

Сумський державний університет
Факультет електроніки та інформаційних технологій
Кафедра комп'ютерних наук
Секція інформаційних технологій проектування
Спеціальність 122 «Комп'ютерні науки»
Освітньо-професійна програма «Інформаційні технології проектування»

ЗАТВЕРДЖУЮ

Зав. секцією ІТП

_____ В. В. Шендрик
« ____ » _____ 2019 р.

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

Гагін Дмитро Олегович

1 Тема роботи Інформаційна система прогнозування результатів тенісних матчів

керівник роботи Парфененко Юлія Вікторівна, к.т.н., доцент _____,

затверджені наказом по університету від «17» травня 2019 р. №0834-III

2 Строк подання студентом роботи «3» червня 2019 р.

3 Вхідні дані до роботи Навчальна вибірка з результатами матчів гравців та їх характеристиками

4 Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) 1) Аналіз предметної області

2) Постановка задачі

3) Проектуванні інформаційної системи прогнозування результатів тенісних матчів

4) Розробка інформаційної системи прогнозування результатів тенісних матчів

5 Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

Мета, задачі, актуальність, проаналізовані системи, порівняльна характеристика систем та методів прогнозування, архітектура мережі, засоби реалізації, діаграми варіантів використання, діаграма процесів з декомпозиціями, навчальна вибірка, висновки

6. Консультанти розділів роботи:

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Оформлення планування робіт	До 08.03.2019	
2.	Оформлення технічного завдання	До 15.03.2019	
3.	Проведення аналізу предметної області	До 22.03.2019	
4.	Проведення структурно-функціонально моделювання процесів	До 10.04.2019	
5.	Розробка навчальної вибірки	До 20.04.2019	
6.	Розробка нейромережі	До 05.05.2019	
7.	Розробка додатку	До 11.05.2019	
8.	Тестування системи	До 17.05.2019	
9.	Здача пояснювальної записки та файлів розробленого проекту	До 03.06.2019	

Студент

(підпис)

Гагін Д.О.

Керівник роботи

(підпис)

к.т.н., доц. Парфененко Ю.В.

РЕФЕРАТ

Тема бакалаврської роботи: «Інформаційна система прогнозування результатів тенісних матчів».

Пояснювальна записка містить вступ, 4 розділи, висновки, додатки та список літератури.

В першому розділі наведено огляд існуючих програмних продуктів, актуальність напрямку дослідження. Проведено аналіз сучасного використання інформаційних систем для вирішення задач прогнозування в обраній предметній області. Визначено всі наявні переваги та недоліки на прикладі існуючих аналогів та обґрунтовано актуальність роботи та потреба в створенні власної інформаційної системи.

Другий розділ включає в себе формулювання мети та основних задач роботи, визначено методи дослідження та вибрано засоби за допомогою яких буде проводитись реалізація поставлених задач.

Третій розділ призначений для опису проектування роботи та розробки архітектури додатку. Проведено структурно-функціональне моделювання та моделювання варіантів використання інформаційної системи. Визначено основні етапи проектування роботи.

У четвертому розділі описується процес розробки інформаційної системи. Наводиться архітектура програмного додатку. Приводяться знімки екрану, що демонструють розробку системи та використання продукту.

Результатом проведеної роботи є розроблена інформаційна система для прогнозування результатів тенісних матчів.

Ключові слова: інформаційна система, машинне навчання, теніс, нейромережі, прогнозування в спорті.

Зміст

Вступ.....	6
1. Аналіз предметної області.....	8
1.1 Складання прогнозів на спортивні події.....	8
1.2 Аналіз підходів до прогнозування тенісних матчів.....	10
1.3 Сервіси прогнозування спортивних подій.....	13
2. Постановка задачі та методи дослідження.....	18
2.1 Мета та задача.....	18
2.2 Вибір методів прогнозування.....	19
2.3 Вибір засобів реалізації.....	28
3. Проектуванні інформаційної системи прогнозування результатів тенісних матчів.....	29
3.1 Структурно-функціональне моделювання діяльності інформаційної системи прогнозування.....	29
3.2 Моделювання варіантів використання інформаційної системи прогнозування результатів тенісних матчів.....	33
4. Розробка інформаційної системи прогнозування результатів тенісних матчів.....	36
4.1 Підготовка навчальної вибірки.....	36
4.2 Імпортування даних в систему.....	37
4.3 Розробка та тестування інформаційної системи прогнозування результатів тенісних матчів.....	38
4.4 Розробка web-додатку прогнозування результатів тенісних матчів.....	43
Висновок.....	45
Список використаних джерел.....	46
Додаток А. Технічне завдання.....	47
Додаток Б. Планування робіт.....	52

ВСТУП

У сучасній науці для дослідження процесів різної природи широкого розповсюдження набули методи інтелектуального аналізу даних. З їх допомогою можна вирішувати різноманітні задачі аналізу даних – прогнозування, класифікації, кластеризації тощо.

Для створення прогнозу на певну подію потрібно враховувати вплив багатьох факторів в системі та факторів зовні системи. Також певні параметри можуть впливати не тільки на вихідний результат, а і чинити вплив на інші параметри. Через це встановлюється внутрішній зв'язок який важко прогнозувати використовуючи тільки експертні думки. Непередбачуваність та неочевидність зв'язків в системі та між собою є головною проблемою прогнозування. З розвитком нейромереж стало можливим враховувати всі ці параметри та їх вплив на результат. Виходячи з цього і збільшилась точність прогнозування результатів певних подій [1].

На сьогоднішній день машинне навчання та нейромережі почали використовуватись у всіх галузях життя за рахунок їх точності, зручності та гнучкості в налаштуванні автоматичному підстроюванні під конкретні задачі.

Основною властивістю нейромереж є паралельна обробка інформації одночасно багатьма нейронами. Завдяки цьому досягається збільшення швидкості обробки інформації. Іншою особливістю нейромереж є здатність до навчання і узагальнення інформації. Завдяки цьому досягається схожість з роботою головного мозку людини.

Метою даної роботи є розробка інформаційної системи прогнозування результатів тенісних матчів. Об'єктом роботи є процес прогнозування результатів спортивних подій. Предметом дослідження є методи та моделі прогнозування результатів матчів.

Для досягнення цієї мети треба вирішити наступні задачі:

- провести аналіз предметної області;
- провести огляд існуючих програмних засобів прогнозування;
- визначити вимоги до інформаційної системи та виконати планування робіт;
- розробити модель прогнозування тенісних матчів;
- спроектувати інформаційну систему прогнозування результатів на основі зібраних даних;
- реалізувати інформаційну систему та провести її тестування.

Створена інформаційна система корисна для спортсменів за рахунок аналізу їх можливостей та можливості керуванням їх тренувального плану та букмекерським конторам для прогнозування результатів матчів.

1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Складання прогнозів на спортивні події

Великий шанс на грошовий прибуток і інтерес в науці зумовили збільшення досліджень алгоритмів прогнозування тенісних матчів. Здобуття балів тенісу являє ієрархічну структуру: матч складається з декількох сетів, що в свою чергу складаються з геймів, на яких гравці заробляють бали – очки для виграшу. В сучасних підходах до прогнозування ця структура використовується щоб отримати ієрархічні вирази ймовірності перемоги гравця на основі марківських ланцюгів [2]. Якщо вважати, що бали в тенісі розподіляються незалежно і однаково то для отримання виразу потрібно знати ймовірність виграшу кожним гравцем балу при подачі. На основі цієї базової статистики, яку можна отримати з історичних даних в інтернеті, обчислюється ймовірність виграшу кожним гравцем гейму, потім сету і, нарешті, матчу.

При всіх перевагах такого підходу, він не може бути визнаний ідеальним. Представляючи якості гравців тільки по одному такий метод не здатен враховувати більш тонкі чинники, які також впливають на результат матчу.

Беручи до уваги величезну кількість історичних даних з тенісу, можна запропонувати альтернативний підхід до прогнозування тенісних матчів - машинне навчання. Параметри гравців і матчу разом з результатом матчу можуть скласти навчальну вибірку.

На результат матчу впливають фактори які можна поділити на такі групи:

- фактори впливу на гравця(стан здоров'я гравця,настрій та поведінка гравця, якість екіпіровки, кількість виграних очків в останніх іграх,відсоток програшних подач);
- фактори впливу на гру(погодні чинники, судійство, випадкові чинники,що можуть порушити гру).

Прогнозування матчів дає позитивний вплив не тільки на букмекерські контори та бізнес ставок, але також ці данні приносять користь в сам спорт, а саме для тренерів та самих спортсменів, бо за допомогою наперед визначених результатів можна визначити слабкі місця та проблемні зони гравця у кількісних значеннях, що дасть змогу прибрати їх до нової гри, або визначити мету якої потрібно досягнути в змаганнях та визначити стратегію за допомогою якої досягнути цю мету. Вивчення структури підготовленості спортсменів в плані вдосконалення підготовки спортивного резерву дозволяє оптимізувати процес відбору та орієнтації спортсменів на основі розробки модельних характеристик, що розширює можливості реалізації індивідуального підходу до планування та корекції тренувального процесу, а також прогнозування спортивного результату.

Основними типами прогнозування спортивних подій по часу є короткострокове, середньострокове прогнозування, довгострокове і наддовге прогнозування.

Короткострокове прогнозування пов'язане з вирішенням задач, що ставляться в ході одного або декількох тренувальних занять.

Середньострокове прогнозування пов'язане з визначенням темпів розвитку тренуваності за рахунок засобів для тренування і методів на окремих етапах. Довгострокове прогнозування направлене на оптимізацію процесу спортивного відбору, підготовки і участі в змаганнях протягом відносного тривалого часу від 1 до 4 років. Наддовгострокове прогнозування направлене на виявлення основних тенденцій розвитку спорту в світі [3].

1.2 Аналіз підходів до прогнозування тенісних матчів

Прогнозування це процес передбачення майбутнього стану предмета чи явища на основі аналізу його минулого і сучасного, систематично оцінювана інформація про якісні й кількісні характеристики розвитку обраного предмета чи явища в перспективі [4]. Результатом прогнозування є прогноз — знання про майбутнє і про ймовірний розвиток сьогочасних тенденцій конкретного явища-об'єкту в подальшому існуванні.

Для знаходження результатів в спорті існує велика кількість підходів та методів, що засновані на різних методиках до вирішення задачі.

Одним з головних підходів до прогнозування є статистичні методи. Вони базуються на зборі та обробці інформації про матчі та знайдення результату на основі експертної оцінки. Цей метод не є точним бо не враховуються багато чинників впливу, а саме прогнозування є інтуїтивне.

За допомогою методів теорії ймовірності та статистики з точністю прогнозу до 55% – 60%, можна передбачити той чи інший результат спортивного змагання, приміром, матчу. Оскільки точність прогнозування мала і не задовольняє потреб букмекерів та самих тренерів, доцільніше використовувати інші методи прогнозування.

Одним з таких методів є машинне навчання – розділ штучного інтелекту, який застосовує статистичні прийоми для надання комп'ютерам здатності «навчатися» (поступово покращувати продуктивність у певній задачі) з даних, без того, щоби бути програмованими явно.

Існує два типи машинного навчання: індуктивне навчання і дедуктивне навчання.

Індуктивне навчання поділяють на чотири основних типи:

- навчання з учителем;
- навчання без учителя;
- навчання з прихованими шарами(з підкріпленням);

– глибоке навчання.

Навчання з вчителем застосовується і добре працює з великими обсягами даних. Беручи інформацію з певними мітками, що означають приналежність елементу до певного класу машина повинна визначити ознаки які підходять до класу та на основі цих даних класифікувати нові отримані дані.

Вчителем виступає особа, що встановила відповідні маркери – мітки. В подальшому знайдений машиною алгоритм може бути швидко переналаштований на рішення будь-якої задачі.

Основне завдання машини при навчанні без вчителя виявити закономірності, знайти зв'язок між окремими даними, упорядкувати дані та підібрати шаблони, виконати класифікацію даних. Навчання без вчителя використовується в рекомендаційних системах – при аналізі попередніх дій користувачеві пропонуються можливості, які можуть зацікавити його з більшою ймовірністю, ніж інші.

Навчання з прихованими шарами використовується для вирішення складних завдань. Це навчання є контрольованим, а в ролі вчителя в ньому виступає середовище. Між машиною і середовищем утворюється система зі зворотним зв'язком. Мережа не має будь-якої інформації про середовище наперед, але вона може здійснювати в ній будь-які дії. Середовище реагує на ці дії і надає машині дані, які дозволяють їй вчитися і реагувати на них [5].

Таке навчання використовується в системах навігації для роботів. Вони навчаються уникати зіткнень отримуючи зворотний зв'язок при кожній аварії і набираються досвіду і самонавчаються. Використовується таке навчання в логістиці, при плануванні завдань чи складанні графіків або при навчанні машини логічним іграм.

В машинному навчанні існує велика кількість алгоритмів за допомогою яких можна навчати систему. Основними з них є дискримінантний аналіз, байєсовські класифікатори мережі з радіально-базисними шарами, нейромережі на основі перцептрону.

Мережа має певну кількість нейронів в залежності від цілі. Деяка кількість нейронів формує собою шар нейронів мережі, яких може бути декілька. Нейрони першого рівня напряму взаємозв'язані з користувачем, вони отримують дані ззовні і після обробки передають сигнали на наступний рівень нейронів. Нейрони на другому рівні формують прихований шар. Кожен нейрон прихованого шару пов'язаний з декількома нейронами вхідного та вихідного шару. Така архітектура здатна до навчання і може знаходити прості взаємозв'язки в даних. Після обробки даних прихованим шаром сигнали передаються на вихідний шар нейронів звідки видається відповідь для конкретної задачі.

Для збільшення точності результату та покращення роботи самої системи може використовуватись глибоке навчання. Воно має декілька прихованих рівнів та поєднує алгоритми навчання з вчителем та без вчителя.

При такому навчанні рівні нейронів можуть чергуватись з шарами, що виконують складні логічні перетворення. Використовуючи таку систему навчання кожен рівень мережі шукає взаємозв'язки в попередньому рівні мережі. Така мережа здатна знаходити не тільки прості взаємозв'язки, а й взаємозв'язки між взаємозв'язками [5].

Прогнозування за допомогою методів машинного навчання має більшу точність за рахунок навчання і можливості модифікації.

В сучасному спорті для аналізу працездатності спортсменів та аналізу їх роботи використовують машинне навчання. Тренери отримують кількісні значення показників підготовленості та можуть редагувати тренувальний процес для досягнення певних цілей.

1.3 Сервіси прогнозування спортивних подій

Прогнозування ставок почало відігравати велику роль в зв'язку з популяризацією букмекерських контор і процесу ставок на спортивні змагання. Основними лідерами в сервісах, що дають можливість для ставок і прогнозування є OhMyBet, MyScore, LiveScore, PariMatch.

