Міністерство освіти і науки України

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Кафедра прикладної математики

Пояснювальна записка

до розрахунково-графічної роботи

з дисципліни “Алгоритми і системи комп’ютерної математики”

на тему:

Створення системи для автоматизації видачі кредиту

Студента групи КМ-2

Федуров Данило

Викладач

Олефір О.С.

Київ – 2020

ЗМІСТ

1 ВІДОМОСТІ ПРО РОЗРАХУНКОВО-ГРАФІЧНУ РОБОТУ

2 МЕТОДИ РОЗВ’ЯЗАННЯ ЗАДАЧІ

3 ОПИС ОБРАНОГО МЕТОДУ

4 СКЛАД ТА ФОРМАТИ ВИХІДНИХ ДАНИХ ТА РЕЗУЛЬТАТІВ

5 СХЕМА ВЗАЄМОДІЇ ПРОГРАМНИХ ЗАСОБІВ

6 ЕКРАННІ ФОРМИ

7 ВІЗУАЛЬНИЙ ІНТЕРФЕЙС

8 ВИСНОВКИ

9 СПИСОК ЛІТЕРАТУРИ

# 

# 1 ВІДОМОСТІ ПРО РОЗРАХУНКОВО-ГРАФІЧНУ РОБОТУ

Назва розрахунково-графічної роботи:

«Платоспроможність клієнта при кредитуванні. Способи та методи прогнозування Місце виконання: місце навчання»

Замовник: університет/ банківська система .

Технічні засоби: ноутбук, бд, середа розроблення.

Програмне забезпечення: Python.

# 

# 2 МЕТОДИ РОЗВ’ЯЗАННЯ ЗАДАЧІ

Створити автоматичний алгоритм, який буде визначати кредитоспроможність клієнта за допомогою методів машинного навчання (accuracy > 75%).

Методи для розв’язання задачі:

1. Decision Tree Classifier

2. Random Forest Classifier

3.  Logistic Regression

4. Linerar Regression

ОПИС ОБРАНОГО МЕТОДУ

Decision Tree Classifier

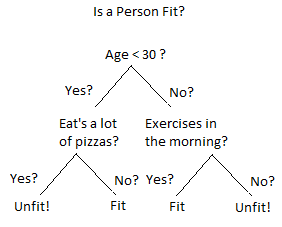
Дерево рішень - це просте представлення для класифікації прикладів. Це контрольоване машинне навчання, де дані безперервно розділяються відповідно до певного параметра.

Дерево рішень складається з:

Вузли: Перевірте значення певного атрибута.

Краї / Гілка: Відповідають результату тесту і підключаються до наступного вузла або аркуша.

Листові вузли: Кінцеві вузли, які передбачають результат (представляють мітки класів або розподіл класів).



Щоб зрозуміти концепцію дерева рішень, розглянемо наведений вище приклад. Скажімо, ви хочете передбачити, чи є людина придатною чи непридатною, враховуючи її інформацію, таку як вік, харчові звички, фізична активність тощо. Вузлами прийняття рішень є питання на зразок „Який вік?“, „Чи займається він?“, „ Він їсть багато піци? А листя представляють результати, подібні до «придатності» або «непридатності».

Є два основних типи дерев рішень:

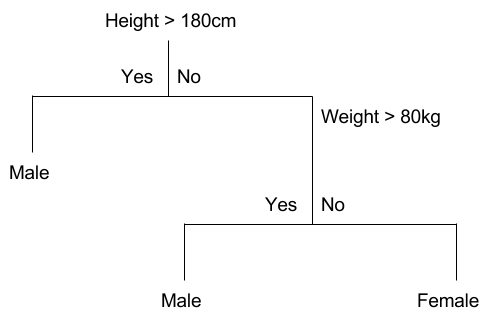
Класифікація дерев.

Дерева регресії.

Класифікаційні дерева (так / ні типи):

Те, що ми бачили вище, є прикладом дерева класифікації, де результатом була така змінна, як „придатність” чи „непридатність”. Тут змінна рішення є категоричною / дискретною.

Таке дерево будується за допомогою процесу, відомого як двійкове рекурсивне розділення. Це ітераційний процес розділення даних на розділи, а потім подальшого розподілу по кожній з гілок.