OhMyBet! задуманий як сервіс з математично обґрунтованими прогнозами на теніс, дає які не людина, а машина - модель штучного інтелекту. За допомогою алгоритмів машинного навчання було проаналізовано велику базу даних зі статистикою минулих матчів і побудована прогностична модель, яка здатна передбачати переможця майбутнього матчу. Алгоритм самонавчальний: чим більше історичних даних (зіграних матчів) він аналізує, тим вище точність моделі. Проаналізовано більш 825 тисяч минулих матчів турнірів Міжнародної федерації тенісу та Асоціації тенісистів-професіоналів з 2000-го по 2014 рік. Перевіряли модель на матчах 2015 року. Результат - точність передбачення переможця 85%
Інтерфейс сервісу показаний на рис. 1.1 та 1.2.

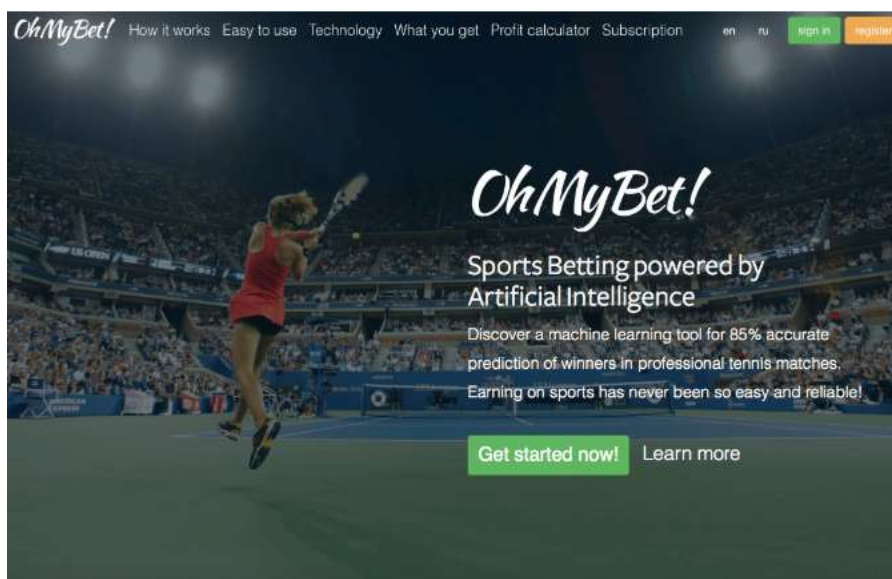


Рисунок 1.1 – Головна сторінка сервісу OhMyBet!

The screenshot displays the OhMyBet! website interface. At the top, there are several statistics: 'Upcoming matches' (2 today), 'Tracked' (3 in progress), 'Average winning odds' (1.73 in total prediction histo...), 'Total placed' (691 since Jan. 2015), 'Max 1-bet profit' (3 470 from a \$1000 bet), and 'Total profit' (30 400 from a \$1000 bet). Below these are tabs for 'Upcoming', 'Tracked', and 'History'. A table lists various tennis tournaments with columns for Date, Tournament, Winner, Loser, Odds, Yield, and Tracked. The table includes search filters for player names and tournaments. A pagination bar at the bottom shows 'First', '<', '0' of 70, '>', and 'Last'.

Date	Tournament	Winner	Loser	Odds	Yield	Tracked
Feb 13, 2017	Memphis Open - Memphis	Benjamin Becker	Nikola Pietrangeli	1.50	\$500	Not tracked
Feb 13, 2017	Argentina Open - Buenos Aires	Rogério Dutra Da Silva	Federico Coria	1.48	\$480	Not tracked
Feb 6, 2017	Budapest Challenger	Jürgen Melzer	Edward Corrie	1.51	\$510	Not tracked
Feb 6, 2017	Budapest Challenger	Edward Corrie	Yannick Maden	2.07	\$1070	Not tracked
Feb 6, 2017	Launceston Challenger	Riccardo Bellotti	Max Purcell	1.80	\$800	Not tracked
Feb 6, 2017	Launceston Challenger	Daniel Nguyen	Marc Polmans	2.45	\$1450	Not tracked
Feb 6, 2017	Budapest Challenger	Gianluigi Quirzi	Attila Balazs	2.15	\$0	Not tracked
Feb 6, 2017	Open Sud de France - Montpellier	Malek Jaziri	Tristan Lamasine	1.56	\$560	Not tracked
Feb 6, 2017	Open Sud de France - Montpellier	Pierre-Hugues Herbert	Calvin Hemery	1.71	\$710	Not tracked
Feb 6, 2017	Launceston Challenger	Tsung-Hua Yang	Maverick Banes	3.07	\$3070	Not tracked

Рисунок 1.2 – інтерфейс сервісу OhMyBet!

Особливість системи в тому, що вона дає дуже обмежену кількість прогнозів: один-два в день, іноді жодного. Модель відбирає тільки матчі з максимальною вірогідністю прогнозу і високими коефіцієнтами на передбаченого переможця - це знижує ризик програшу і забезпечує високу прибутковість з кожної ставки. Після реєстрації користувач отримує сім днів пробного періоду, протягом яких він отримує доступ до нових прогнозів та історії прогнозів з 2015 року. Є різні варіанти підписки: на 1, 3, 6 місяців.

Інший не менш популярний сервіс для прогнозування та ставок MyScore, на якому міститься інформація по 32 видам спорту. Результати спортивних подій оновлюються в режимі реального часу. Користувач сервісу може оперативно знаходити необхідні його події по країнам і лігам. До його послуг - результати минулих ігор, розклад майбутніх матчів, а також рух коефіцієнтів в букмекерських конторах. В папку «Мої ігри» можна розміщувати окремі події - для отримання

повідомлень за цими іграм. Сайт MyScore дуже зручний для гравців, що роблять ставки в режимі Live - завдяки наявності текстових і відео трансляцій в поєднанні з докладними статистичними даними. Також є наявність повнофункціональної мобільної версії сайту і додатків для iOS і Android.

Інтерфейс сервісу показаний на рис. 1.3



Рисунок 1.3 – інтерфейс сервісу MyScore

LiveScore це популярний сервіс, що не володіє дуже широким функціоналом, але є цілком інформативним. У меню представлені всього 10 спортивних дисциплін (футбол, хокей, теніс, баскетбол, бейсбол, волейбол, американський футбол, гандбол, регбі-юніон і крикет), проте відомості по локальних чемпіонатів дуже глибокі. Тут можна ознайомитися з котируваннями букмекерів і рухом ліній від сервісу OddsPortal, а при необхідності - відразу перейти на сайт обраної букмекерської контори. Передбачена можливість додавання подій в «Вибране». Головна сторінка сервісу показана на рис. 1.4.



Рисунок 1.4 – інтерфейс сервісу LiveScore

Букмекерська контора "Парі-Матч» входить в число лідерів пострадянського ігрового ринку. Компанія має багаторічний досвід роботи з онлайн-ставками, розвинену мережу стаціонарних офісів і сайт з інтуїтивно зрозумілим дизайном. Є можливість віддаленої ідентифікації клієнта і компетентна служба підтримки.

З мінусів клієнти компанії відзначають середні коефіцієнти і недостатньо опрацьовану, в порівнянні з іншими конторами, підпис на найбільші турніри Північної Америки. Ще один мінус контори - відсутність додатків для мобільних гаджетів. Букмекерська контора Парі-Матч відноситься до компаній з середніми коефіцієнтами і невеликою маржею. Маржа на топові спортивні події може скласти 2-5 відсотка, на менш імениті - 5-7 відсотків. Букмекерська лінія Парі-Матч включає всі основні чемпіонати і змагання. Тут можна робити ставки на футбол, баскетбол, теніс, значущі події в світі шоу-бізнесу і кіберспортивні заходи. Головна сторінка сервісу показана на рис.1.5.

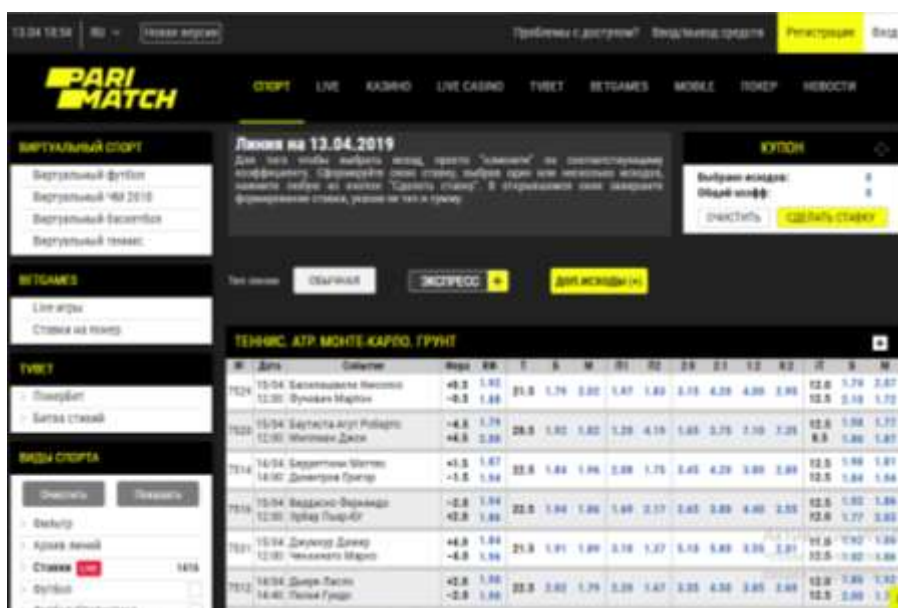


Рисунок 1.5 – інтерфейс сервісу Parimatch

Порівняння основних характеристик сервісів для прогнозування показано на таблиці 1.1.

Таблиця 1.1 – порівняльна характеристика сервісів прогнозування

	OhMyBet!	MyScore	LiveScore	Parimatch
Точність прогнозування	85%	80%	70%	65%
Мобільний додаток	+	+	–	–
Зручність	+	+	+	+/-
Велика кількість ставок	+	+	+	+
Наявність великої бази даних для прогнозів	+	+	–	–
Велика вартість прогнозування	+	+	+/-	+/-
Швидкість прогнозування	–	+/-	–	–

Отже для конкурентоспроможності потрібно створити інформаційну систему для прогнозування результатів матчів з точністю більше 70% та великою базою даних для більш кращого прогнозування. Ця система буде корисна спортивним тренерам для регулювання тренувальних навантажень аби досягнути певної поставленої мети.

2. ПОСТАНОВКА ЗАДАЧІ ТА МЕТОДИ ДОСЛІДЖЕННЯ

2.1 Мета та задача

Інформаційна система повинна забезпечувати автоматизацію роботи прогнозування, а саме прийом даних по гравцям або матчам, обробку цих даних та занесення їх до загальної бази даних та сам процес прогнозування із зручним виводом інформації користувачу.

В ролі вхідних даних для інформаційної системи виступають характеристики гравців та їхня статистика по останнім зіграним матчам, а на виході отримаємо значення результату матчу стосовно одного гравця – виграш , програш або нічия.

Використання інформаційної системи спрямоване на зменшення роботи та підвищення продуктивності праці тренерського складу спортсменів та забезпечення достовірної інформації щодо статистики матчів для букмекерських контор та інших закладів.

Впровадження інформаційної системи передбачає такі результати:

- введення і зберігання інформації по гравцям та матчам, що проводилися та щодо їх характеристик;
- редагування інформації про гравців та матчів;
- пошук інформації по певним користувачами критеріям;
- отримання вірного прогнозу по матчу для конкретного замовника.

Також було сформоване та затверджене технічне завдання, яке було розміщене у додатку А.

1.4 Вибір методів прогнозування

Одним з методів є розрахунок за статистичними даними заснований на формулі Байеса. Вона має вигляд:

$$P_a(H_i) = \frac{P(A|H_i) * P(H_i|A)}{\sum_{i=1}^n P(H_i) * P(H_i|A)}$$

Розглянемо працездатність на реальному прикладі. Для цього візьмемо будь-який матч з будь-якого чемпіонату з тенісу.

Контрольним турніром буде матч Вімблдону 2018 року між Новаком Джоковичем та Кевіном Андерсоном. Для зрозумілості позначимо подію(матч відбувся) як А, Н1 – твердження про те що переможе 1 гравець, Н2 – твердження що переможе другий гравець.

Знайдемо ймовірність кожної події, а саме $P(H1) = P(H2) = 1/2$. Тобто з шансом 1/2 виграє один з гравців.

Введемо параметр статистичної ймовірності $P_{H1}(A)$ та $P_{H2}(A)$ – відношення кількості виграних матчів першого гравця до загальної кількості матчів. $P(A)$ - повна ймовірність настання події;

Метою даної задачі є спрогнозувати перемогу першого гравця – $P_a(H1)$. Для цього, скориставшись статистичною інформацією про ігри за попередній сезон, потрібно знайти $P_{H1}(A)$, $P_{H2}(A)$:

$$P_{H1}(A) = 13 \text{ перемог} / 21 \text{ (загальна кількість матчів першого гравця);}$$

$$P_{H2}(A) = 5 \text{ перемог} / 21 \text{ (загальна кількість матчів другого гравця);}$$

відповідно:

$$P_{H1}(A) = 13/21=0,61904;$$

$$P_{H2}(A) = 5/21= 0,23809;$$

Потім знайдемо $P(A)$:

$$P(A) = P(H1) * P_{H1}(A) + P(H2) * P_{H2}(A)$$

$$P(A) = 0,465079$$

Тепер можемо знайти $P_a(H1)$:

$$P_a(H1) = P(H1) * P_{H1}(A) / P(A)$$

$$P_a(H1) = 0,678261 \text{ (або приблизно 68\%)}$$

Аналогічно знайдемо $P_a(H2) = 32\%$

Отже, 68% того, що Новак Джокович переможе Кевіна Андерсона 32% - Кевіна Андерсона переможе Новака Джоковича .

Точність такого прогнозування становить приблизно 50% – 55% , що не задовольняє поставленій задачі [14].

Для прогнозування результатів тенісних матчів виходячи з недоліків і переваг основних методів аналізу було прийнято рішення використовувати машинне навчання за допомогою нейромереж.

Один з методів машинного навчання це використання нейромережі на основі перцептрону.

Перцептрон – найпростіший вид нейронних мереж. В основі лежить математична модель сприйняття інформації мозком, що складається з сенсорів, асоціативних і реагуючих елементів [6].

У найзагальнішому своєму вигляді він представляє систему з елементів трьох різних типів: сенсорів, асоціативних елементів і реагуючих елементів.

Першими в роботу включаються S-елементи. Вони можуть перебувати або в стані спокою (сигнал дорівнює 0), або в стані збудження (сигнал дорівнює 1). Схематично робота перцептрона показана на рис. 2.2.

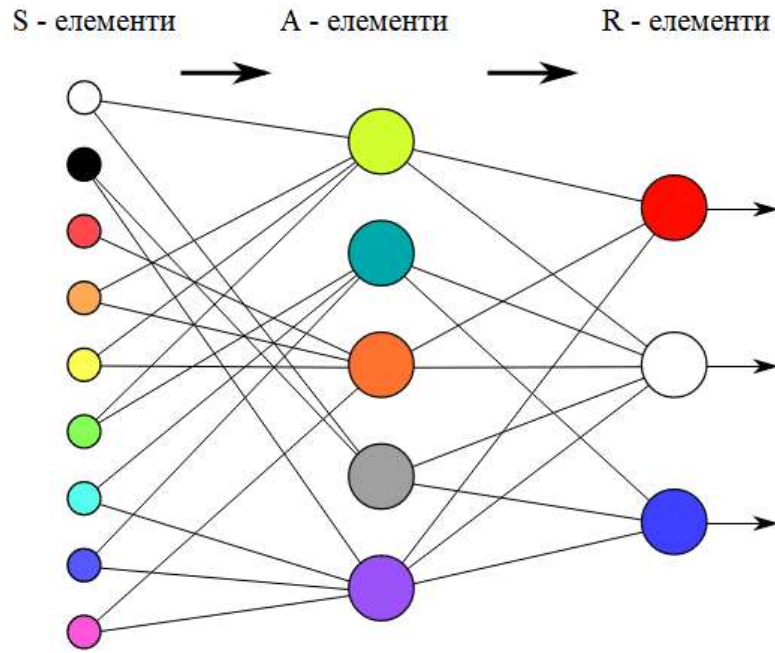


Рисунок 2.2 – схема роботи перцептрона

В ході роботи був проведений розрахунок на навчальній вибірці з 4 параметрів. Результат показаний на рис. 2.3.

В ході моделювання були знайдені недоліки роботи з перцептроном, а саме неточність у прогнозуванні у випадках з малими навчальними вибірками. Точність результатів приблизно 50%.

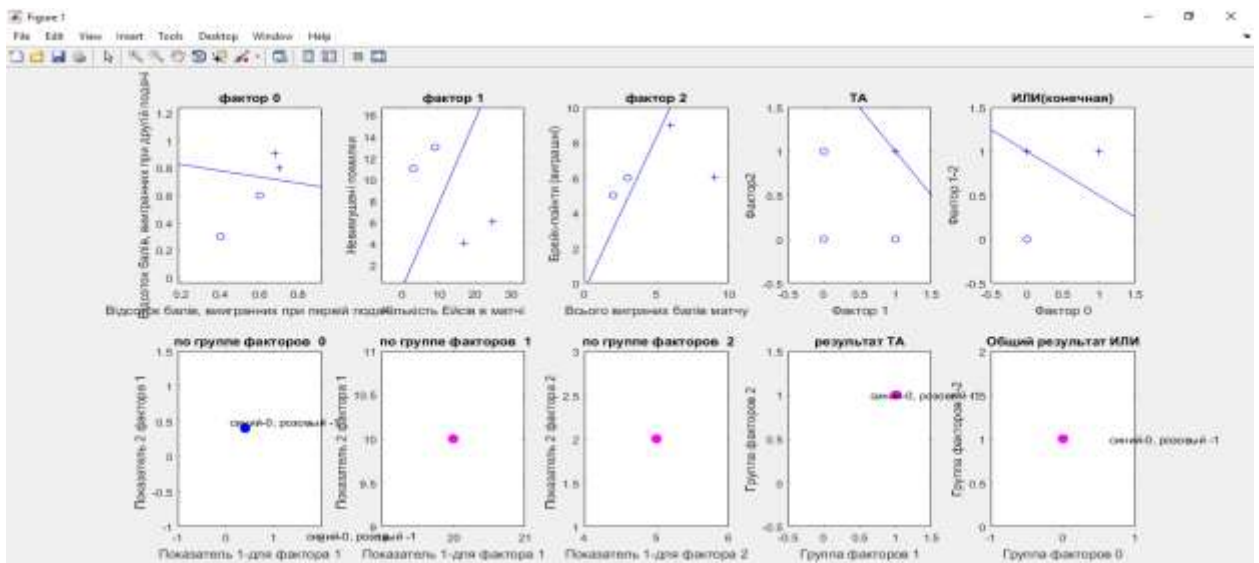


Рисунок 2.3 – моделювання мережі за допомогою перцептрону

Перевагами персептрону є його простота у створенні,але недоліком є мала точність при малих навчальних вибірках.

PNN (probabilistic neural network) призначені для вирішення вірогідностних задач (задач класифікації). Архітектура мережі PNN базується на архітектурі радіальної базисної мережі, але в якості другого шару використовує так званий конкуруючий шар, який підраховує ймовірність приналежності вхідного вектора до того чи іншого класу і в кінцевому рахунку зіставляє вектор з тим класом, ймовірність приналежності до якого вище. Структура мережі PNN представлена на рис. 2.4.

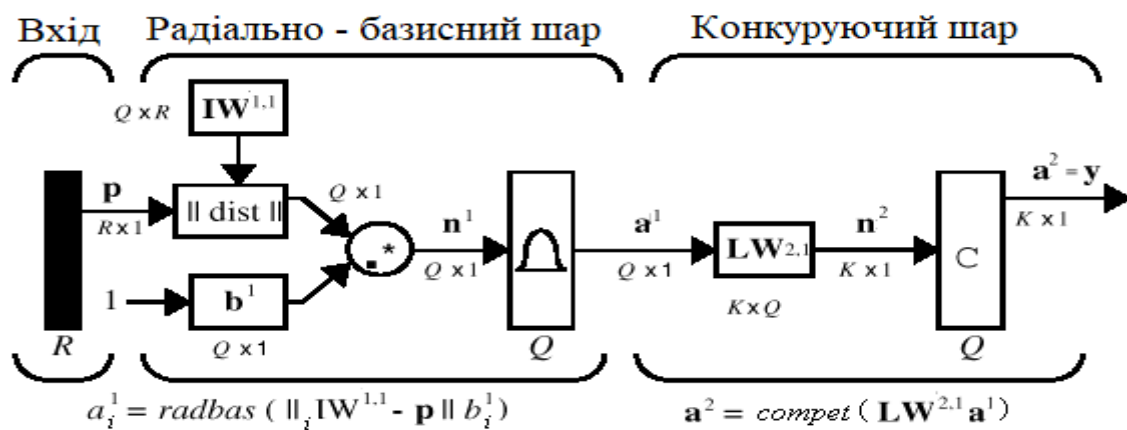


Рисунок 2.4 – структура мережі PNN

На основі навчальної вибірки що складалась з 4 параметрів гравців і 5 класів (великий виграш, виграш ,нічия,програш,великий програш) була створена мережа PNN та проаналізована її робота та її точність. Робота мережі показана на рис.2.5.

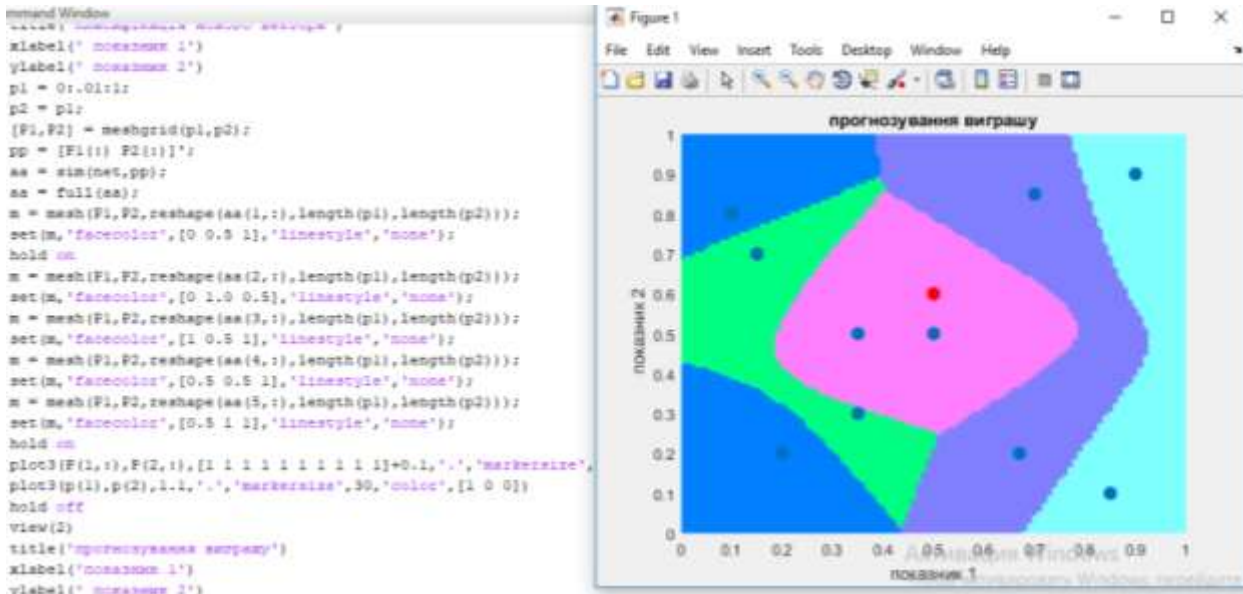


Рисунок 2.5 – робота мережі PNN

Перевагами моделі PNN є її точність та функціональність за рахунок великої кількості настановлених параметрів. Недоліком мережі є проблемні точки на гранях кожного класу. Точність моделі приблизно 70 % – 75%.

Наївний баєсівський класифікатор використовує теорему Баєса для визначення ймовірності приналежності кожного елемента вибірки до одного з класів C за умови того, що залежні змінні приймають задані значення : $P(C | F_1, \dots, F_n)$.

Тобто використовуючи наївний баєсівський класифікатор ми отримаємо результат приналежності до певного класу, який являтиме собою 1 – якщо елемент вибірки на 100% належить класу або вектор, компоненти якого є ймовірностями приналежності до того чи іншого класу якщо елемент на певний відсоток належить одному класу і на інший відсоток другому класу. [15]

Було проведено моделювання мережі і визначена її точність що дорівнює приблизно 75%. Результат на рис. 2.6.

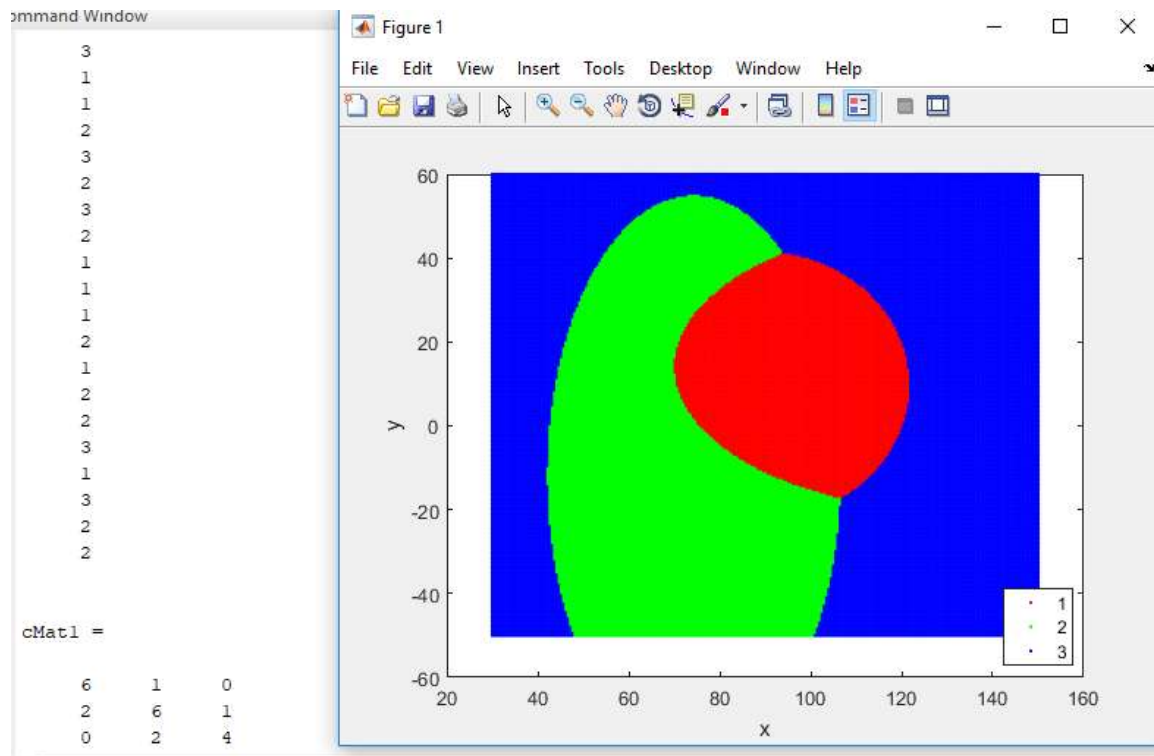


Рисунок 2.6 – робота Баєсівського класифікатора

Перевагою цього підходу є те, що вимоги до розміру вибірки скорочуються від експоненційних до лінійних. Недоліком те, що модель показує точний результат лише у випадку, коли виконується припущення про незалежність. В іншому випадку обчислені ймовірності вже не є точними.

Нейронна мережа Елмана - один з видів рекурентної мережі, яка так само як і мережа Джордана виходить з багатошарового перцептрона введенням зворотних зв'язків, тільки зв'язку йдуть не від виходу мережі, а від виходів внутрішніх нейронів. Це дозволяє врахувати передісторію процесів за якими іде спостереження і накопичити інформацію для вироблення правильної стратегії управління.

Ці мережі можуть застосовуватися в системах управління рухомими об'єктами, так як їх головною особливістю є запам'ятовування послідовностей [7]. Мережа має хорошу точність в 80% ,але складна в реалізації. Її робота показана на рис. 2.7.

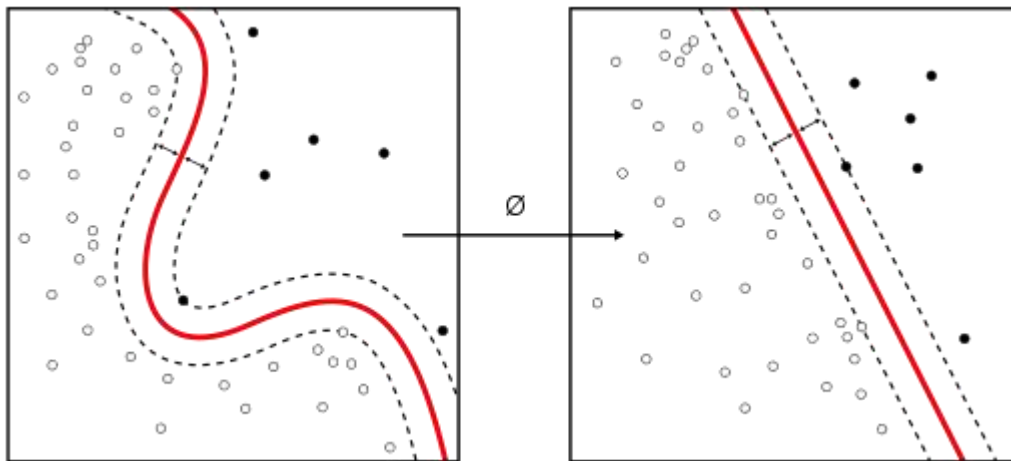


Рисунок 2.7 – моделювання мережі Елмана

При моделюванні часових рядів часто використовується модель NARX — це нелінійна авторегресійна модель, яка має екзогенні входи.

Система на основі цього методу порівнює поточне значення часового ряду з минулим значенням того самого ряду, а також з поточним та минулим значенням ряду.