Дерева регресії (безперервні типи даних):

Дерева рішень, де цільова змінна може приймати безперервні значення (як правило, реальні числа), називаються деревами регресії. (наприклад, ціна будинку або тривалість перебування пацієнта в лікарні)

Основний алгоритм поділу та перемоги:

Виберіть тест для кореневого вузла. Створіть гілку для кожного можливого результату тесту.

Розбийте екземпляри на підмножини. По одному на кожну гілку, що йде від вузла.

Повторіть рекурсивно для кожної гілки, використовуючи лише ті екземпляри, які досягли гілки.

Зупиніть рекурсію для гілки, якщо всі її екземпляри мають однаковий клас.

Класифікатор дерева рішень

Використовуючи алгоритм прийняття рішень, ми починаємо з кореня дерева і розділяємо дані щодо ознаки, що призводить до найбільшого прибавлення інформації (IG) (зменшення невизначеності до остаточного рішення).

В ітераційному процесі ми можемо повторити цю процедуру розділення на кожному дочірньому вузлі, поки листя не стануть чистими. Це означає, що зразки на кожному листовому вузлі належать до одного класу.

На практиці ми можемо встановити обмеження глибини дерева, щоб запобігти переобладнанню. Ми дещо йдемо на компроміс щодо чистоти, оскільки остаточне листя все ще може мати певну домішку.

Переваги класифікації з деревами рішень:

Недорого побудувати.

Надзвичайно швидко класифікує невідомі записи.

Легко інтерпретувати для невеликих дерев

Точність порівнянна з іншими методами класифікації для багатьох простих наборів даних.

Виключає неважливі функції.

Недоліки класифікації з деревами рішень:

Легко переобладнати.

Межа рішення обмежена паралельністю осям атрибутів.

Моделі дерев рішень часто схильні до розбиття на особливості, що мають велику кількість рівнів.

Невеликі зміни в навчальних даних можуть призвести до значних змін в логіці прийняття рішень.

Великі дерева може бути важко інтерпретувати, і рішення, які вони приймають, можуть здатися протилежними інтуїтивно зрозумілим.

Застосування дерев рішень у реальному житті:

Біомедична інженерія (дерева рішень для ідентифікації особливостей, що використовуються в імплантованих пристроях).

Фінансовий аналіз (Задоволення споживача товаром чи послугою).

Астрономія (класифікуйте галактики).

Керування системою.

Виробництво та виробництво (контроль якості, виготовлення напівпровідників тощо).

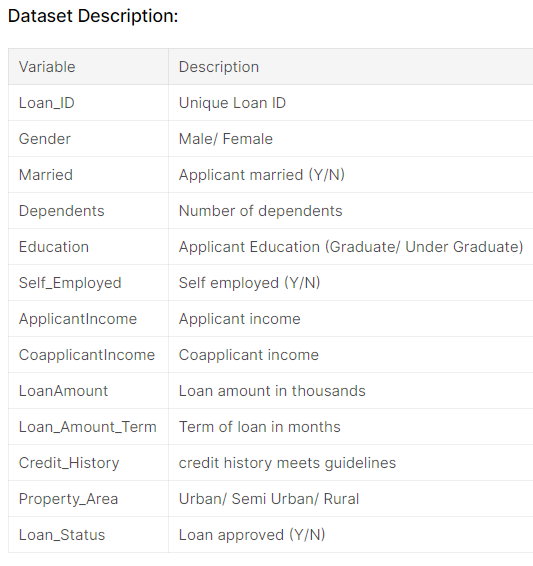
Ліки (діагностика, кардіологія, психіатрія).

Фізика (Виявлення частинок).

# СКЛАД ТА ФОРМАТИ ВИХІДНИХ ДАНИХ ТА РЕЗУЛЬТАТІВ

Вхідними даними буде база даних, що матиме наступні поля

(Дані були взяті з сайту kaggle)



Результатом буде текстовий результат True/False (видавати кредит чи ні).

# СХЕМА ВЗАЄМОДІЇ ПРОГРАМНИХ ЗАСОБІВ

# ЕКРАННІ ФОРМИ

Алгоритм роботи програми:

# 

Вивід результатів

Обрахунок прогнозної можливості клієнта

Дані коректні?