Модель включає в себе елемент похибки, який відповідає тому фактові, що знання інших елементів не дає можливості передбачувати поточне значення часового ряду точно. Таку модель може бути сформульовано алгебраїчно як

$$y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \dots, u_t, u_{t-1}, u_{t-2}, u_{t-3}, \dots) + \varepsilon_t \quad (2.1)$$

Тут Y_t є цільовою змінною, а U_t є змінною, що визначається ззовні. У цій схемі інформація про u допомагає передбачувати y , як це роблять і попередні значення самої y . ε_t тут є членом похибки (шум). Наприклад, y може бути температурою повітря опівдні, а u може бути номером дня в межах року.

Функція F — це деяка нелінійна функція, наприклад, поліномна. F може бути нейронною мережею, сигмоїдною мережею тощо. [8].

Мережа NARX добре працює з великими навчальними вибірками та дуже добре прогнозує часові ряди. Точність даної мережі приблизно 85% – 90%. Моделювання показане на рис. 2.8.

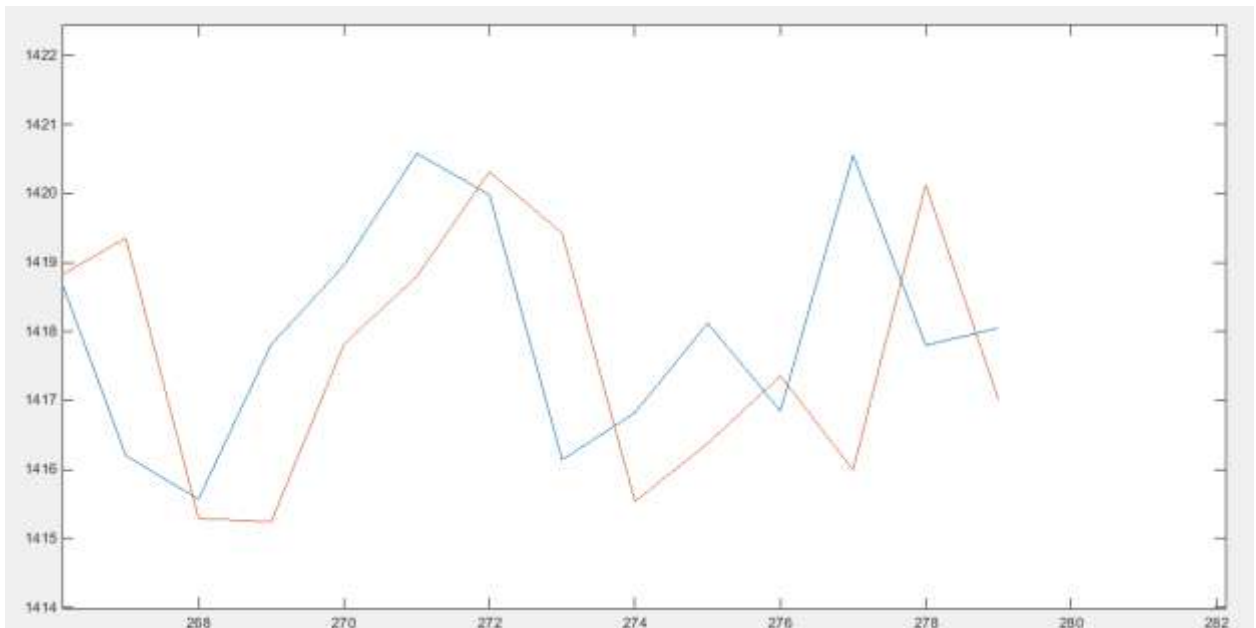


Рисунок 2.8 – прогнозування NARX

Недоліком мережі є погана точність та швидкість прогнозування з навчальними вибірками у яких багато параметрів.

Порівняльна характеристика мереж показана в таблиці 2.1.

Таблиця 2.1 – порівняльна характеристика мереж

Назва мережі	Точність мережі
Розрахунок за статистичними даними	50 – 55%
Нейромережа на основі персептрону	50 – 60%
PNN - мережа	70%
Мережа на основі Баєсівського класифікатора	75%
Мережа Елмана	80%
NARX- мережа	85%
Багатошаровий персептрон	85%

Виходячи з аналізу всіх мереж прийнято рішення використовувати мережу на основі багатошарового персептрону. NARX – мережа більше підходить для прогнозування часових рядів з невеликою кількістю впливаючих на результат параметрів. Вона має найбільшу точність та найкраще підходить для задачі прогнозування. Також бібліотеки з цією мережею є в широкому доступі, що полегшить роботу з нею у всіх середовищах.

2.3 Вибір засобів реалізації

Для створення нейронної мережі існує багато мов та середовищ програмування. В ході аналізу засобів реалізації було розглянуто основні мови та додатки.

Програмування на мові C зручне та легке в розумінні. Середовище Visual Studio має зручний інтерфейс та наявність великої кількості вбудованих бібліотек. Недоліками цієї мови для програмування нейромережі є потрібність підключення нових бібліотек та платна версія програми для програмування.

Середовище для програмування MATLAB має зрозумілий інтерфейс та вже вбудовані бібліотеки, що знадобляться для роботи з нейромережами всіх видів. Середовище є безкоштовне та невелике за розміром. MATLAB має багато утиліт, що можуть спростити навчання та тестування нейромережі. Мова для створення мереж зрозуміла та не потребує окремих навичок. Середовище має вбудований опис всіх команд та приклади застосування основних мереж.

Мова програмування PHP полегшує роботу в програмуванні за рахунок невеликого розміру коду потрібного для реалізації поставленої задачі та зрозумілості його в подальшому для реалізації під інші задачі.

Для створення інформаційної системи прогнозування тенісних матчів вирішено використовувати середовище MATLAB для зрозумілості роботи системи, а для створення зручного додатку буде використано програмування на мові PHP.

3 ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ТЕНІСНИХ МАТЧІВ

3.1 Структурно-функціональне моделювання діяльності інформаційної системи прогнозування

Для моделювання структури та функціоналу певного процесу використовуються методи на основ структурного та об'єктно-орієнтованого підходу до моделювання. Основним методом моделювання цих процесів є функціональне модулювання SADT (IDEF0).

SADT-метод є класичним методом процесного підходу до управління. Він працює за принципом структуризації діяльності процесу відповідно до його внутрішніх підпроцесів [10].

Структурно-функціональне моделювання це певні правила і процедури, за допомогою яких декомпозиується основний процес на більш менші його частини і завдяки цьому з'являється можливість графічного відображення алгоритму управління та представлення послідовності етапів модельованого процесу та виділити проблемні ділянки [11].

За допомогою IDEF0 – моделі можна отримати специфікацію усіх дій, які відбуваються у процесі і характер взаємозв'язку між ними. Створити модель вирішено в програмному продукті Erwin Process Modeler.

Контекстну діаграму головного процесу інформаційної системи прогнозування показано на рис 3.1.

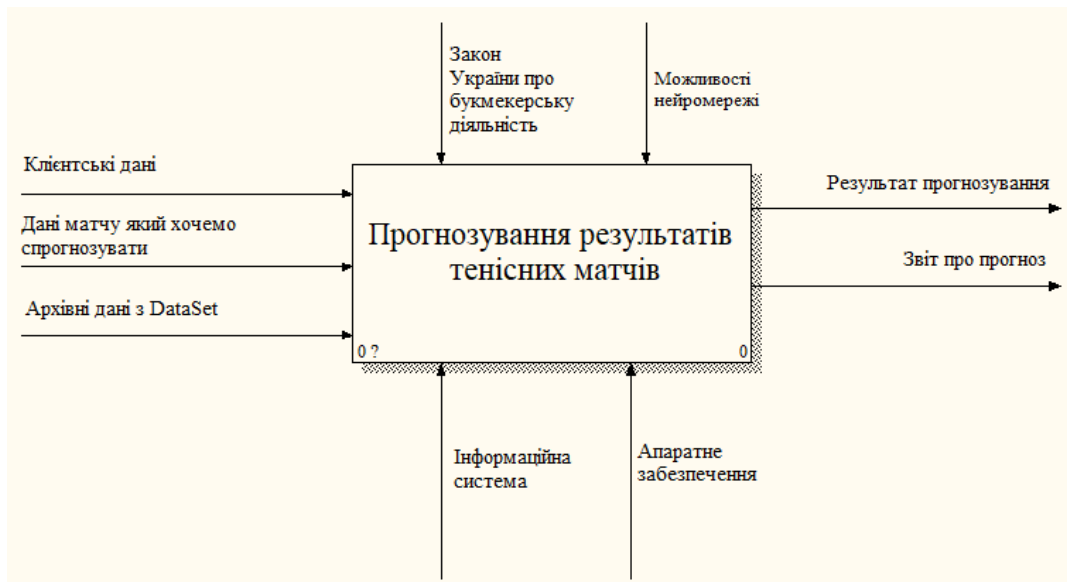


Рисунок 3.1 – контекстна діаграма процесу прогнозування

В центрі діаграми знаходиться основний блок діяльності. Для нашої задачі це прогнозування результатів тенісних матчів, вона подається у вигляді дієслова. Ця дія показує що протягом певного часу та використовуючи певні ресурси буде виконуватись ця головна задача і по досягненні її мети буде отримано вихідна дані.

В ролі вхідних даних є:

- клієнтські дані;
- дані матчу який хочемо спрогнозувати;
- Архівні дані з Dataset

Основними механізмами за допомогою яких буде виконуватись процес є сама інформаційна система, що виконує прогнозування виходячи з даних та апаратне забезпечення на якому та за допомогою якого виконуються обчислення.

Прогнозування спортивних подій керується Законом України про букмекерську діяльність та можливостями самої нейромережі котру було вибрано для системи.

Виконавши аналіз та прогнозування інформаційна система видасть результат у вигляді прогнозу виграшу або нічиєї відносно першого гравця.

Контекстна діаграма декомпозується на діаграми нижчого рівня для того аби показати внутрішні процеси, що проходять в ній. Всі дуги та зв'язки з батьківської діаграми переносяться на діаграму декомпозиції.

Головна батьківська діаграма декомпозується на наступні блоки , вона представлена на рис. 3.2.

- введення даних для прогнозування та клієнтських даних в ІС;
- обчислення прогнозу інформаційною системою;
- занесення результату прогнозування та клієнтських даних в базу даних;
- формування звіту щодо результату прогнозування.

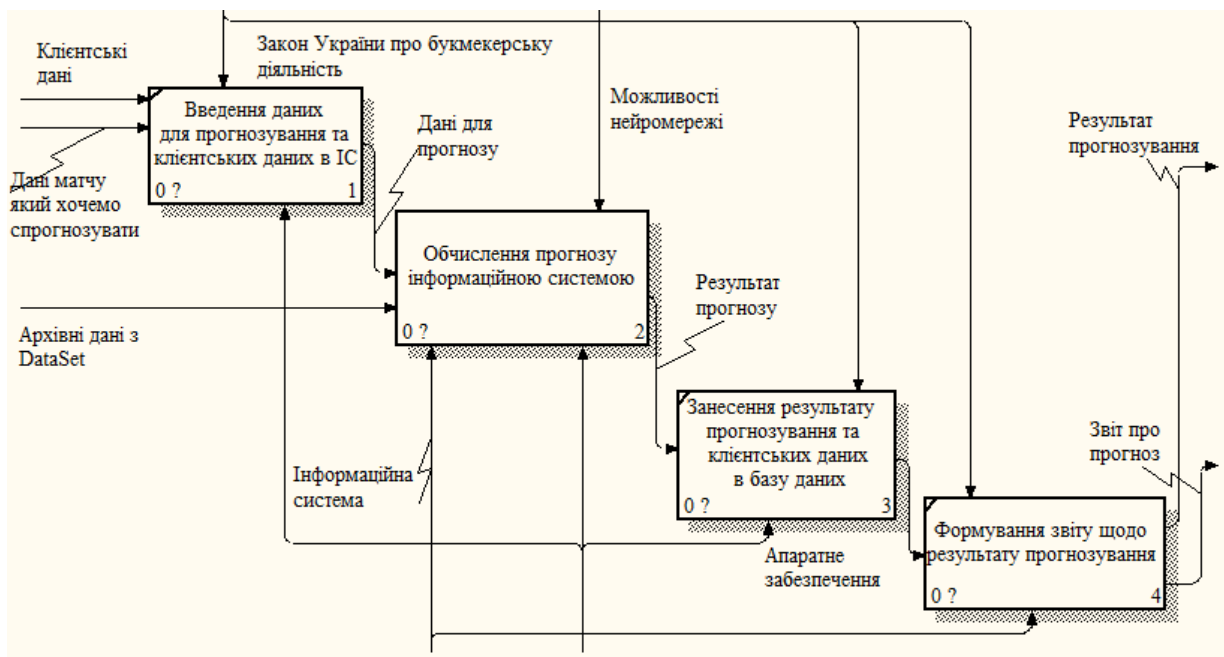


Рисунок 3.2 – Діаграма декомпозиції IDEF0

Вхідними даними до діяльності «введення даних для прогнозування та клієнтських даних в ІС» є: клієнтські дані, дані матчу який хочемо спрогнозувати. Для «обчислення прогнозу» вхідними даними є архівні дані з Dataset. Стрілками контролю для діяльностей «введення даних для прогнозування та клієнтських даних в ІС», «занесення результату прогнозування та клієнтських даних в базу даних» та «формування звіту щодо результату прогнозування» – «Закон України про букмекерську діяльність». Стрілками контролю для діяльності «Обчислення

прогнозу інформаційною системою» є «можливості нейромережі». Стрілками механізмів для діяльностей «введення даних для прогнозування та клієнтських даних в ІС», «занесення результату прогнозування та клієнтських даних в базу даних» та «формування звіту щодо результату прогнозування» – «Апаратне забезпечення», а для діяльності «Обчислення прогнозу» – «Інформаційна система».

Для розуміння процесів що відбуваються в системі декомпозовано діяльність «Обчислення прогнозу інформаційною системою». Діаграма декомпозиції показана на рис 3.3.

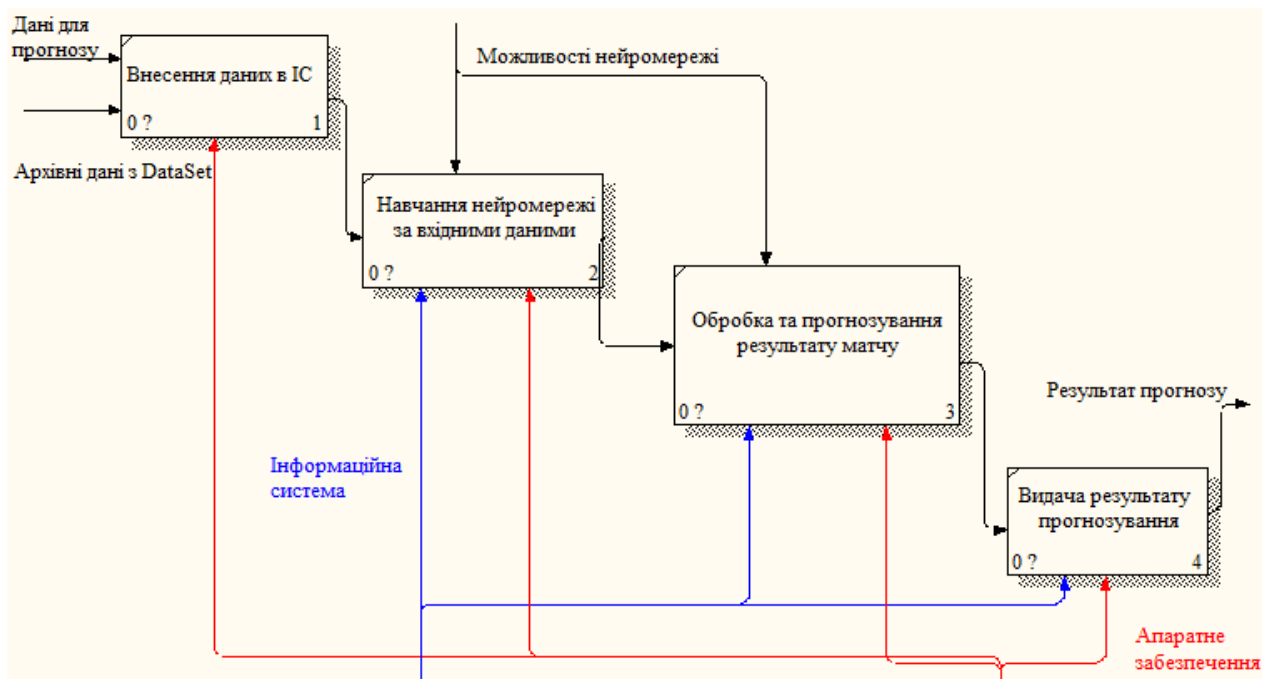


Рисунок 3.3 – Діаграма декомпозиції процесу «Обчислення прогнозу інформаційною системою»

Було виконано аналіз всіх процесів, що проходять в системі та показано їх на діаграмах. Проведена декомпозиція для більш точного розуміння зв'язків між процесами.

3.2 Моделювання варіантів використання інформаційної системи прогнозування результатів тенісних матчів

Для того аби описати дії, що може виконувати інформаційна система у відповідь на прохання користувачів була створена діаграма варіантів використання. Вона демонструє функціональність системи і те як вона повинна взаємодіяти з клієнтами.

Діаграма варіантів використання складається з акторів, елементів зовнішнього зв'язку з системою, самих варіантів використання та зв'язків між всіма частинами діаграми.

Діаграма варіантів використання показана на рис.3.4.

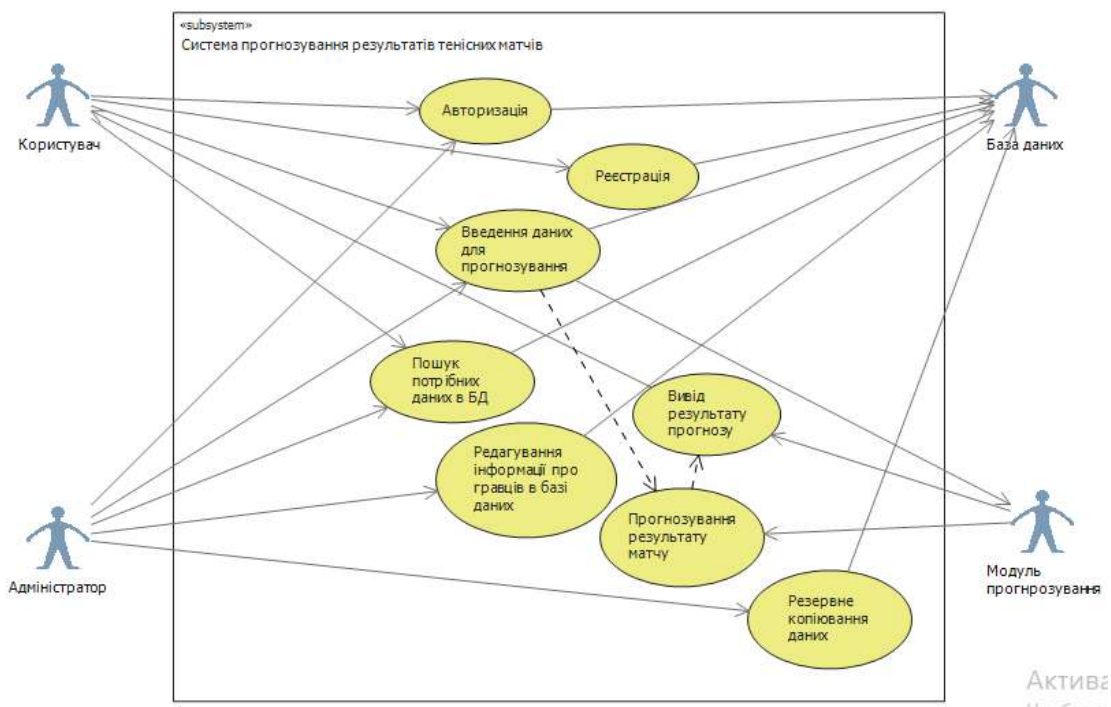


Рисунок 3.4 – діаграма варіантів використання інформаційної системи прогнозування результатів тенісних матчів

Елементами зовнішнього зв'язку в діаграмі виступають «База даних» та «Модуль прогнозування». До цих елементів інформаційна система звертається при потребі користувача.

До бази даних заносяться:

- інформація про користувача(його логін, пароль);
- інформація щодо матчу який користувач хоче спрогнозувати;
- результат прогнозованого матчу;
- заявки щодо прогнозування конкретним користувачем.

Модуль прогнозування приймає дані від користувача та бази даних та оброблюючи цю інформацію видає користувачеві прогноз з результатом конкретного матчу.

Акторами на даній діаграмі виступають користувач та адміністратор.

Користувач має можливість реєструватись та авторизовуватись в системі та викликати прогнозуючий модуль для знаходження результату матчу. Дані для прогнозування заносяться в базу даних.

Адміністратор має можливість редагувати діяльність прогнозуючого модулю та бази даних та виконувати резервне копіювання даних системи.

Варіанти використання:

- ВВ авторизація – дозволяє користувачу та адміністратору працювати з системою в залежності від наданих прав доступу;
- ВВ реєстрація – дозволяє користувачеві та адміністратору додавати свої дані для входу в систему до бази даних та подальшого входу по ним;
- ВВ введення даних для прогнозування – дозволяє користувачеві або адміністратору додавати до бази даних інформацію про гравців результат матчу яких потрібно визначити та подальшого передання цих даних до модуля прогнозування;
- ВВ пошук потрібних даних в БД – дозволяє користувачеві та адміністратору проводити пошук останніх матчів, які вони прогнозували в своєму профілі;

- ВВ редагування інформації про гравців в базі даних – дає можливість адміністратору редагувати дані про ігри гравців за останній їх період активної гри для більш точних результатів прогнозування;
- ВВ прогнозування результату матчу – при введенні користувачем або адміністратором даних гравців дає змогу модулю прогнозування проаналізувати їх та визначити результат матчу відносно одного гравця;
- ВВ вивід результату прогнозу – можливість модуля прогнозування видавати результат зпрогнозованого матчу користувачеві або адміністратору.
- ВВ резервне копіювання даних – дає можливість адміністратору виконувати резервне копіювання даних з бази для збільшення безпеки системи у разі виникнення ризикових ситуацій.

4. РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ТЕНІСНИХ МАТЧІВ

4.1 Підготовка навчальної вибірки

В прогнозуванні будь чого важливими параметрами є характеристики об'єктів з якими ведеться робота та історичні особливості поведінки цих об'єктів в системах. Всі ці дані формулюють навчальну вибірку за допомогою якої виконується прогнозування інформаційною системою.

Для прогнозування тенісних матчів було вирішено в якості навчальної вибірки для інформаційної системи використовувати характеристики гравців, що є числовими даними такими як вік гравця, ранг гравця в балах, швидкість подачі гравця та відсотки виграшів геймів при подачі та попадання м'ячем, середня кількість помилок за гру та інші характеристики. Всі вони впливають на результат гри між двома гравцями, але для більшої точності потрібно враховувати результати минулих ігор та рахунки з якими вони були закінчені. Всі ці дані дають навчальну вибірку, на якій буде навчатися інформаційна система. Фрагмент навчальної вибірки показаний на рис. 4.1.

pid2	3 sets							2 sets		Невимушені помилки		швидкість подачі м'яча		відсоток виграшів ударів своєю подачею		кількість ударів при розіграші у білу		Відсоток попадань першої подачі	
	1	1	2	2	3	3	sets	sets	кількість виграшів що йдуть вперед	Відсоток виграшів першого подачі	швидкість подачі м'яча	вік гравця	відсоток виграшів своєю подачею	кількість ударів при розіграші у білу	кількість виграшів що йдуть вперед	Відсоток попадань першої подачі			
596	6	4	6	3	0	0	2	0	3	63	2	2	169	24	60	5	1		
635	4	6	6	4	4	6	2	1	3	63	2	2	169	24	60	5	2		
797	6	4	6	3	0	0	2	0	3	63	2	2	169	24	60	5	1		
625	4	6	2	6	0	0	2	0	1	28	4	5	162	30	32	8	2		
797	6	4	4	6	4	6	1	2	1	28	4	5	162	30	32	8	1		
650	6	4	6	4	0	0	2	1	1	28	4	5	162	30	32	8	3		
732	4	6	6	4	6	4	2	1	2	58	2	4	155	27	58	4	2		
462	4	6	4	6	0	0	0	2	2	42	3	4	152	32	42	6	3		
619	4	6	6	4	4	6	1	2	1	40	4	5	134	36	36	6	3		
625	6	4	6	4	0	0	2	0	3	60	2	2	168	24	62	5	2		
489	6	4	6	4	0	0	2	0	3	37	3	3	169	27	58	4	1		
546	4	6	6	4	6	4	1	2	1	51	3	4	155	26	40	7	3		
500	4	6	4	6	0	0	0	2	0	32	7	8	122	37	21	9	8		
421	6	4	6	4	0	0	2	0	4	70	1	1	170	24	70	2	0		
325	6	4	4	6	4	6	1	2	2	62	3	2	165	25	50	4	3		
428	6	4	4	6	6	4	2	1	1	41	4	5	152	29	43	6	1		
485	4	6	4	6	0	0	0	2	4	62	3	3	172	25	53	4	4		

Рисунок 4.1 – фрагмент навчальної вибірки для інформаційної системи

4.2 Імпортування даних в систему

Для того щоб дані з Excel перенести у середовище розробки MATLAB потрібно натиснути кнопку Import data та вибрати потрібні колонки з навчальної вибірки, які в подальшому будуть входами нейронної мережі. Процес імпортування показаний на рис. 4.2. Таблиця з даними перетворюється в формат змінної Матлаб Numeric Matrix. Так само, але в іншу змінну імпортуються значення класів до яких відносяться дані.

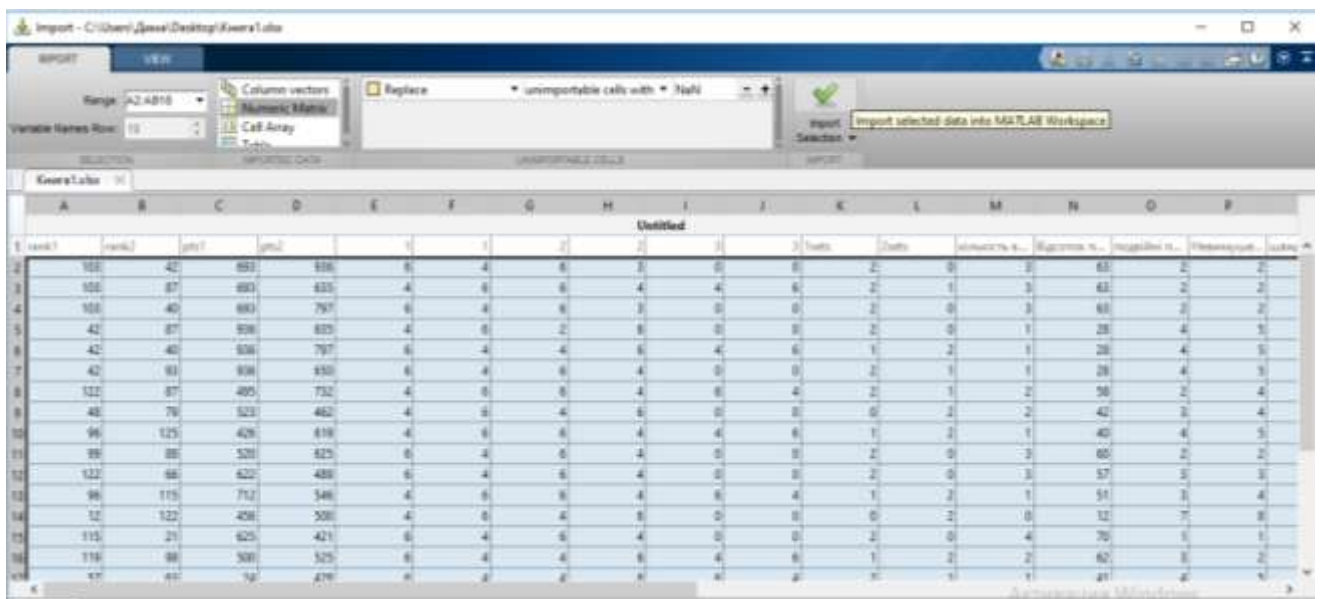


Рисунок 4.2 – процес імпортування даних в Матлаб

4.3 Розробка та тестування інформаційної системи прогнозування результатів тенісних матчів

Після імпортування даних в МАТЛАБ дані потрібно транспонувати для коректного відображення входів для мережі.

За створення багат шарового перцептрон у системі Матлаб відповідає команда `newff`. Команда має набір параметрів, що відповідають за характеристики мережі.

`NEWFF (PR, [S1 S2 ... SN], {TF1 TF2 ... TFN1}, BTF, LF, PF)` в якості вхідних параметрів використовує [12]:

- `PR` - $R \times 2$ матриця мінімальних і максимальних значень рядків вхідної матриці з розмірністю $R \times Q$. Для отримання матриці `PR` можна використовувати функцію `minmax`;

- `Si` - кількість нейронів в i -му шарі, `N1` - кількість шарів;

- `TFi` - функція активації i -го шару, за замовчуванням `'tansig'`;

- `BTF` - навчальна функція зворотного поширення, за замовчуванням `'trainlm'`;

- `BLF` - алгоритм підстроювання ваг і зміщень (навчальний алгоритм), за замовчуванням `'learngdm'`;

- `PF` - функція оцінки функціонування мережі, за замовчуванням `'mse'`;

Навчання мережі виконується за допомогою системної команди Матлаб `train`. `TRAIN` викликає функцію, яка задається `NET.trainFcn`, що використовує параметри тренування, що задаються `NET.trainParam`.

Зазвичай одна епоха тренування визначається як одноразове пред'явлення всіх вхідних векторів. Потім мережа модифікується відповідно до результатів всіх таких представлень.