так

ні

Введення вхідних даних

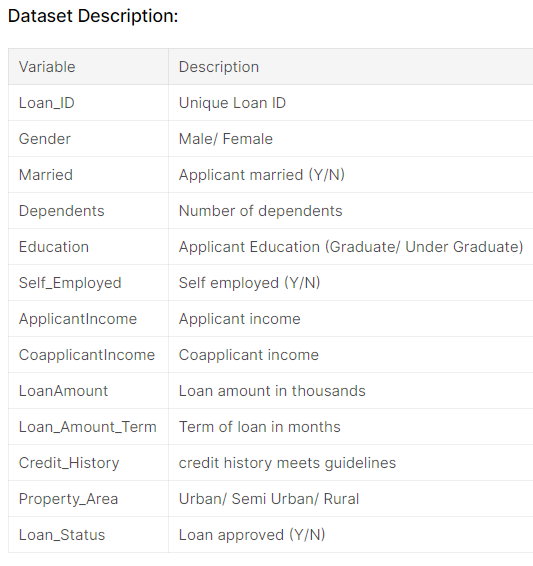
# ВИСНОВОК

так

Сама програма буде представлена в вигляді веб сторінки.

Початкова сторінка буде містити поля для внесення початкових данних, що використовуються в программі

Після натисканні на кнопку Start, користувач перейде на наступну сторінку, яка буде містити форму для заповнення, що матиме наступні поля:



Після заповнення форми, користувач отримає відповідь про кредитоспроможність, він отримає сторінку з написом видавати йому кредит чи ні.

Візуальний вигляд:

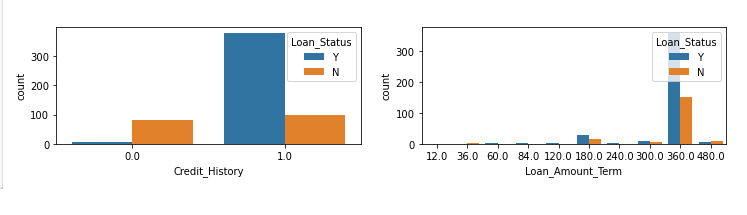
ТАК/НІ

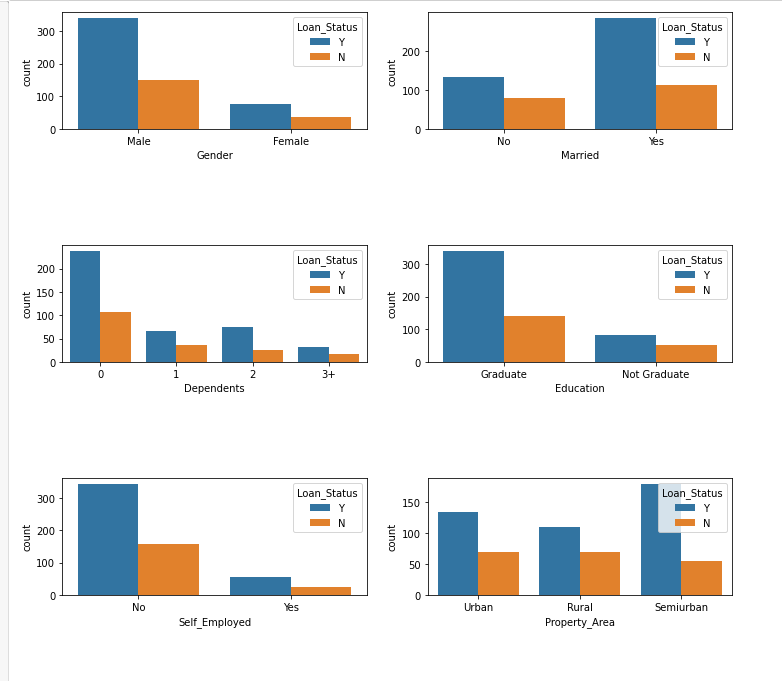
\*Ймовірність, що користувачу потрібно видати кредит