Тренування триває до тих пір, поки не буде досягнуто задане значення помилки функціонування або поки не буде досягнуто одна з умов зупинки із заданих відповідно до виду обраної функції `NET.trainFcn` [13].

Після навчання та симуляції мережі були проведені дослідження зі зміною параметрів мережі для знаходження оптимального набору змінних та досягнення найкращої точності та швидкості навчання мережі.

При аналізі мереж було виділено дві характеристики, що впливають на точність навчання – кількість прихованих шарів та нейронів в них та алгоритм навчання мережі. Порівняльна характеристика мереж, що показує точність прогнозу показана в таблиці 4.1

Таблиця 4.1 – порівняльна характеристика мереж в залежності від їхніх характеристик

	trainlm	trains	traingdx	trainbr
2 шари	30%	15%	20%	25%
3 шари	65-70%	40%	40%	45%
4 шари	50%	35%	30%	40%

Найкращу точність показує система з 3 прихованих шарів та алгоритмом прогнозування trainlm. TRAINLM може навчати будь-яку мережу, якщо її вагові, вхідні і активаційні функції мають похідні.

Також на точність прогнозування прямо впливає кількість епох протягом яких навчалась система. Чим більша кількість епох – тим більша точність прогнозування.

Залежність епох від точності показана в таблиці 4.2.

Таблиця 4.2 – залежність точності від кількості навчальних епох.

Кількість епох	Точність
279	20%
1235	35%
4456	70%

Для обчислення ефективності функціонування щодо вагових змінних використовується алгоритм зворотного поширення. Кожна змінна модифікується відповідно до методу Левенберга-Маркара – він є альтернативою методу Ньютона. Може розглядатися як комбінація останнього з методом градієнтного спуску. Залежність точності від алгоритму показана на рис. 4.3.



Рисунок 4.3 – вплив шарів мережі на точність прогнозу

Для перевірки точності розробленої системи була створена тестова вибірка, що складається з 10 матчів між двома гравцями. Результат 5 матчів був відомий, але не занесений в навчальну вибірку, а результат інших 5 ігор був невідомий, бо гравці не зустрічались між собою. Для знаходження приблизного результату невідомих ігор була взята експертна думка, та її результат порівнювався з результатом що видавала мережа. Схема мережі показана на рис.4.4.

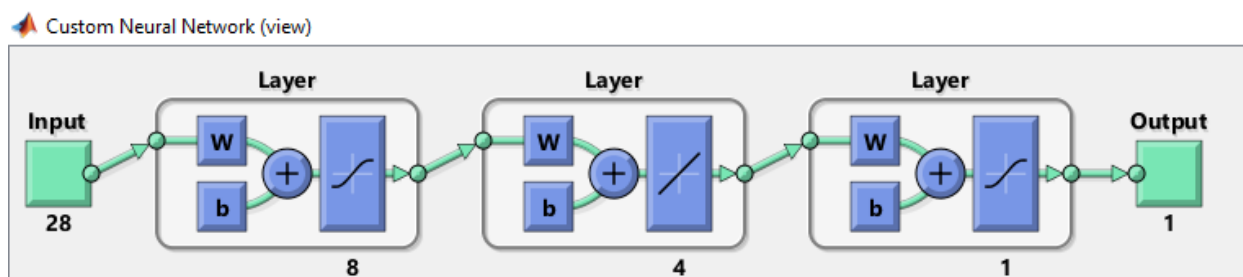


Рисунок 4.4 – схема мережі для прогнозування

Результатом тестування мережі є 7 вірних відповідей, та 3 невірних прогнози. Точність системи складає приблизно 65 – 70%. Це пов'язане з малою кількістю навчальних даних та можливою модифікацією алгоритму системи для збільшення точності. Результати навчання мережі показані на рис.4.5 , а результат тестування на рис 4.6.

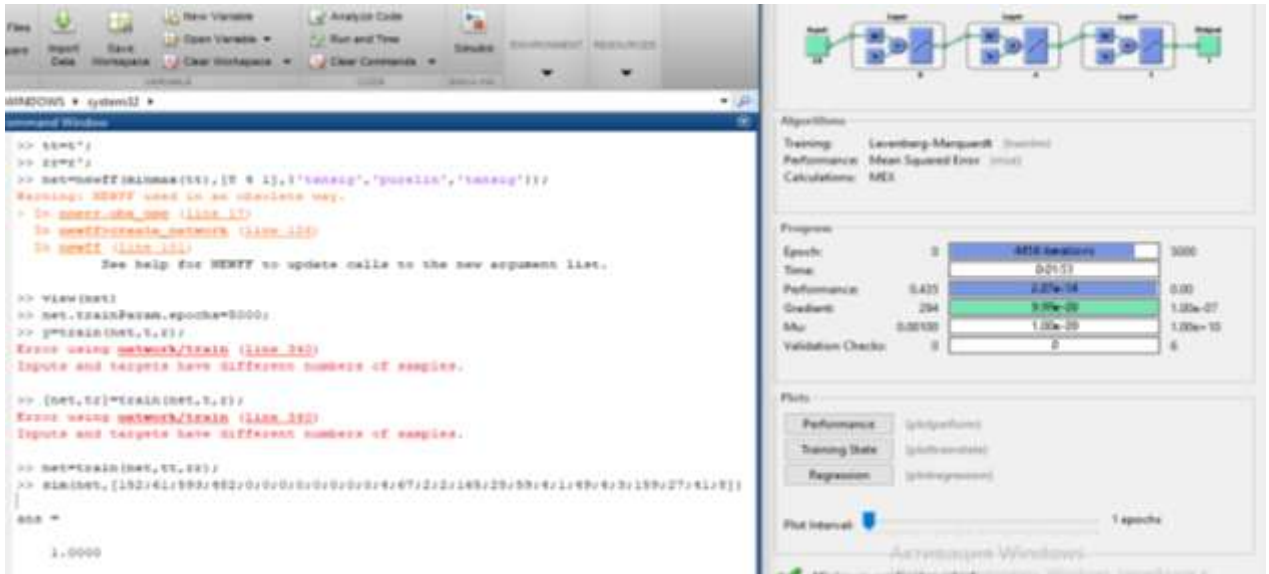


Рисунок 4.5 – результат навчання мережі з відповідними параметрами

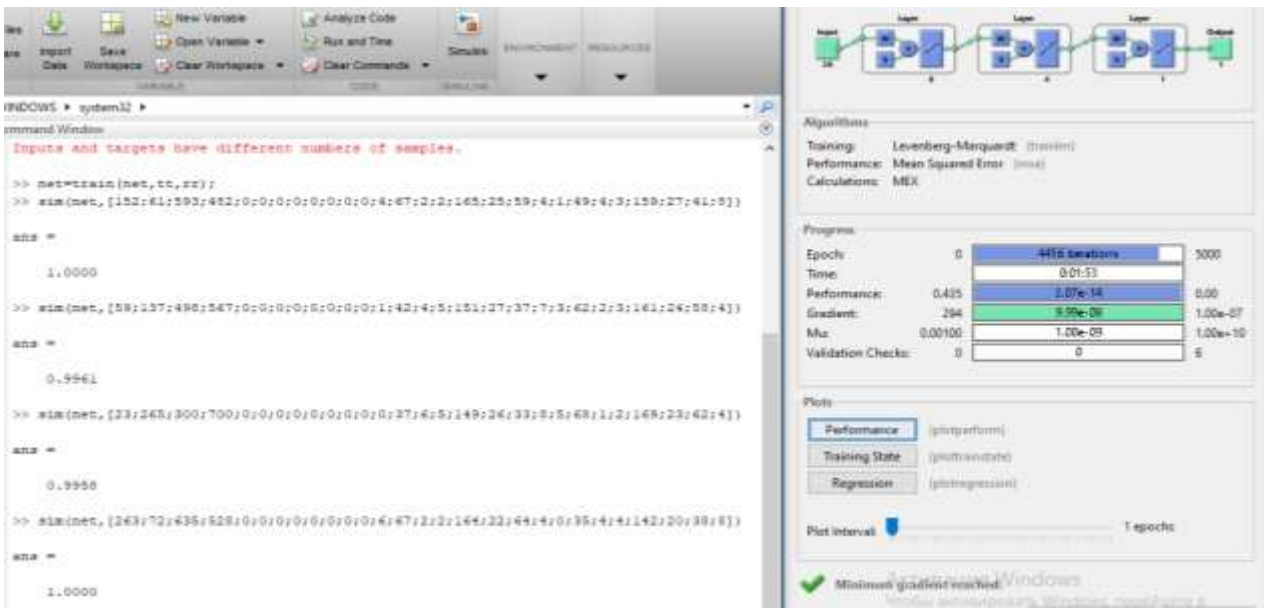


Рисунок 4.6 – результат тестування мережі

При тестуванні система видає значення 1 для першого класу (переможе гравець 1) та менше 1 для 2 класу(переможе гравець 2). Такі виходи системі надає функція активації вихідного шару Tansig, що являє собою гіперболічний тангенс.

Код програми та тестування мережі було виконано за допомогою команд:

```
tt='t';
rr='r';
net=newff(minmax(tt),[8 4 1],{'tansig','purelin','tansig'});
view(net)
net.trainParam.epochs=5000;
net=train(net,tt,rr);
sim(net,[152;61;593;482;0;0;0;0;0;0;0;0;4;67;2;2;165;25;59;4;1;49;4;3;159;27;41;8])
ans = 1.0000
sim(net,[59;137;498;567;0;0;0;0;0;0;0;0;1;42;4;5;151;27;37;7;3;62;2;3;161;24;58;4])
ans = 0.9961
sim(net,[23;265;300;700;0;0;0;0;0;0;0;0;37;6;5;149;26;33;8;5;68;1;2;169;23;62;4])
ans = 0.9958
sim(net,[263;72;635;528;0;0;0;0;0;0;0;0;6;67;2;2;164;22;64;4;0;35;4;4;142;20;38;8])
ans = 1.0000
sim(net,[179;57;712;428;0;0;0;0;0;0;0;0;7;72;2;2;159;22;62;4;1;33;5;4;138;21;35;7])
ans = 1.0000
sim(net,[63;118;532;599;0;0;0;0;0;0;0;0;2;41;5;4;152;25;38;6;4;61;1;3;162;23;63;4])
ans = 0.9970
```

Подальша модифікація системи можлива та доцільна, а саме збільшення навчальної вибірки мережі, зміна параметрів та функцій активації та алгоритмів навчання для збільшення точності, але як наслідок збільшення часу навчання мережі. Можлива зменшення кількості вхідних параметрів для зменшення часу обробки при навчанні.

4.4 Розробка web-додатку прогнозування результатів тенісних матчів

За допомогою PHP був створений web-додаток який звертається до серверу на якому встановлений Матлаб з М - файлом програми прогнозування. Архітектура додатку показана на рис. 4.7.



Рисунок 4.7 – архітектура інформаційної системи прогнозування результатів тенісних матчів

Після введення даних користувачем в додатку на форму, форма звертається до PHPсерверу який в свою чергу викликає М-file Матлабу, що виконує прогнозування матчу. Після цього додаток повертає на сервер прогноз, а сервер видає його користувачеві. Після виконання файлу видається прогноз результату матчу, виходячи з вхідних параметрів. Демонстрація роботи системи показана на рис.4.8 та рис.4.9.



Рисунок 4.8 – головна сторінка додатку прогнозування



Рисунок 4.9 – процес видачі результату прогнозування користувачеві

ВИСНОВОК

В ході виконання дипломного проекту було зроблено аналіз проектної області прогнозування результатів спортивних подій за допомогою неймереж. В результаті була поставлена задача створити власну інформаційну систему прогнозування результатів тенісних матчів.

В аналітичній частині дипломної роботи були описані вимоги до інформаційної системи, обґрунтовано вибір програмного забезпечення для реалізації завдання та методів і алгоритмів на основі яких буде створюватись інформаційна система.

У проектній частині роботи для реалізації були розроблені: модель процесів, діаграма варіантів використання.

На основі методів розробки систем прогнозування та цілей які потрібно досягнути при створенні мережі в системі Матлаб була розроблена неймережа для прогнозування результатів тенісних матчів. Створена система виконує прогнозування і має достатню точність. Може бути покращена для збільшення результативності та модифікована для прогнозування в інших сферах життя.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. http://www.irbisnbuv.gov.ua/cgi-bin/irbis_nbuv/.– [Електронний ресурс].
2. http://ela.kpi.ua/bitstream/123456789/23823/4/Didkivska_magistr.pdf
[Електронний ресурс].
3. https://studopedia.com.ua/1_69143_tema--prognozuvannya-v-sportI.html. [Електронний ресурс].
4. <https://uk.wikipedia.org/wiki/Прогнозування> – [Електронний ресурс].
5. <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>
[Електронний ресурс].
6. <https://neuralnet.info/chapter/персептрони/> – [Електронний ресурс].
7. https://ru.wikipedia.org/wiki/Нейронная_сеть_Элмана – [Електронний ресурс].
8. https://uk.wikipedia.org/wiki/Нелінійна_авторегресійна_екзогенна_модел
[Електронний ресурс].
9. <http://qaru.site/questions/10332542/narx-neural-network-prediction> [Електронний ресурс].
10. http://www.rusnauka.com/18_EN_2009/Economics/48755.doc.htm Структурно-функціональне моделювання процесу управління [Електронний ресурс].
11. <http://molodyvcheny.in.ua/files/journal/2016/6/24.pdf> Структурно-функціональне моделювання – [Електронний ресурс].
12. <http://matlab.exponenta.ru/neuralnetwork/book2/11/newff.php> [Електронний ресурс].
13. <http://matlab.exponenta.ru/neuralnetwork/book2/10/train.php> [Електронний ресурс].
14. <http://jak.iblog.in.ua/articles/metodi-teorii-jmovirnostej-i-statistiki-v-stavkah.html> [Електронний ресурс].
15. <https://int-konf.org/ru/2016/naukovij-potentsial-2016-16-18-03-2016/1230-isaev-d-v-intelektualnij-analiz-danikh-v-medsini> [Електронний ресурс].

ТЕХНІЧНЕ ЗАВДАННЯ
на розробку інформаційної системи «Прогнозування тенісних матчів»

3. ПРИЗНАЧЕННЯ Й МЕТА СТВОРЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Призначення інформаційної системи

Інформаційна система повинна представляти додаток для прогнозування результатів тенісних матчів.

Мета створення інформаційної системи –

Полегшення роботи тренерського складу в роботі з спортсменами та автоматизація прогнозування в букмекерських конторах.

Цільова аудиторія

У цільовій аудиторії інформаційної системи можна виділити наступні групи:

1. Спортсмени.
2. Тренерський склад.
3. Букмекерські контори.
4. Інші зацікавлені особи.

4. ВИМОГИ ДО ІНФОРМАЦІЙНОЇ СИСТЕМИ

Вимоги до інформаційної системи в цілому

Вимоги до структури й функціонування інформаційної системи

Інформаційна система повинна бути реалізована у вигляді додатку. Він повинен складатися із взаємозалежних розділів із чітко розділеними функціями.

Вимоги до персоналу

Для підтримки додатку й експлуатації системи від персоналу не повинно вимагатися спеціальних технічних навичок, знання технологій або програмних продуктів, за винятком загальних навичок роботи з персональним комп'ютером.

Вимоги до збереженні інформації

У системі керування додатком повинен бути передбачений механізм резервного копіювання структури й умісту бази даних. Процедура резервного копіювання повинна проводитися співробітником, відповідальним за підтримку додатку, не рідше 1 рази в місяць.

Вимоги до розмежування доступу

Інформація, розташовувана в додатку, є загальнодоступною.

Користувачів можна розділити на 2 групи відповідно до прав доступу:

1. Відвідувачі
2. Редактор

Відвідувачі мають доступ тільки до загальнодоступної частини додатку.

Редактор може редагувати матеріали розділів.

Вимоги до функцій, виконуваних сайтом

Основні вимоги

Структура додатку

Додаток повинен складатися з наступних розділів:

- Головна – виводиться довідкова інформація згідно запиту.
- Зпрогнозовані матчі – інформація про вже зпрогнозовані системою матчі.

Навігація

Користувацький інтерфейс повинен забезпечувати наочне, інтуїтивно зрозуміле представлення структури розміщеної на ньому інформації, швидкий і логічний перехід до розділів. Навігаційні елементи повинні забезпечувати однозначне розуміння користувачем їх змісту: посилання на сторінки повинні бути мати заголовки, умовні позначки відповідати загальноприйнятим. Графічні елементи навігації повинні бути мати альтернативний підпис.

Вимоги до видів забезпечення

2.3.1 Вимоги до інформаційного забезпечення

Реалізація додатку відбувається з використанням:

- Python 3.7.2
- Matlab 2016b
- MySQL 5.6

2.3.2 Вимоги до програмного забезпечення

Програмне забезпечення клієнтської частини повинне задовольняти наступним вимогам:

- Веб-браузер: Internet Explorer 7.0 і вище, або Firefox 3.5 і вище, або Opera 9.5 і вище, або Safari 3.2.1 і вище, або Chrome 2 і вище;
- Включена підтримка javascript, Flash і cookies.

3 Склад і зміст робіт зі створення сайту

Докладний опис етапів роботи зі створення додатку наведено в табл. 1.

Таблиця 1 – Етапи створення сайту

№	Склад і зміст робіт	Строк розробки (у робочих днях)
1	Знаходження Dataset: Знаходження бази даних з характеристиками гравців	1 день
2	Обробка Dataset: Обробка та усунення непотрібних даних для навчальної вибірки	1 день
3	Створення в Matlab: Програмування мережі в середовищі Matlab та її тестування	7 днів
4	Створення на Python: Створення мережі на мові Python і фреймворку flask	5 день
5	Завершення роботи: Проведення стилістичних виправлень додатку, перевірка (тестування) реалізованого функціоналу	2 дні
ЗАГАЛЬНА ТРИВАЛІСТЬ РОБІТ (З УРАХУВАННЯМ РЕЗЕРВНОГО СТРОКУ НА НАЛАГОДЖЕННЯ Й ВИПРАВЛЕННЯ ПОМИЛОК) І СТРОК ЗАКІНЧЕННЯ ПРОЕКТУ		18

Додаток Б. Планування робіт

Деталізація мети проекту методом SMART. Продуктом дипломного проекту є додаток для прогнозування результатів тенісних матчів. Результати деталізації методом SMART розміщені у табл. Б.1.

Таблиця Б.1 – Деталізація мети методом SMART

Specific (конкретна)	Створити додаток для прогнозування результатів тенісних матчів.
Measurable (вимірювана)	Результатом роботи проекту є оцінка замовника.
Achievable (досяжна)	Реалізації системи здійснюється за допомогою середовища Matlab 2016b та мови Python з фреймворком flask.
Relevant (реалістична)	У наявності є всі необхідні технічні та програмні засоби. Розробники достатньо кваліфіковані для виконання поставлених задач.
Time-framed (обмежена у часі)	Ціль має часове обмеження. Робота повинна бути виконана у терміни, що були оговорені замовником проекту. Проект повинен бути виконаний згідно з календарним планом.

Планування змісту структури робіт. Основним інструментом для планування змісту структури робіт служить WBS діаграма – графічне подання згрупованих елементів проекту у вигляді пакета робіт, які ієрархічно пов'язані з продуктом проекту. Побудуємо структуру WBS, у якій детально опишемо роботи, які потрібно виконати на кожному етапі створення проекту. Виконаємо декомпозицію робіт для даного проекту. Діаграма WBS зображена на рис. Б.1.

Планування структури організації, для впровадження готового проекту (OBS). Після побудови WBS розробимо організаційну структуру виконавців OBS. Організаційна структура проекту стосується тільки внутрішньої організаційної структури проекту і не стосується відносин проектних груп чи учасників з батьківськими організаціями. Діаграма OBS зображена на рис. Б.2. Список виконавців, що функціонують в проекті знаходиться в табл. Б.2.

Таблиця Б.2 – Виконавці проекту

Роль	Ім'я	Проектна роль
Розробник	Гагін Д.О.	Виконує розробку основного функціоналу проекту та інтерфейс користувача
Тестувальник	Незалежна особа	Відповідає за тестування функціоналу та дизайну додатку, перевірку моделі на адекватність.
Косультант проекту	Парфененко Ю.В.	Формує завдання на розробку проекту.

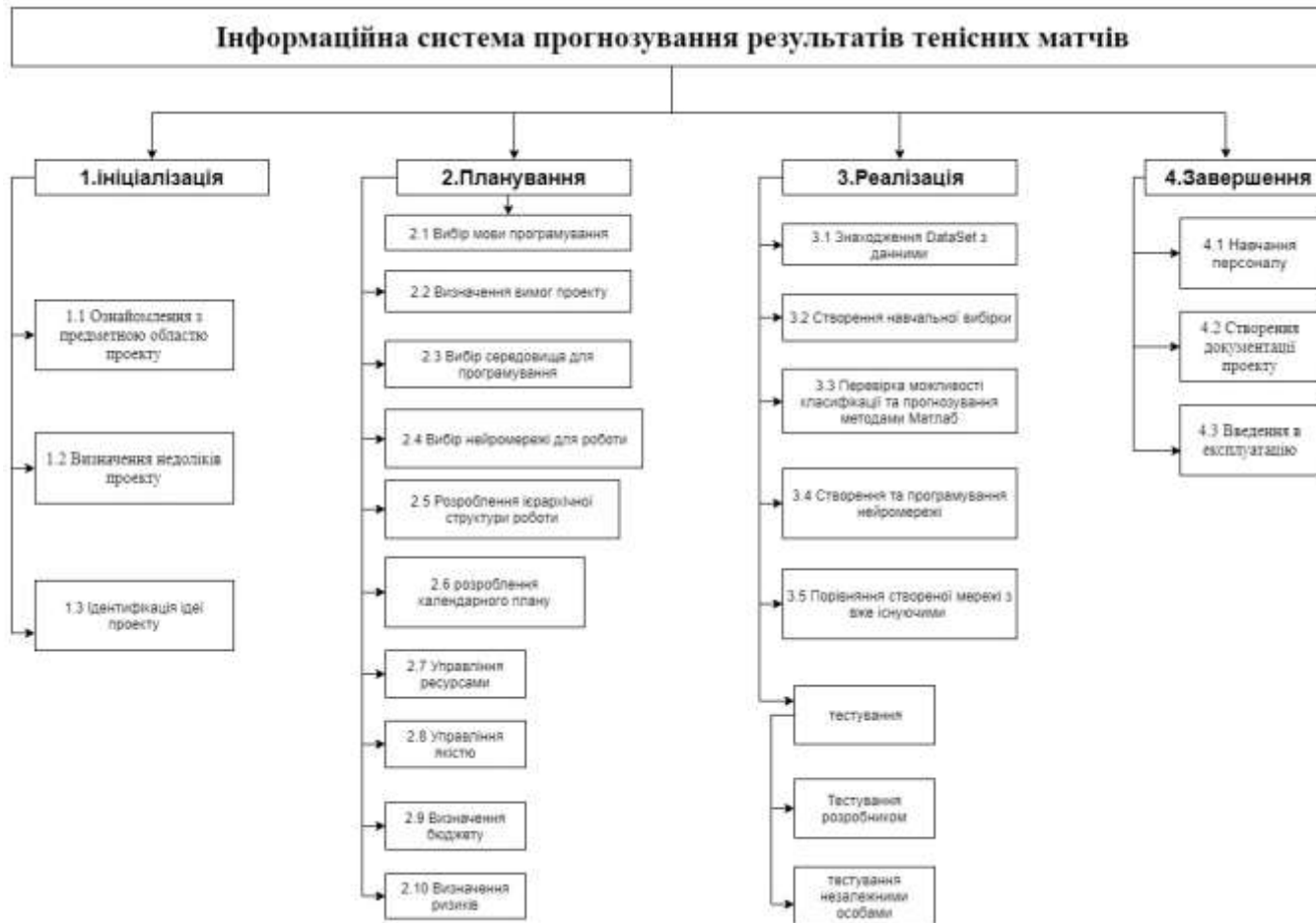


Рисунок Б.1 – WBS. Структура робіт проекту

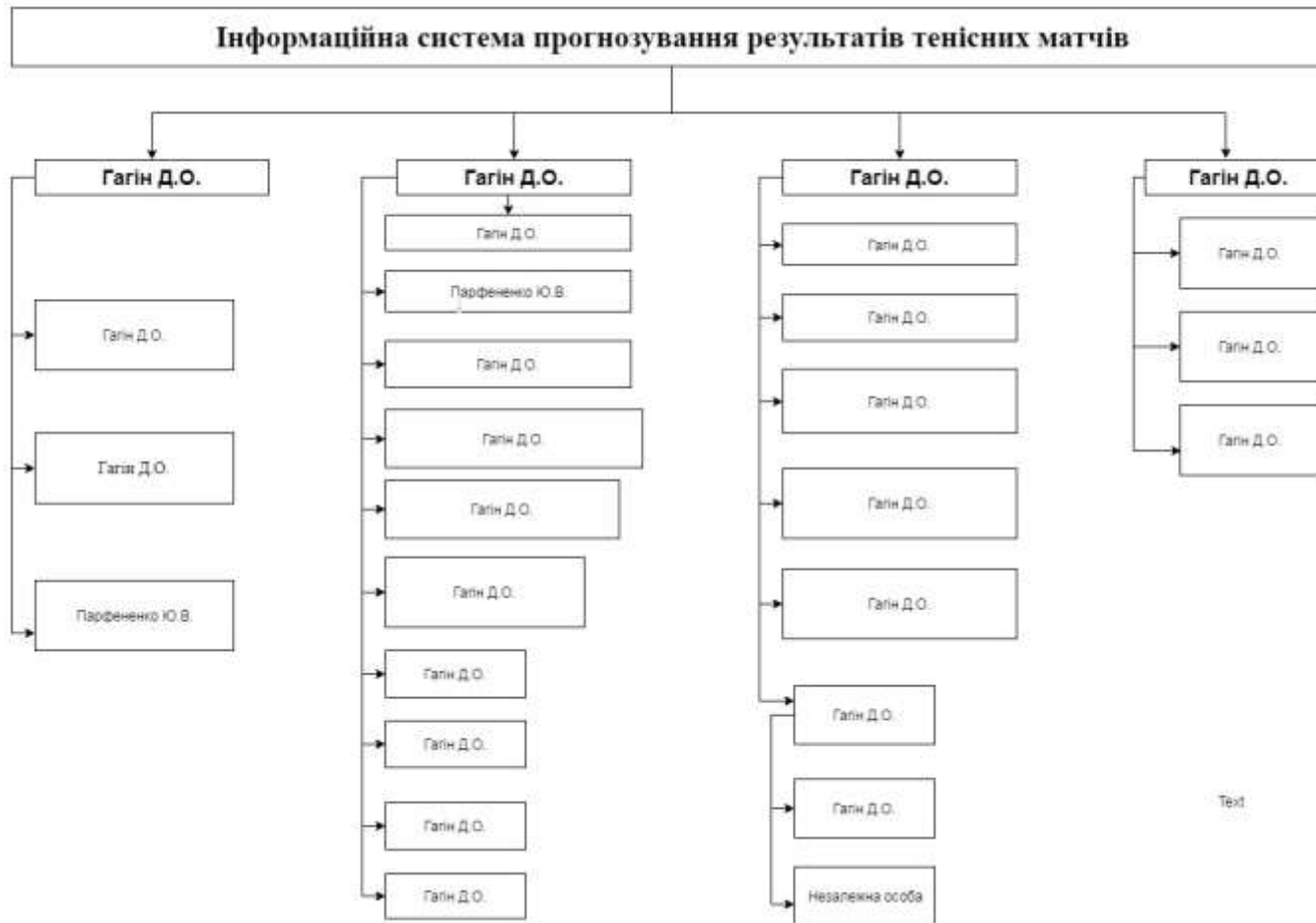


Рисунок Б.2 – Організаційна структура проекту (OBS)

Діаграма Ганта. Далі побудуємо календарний план виконання дипломного проекту. Найпоширеніший формат графіка в будь-якій галузі — діаграма Ганта. Цей графік дозволяє менеджерам проекту і всій команді розробників візуалізувати графіки часу і взаємозв'язок між окремими завданнями та етапами роботи над проектом. Тривалість виконання робіт зазначена в днях, але фактична тривалість виконання робіт приблизно дорівнює 4-6 години на день. Для того щоб мати реальне уявлення про тривалість виконання робіт з урахуванням обмеженості у використанні ресурсів, з урахуванням вихідних та святкових днів, побудовано календарний графік. Діаграма Ганта та список робіт діаграми Ганта та мережевий графік проекту зображені на рис. Б.3-Б.5.

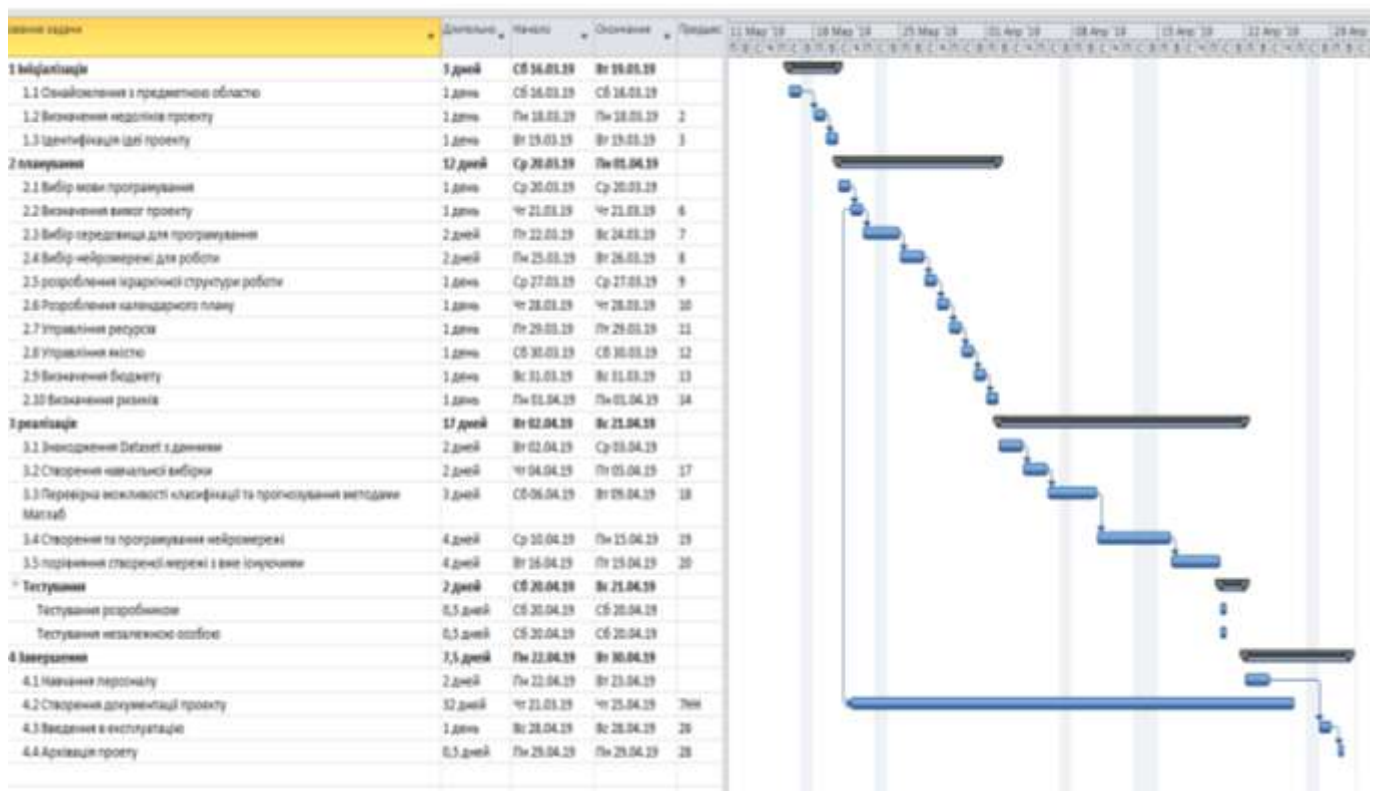


Рисунок Б.3 – Діаграма Ганта

№	Режим задачі	Назва задачі	Длительность	Начало	Окончание	Преддес
1		1 Ініціалізація	3 дней	Сб 16.03.19	Вт 19.03.19	
2		1.1 Ознайомлення з предметною областю	1 день	Сб 16.03.19	Сб 16.03.19	
3		1.2 Визначення недоліків проекту	1 день	Пн 18.03.19	Пн 18.03.19	2
4		1.3 Ідентифікація ідей проекту	1 день	Вт 19.03.19	Вт 19.03.19	3
5		2 планування	12 дней	Ср 20.03.19	Пн 01.04.19	
6		2.1 Вибір мови програмування	1 день	Ср 20.03.19	Ср 20.03.19	
7		2.2 Визначення вимог проекту	1 день	Чт 21.03.19	Чт 21.03.19	6
8		2.3 Вибір середовища для програмування	2 дней	Пт 22.03.19	Вс 24.03.19	7
9		2.4 Вибір нейромережі для роботи	2 дней	Пн 25.03.19	Вт 26.03.19	8
10		2.5 розроблення ієрархічної структури роботи	1 день	Ср 27.03.19	Ср 27.03.19	9
11		2.6 Розроблення календарного плану	1 день	Чт 28.03.19	Чт 28.03.19	10
12		2.7 Управління ресурсів	1 день	Пт 29.03.19	Пт 29.03.19	11
13		2.8 Управління якістю	1 день	Сб 30.03.19	Сб 30.03.19	12
14		2.9 Визначення бюджету	1 день	Вс 31.03.19	Вс 31.03.19	13
15		2.10 Визначення ризиків	1 день	Пн 01.04.19	Пн 01.04.19	14
16		3 реалізація	17 дней	Вт 02.04.19	Вс 21.04.19	
17		3.1 Знаходження Dataset з даними	2 дней	Вт 02.04.19	Ср 03.04.19	
18		3.2 Створення навчальної вибірки	2 дней	Чт 04.04.19	Пт 05.04.19	17
19		3.3 Перевірка можливості класифікації та прогнозування методами Матлаб	3 дней	Сб 06.04.19	Вт 09.04.19	18
20		3.4 Створення та програмування нейромережі	4 дней	Ср 10.04.19	Пн 15.04.19	19
21		3.5 порівняння створеної мережі з вже існуючими	4 дней	Вт 16.04.19	Пт 19.04.19	20
22		Тестування	2 дней	Сб 20.04.19	Вс 21.04.19	
23		Тестування розробником	0,5 дней	Сб 20.04.19	Сб 20.04.19	
24		Тестування незалежною особою	0,5 дней	Сб 20.04.19	Сб 20.04.19	
25		4 Завершення	7,5 дней	Пн 22.04.19	Вт 30.04.19	
26		4.1 Навчання персоналу	2 дней	Пн 22.04.19	Вт 23.04.19	
27		4.2 Створення документації проекту	32 дней	Чт 21.03.19	Чт 25.04.19	7НН
28		4.3 Введення в експлуатацію	1 день	Вс 28.04.19	Вс 28.04.19	26
29		4.4 Архівізація проекту	0,5 дней	Пн 29.04.19	Пн 29.04.19	28

Рисунок Б.4 – Список робіт для побудови діаграми Ганта

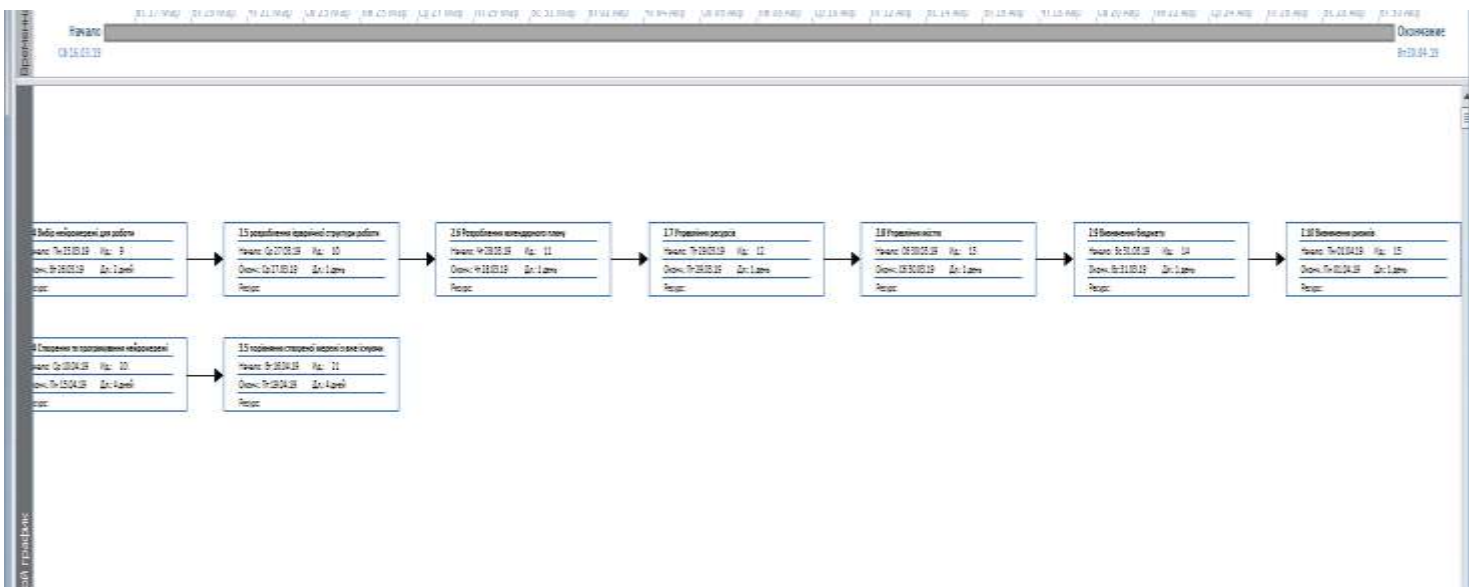
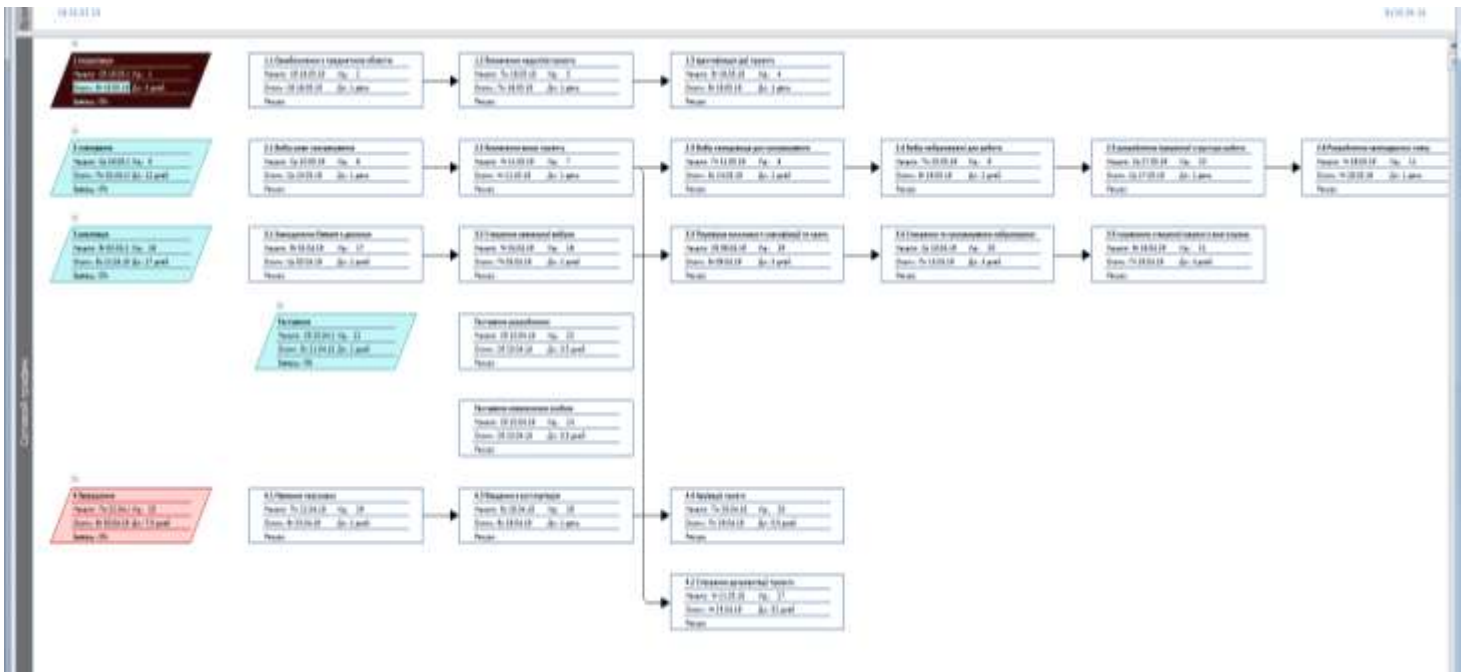


Рисунок Б.5– Мережевий графік проекту

Аналіз ризиків. Ретельне і детальне планування управління ризиками спрямовано на підвищення вірогідності досягнення цілей проекту. Даний процес повинен бути завершений на ранній стадії планування.

Якісний аналіз ризиків включає розстановку пріоритетів для ідентифікованих ризиків, результати якої використовуються потім у ході кількісного аналізу ризиків і планування реагування на ризики.

У ході якісного оцінювання використовується нечислова шкала вірогідності, наприклад:

1 - дуже слабкий вплив;

2 - слабкий вплив;

3 - середній вплив;

4 - сильний вплив;

5 - дуже сильний вплив.

Оцінку ризиків проводять за допомогою оцінки вірогідності і наслідків. Один з можливих прикладів подібної матриці, в якій показані тільки оцінки наслідків, представлений у таблиці Б.3.

Таблиця Б.3 – Матриця оцінки наслідків

Ціль проекту	Дуже слабкий вплив – 0,05	Слабкий вплив - 0,01	Середній вплив - 0,2	Сильний вплив - 0,4	Дуже сильний вплив - 0,8
Вартість	Несуттєве збільшення бюджету	Збільшення бюджету до 10%	Збільшення бюджету на 10-20%	Збільшення бюджету на 20-30%	Збільшення бюджету більше, ніж на 40%
Терміни	Несуттєве збільшення календарного плану	Порушення календарного плану не більше ніж на 5%	Порушення календарного плану на 5-10%	Порушення календарного плану на 10-20%	Порушення календарного плану більш ніж на 20%
Якість	Несуттєве зниження якості	Суттєве зниження якості	Зниження якості потребує узгодження з замовником	Зниження якості неприйнятне для замовника	Результат проекту повністю даремний

Існують три стратегії реагування на появу негативних ризиків, до яких відносяться:

- Ухилення від ризику передбачає змінення плану управління проектом таким чином, щоб виключити ризик, убезпечити цілі проекту від наслідків ризику.

- Передача ризику має на увазі перекладення негативних наслідків на третю сторону. Передача ризику просто переносить відповідальність за його управління іншій стороні, але ризик при цьому не зникає. Передача ризику є найбільш ефективною у відношенні фінансових ризиків. Передача ризику практично завжди передбачає виплату премії за ризик стороні, яка прийняла ризик на себе. В якості інструментів передачі ризиків використовуються страховки, гарантійні зобов'язання і т.д. У ряді випадків витрати на ризики можуть перекладатися на покупця або продавця, що оговорюється у контракті.

- Зниження ризику передбачає зниження вірогідності наслідків негативної ризикованої події до прийнятних границь. Прийняття попереджувальних заходів зі зниження вірогідності настання ризику або його наслідків часто буває більш ефективним, ніж зусилля з усунення негативних наслідків, що здійснюються після настання події ризику. В якості прикладів заходів зі зниження ризиків можна привести: упровадження менш складних процесів, проведення великої кількості випробувань або вибір постачальника, поставки якого мають більш стабільний характер.

До стратегій реагування на позитивні ризики (сприятливі можливості) відносяться:

- Використання. Дана стратегія призначена для усунення всіх невизначеностей, пов'язаних з ризиком верхнього рівня. До числа заходів прямого реагування відноситься, наприклад, залучення до участі у проекті

більш талановитих спеціалістів для того, щоб скоротити строки виконання проекту або домогтися більш високої якості.

- **Спільне використання.** Спільне використання позитивних ризиків передбачає передачу відповідальності третій стороні, яка здатна найкращим чином скористатися сприятливою можливістю в інтересах проекту. До числа таких заходів відносяться: утворення партнерств зі спільною відповідальністю за ризики команд, спеціалізованих компаній або спільних підприємств.

- **Посилення.** При застосуванні цієї стратегії змінюється «розмір» сприятливої можливості шляхом підвищення вірогідності виникнення позитивного впливу.