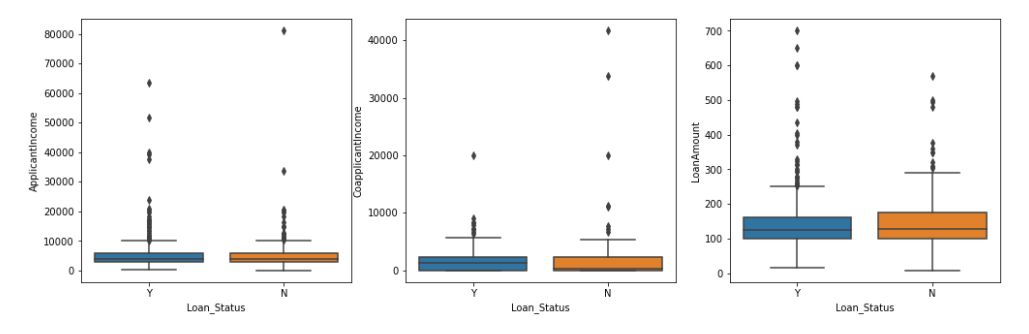
Точність моделі

При натисканні на кнопку «чому була отримана така відповідь?», користувач перейде на наступну сторінку, де будуть відповідні графікі рішення, точність моделі, документація до продукта і тд.

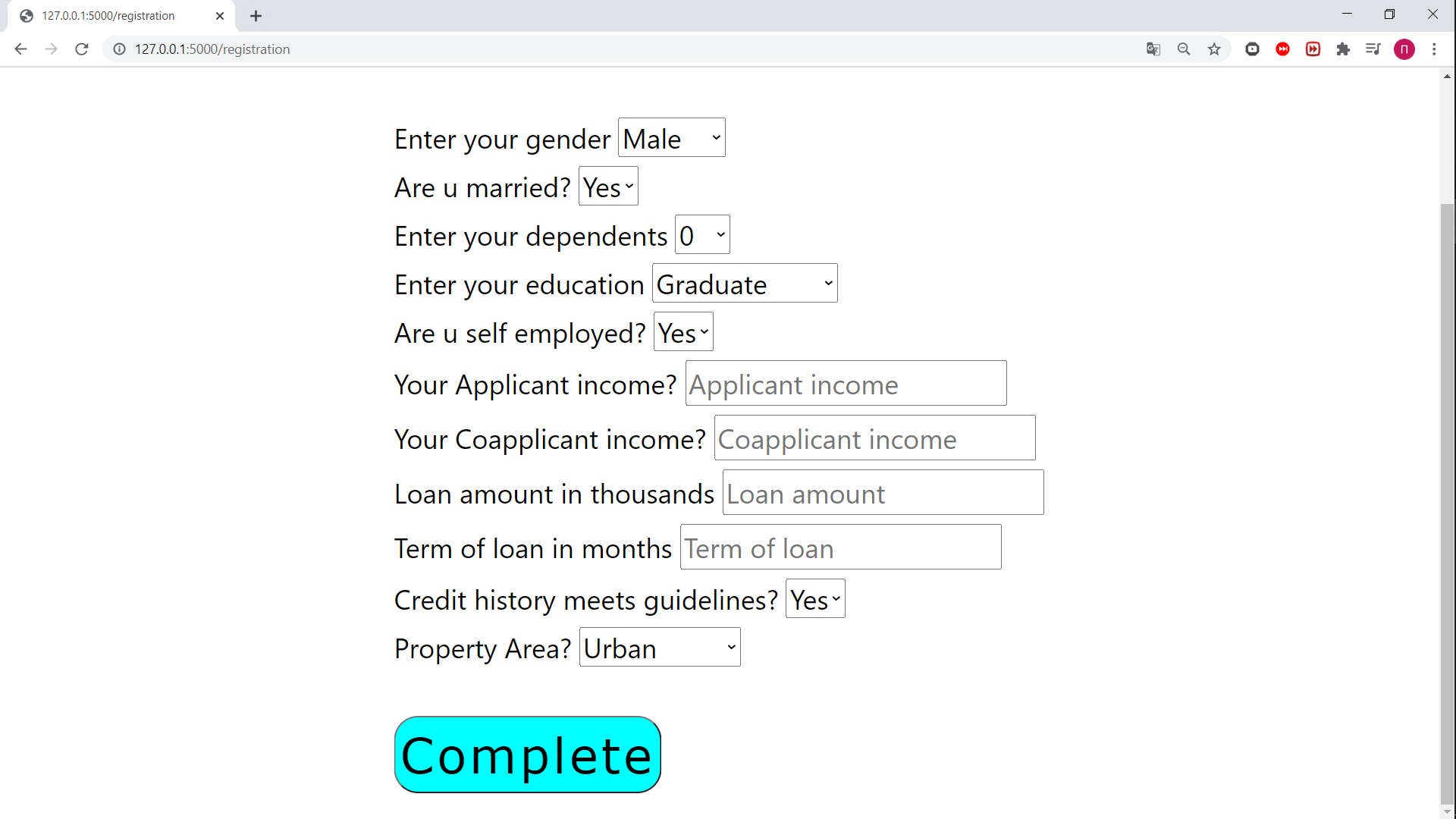
Наближені графікі:

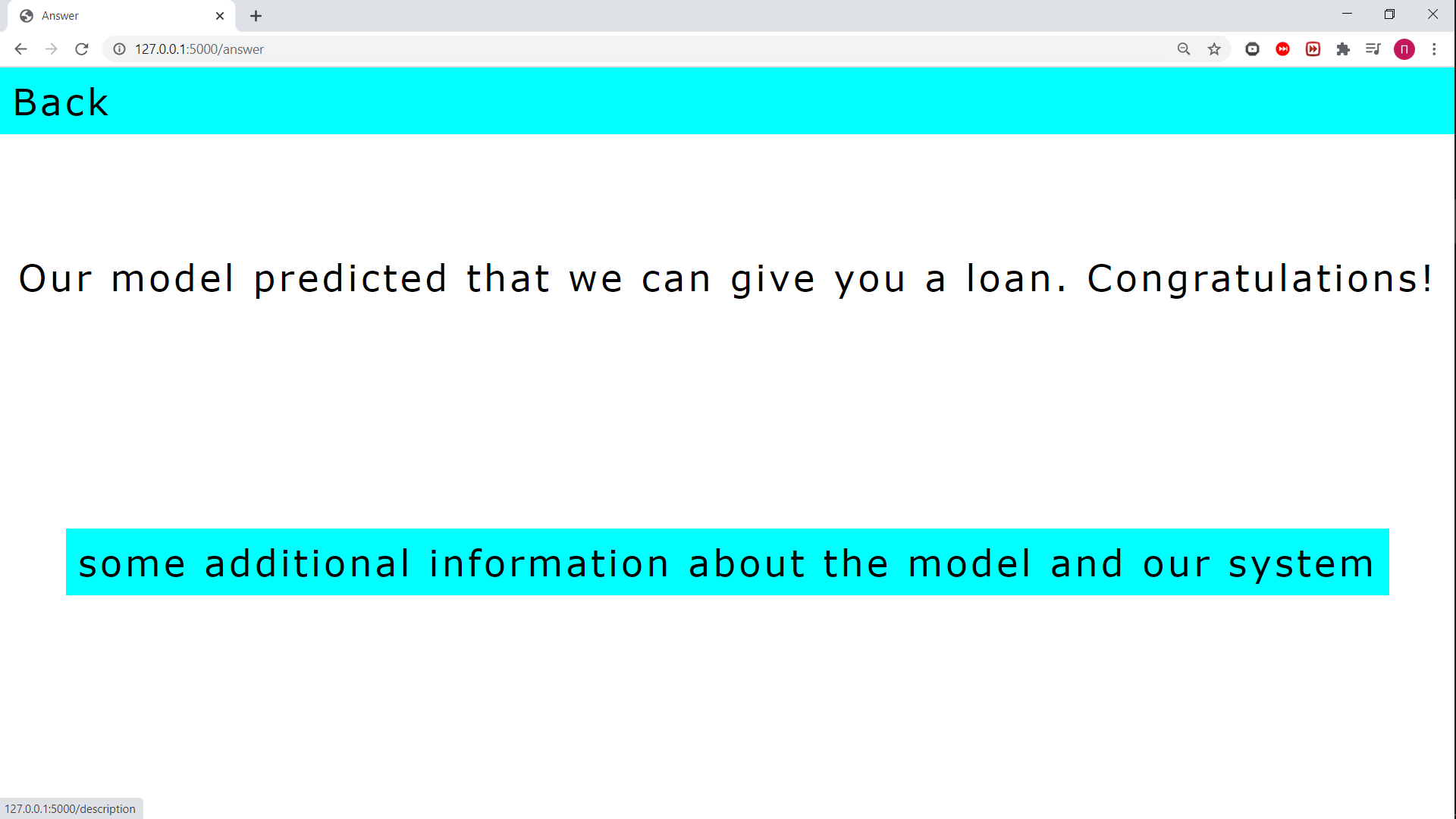


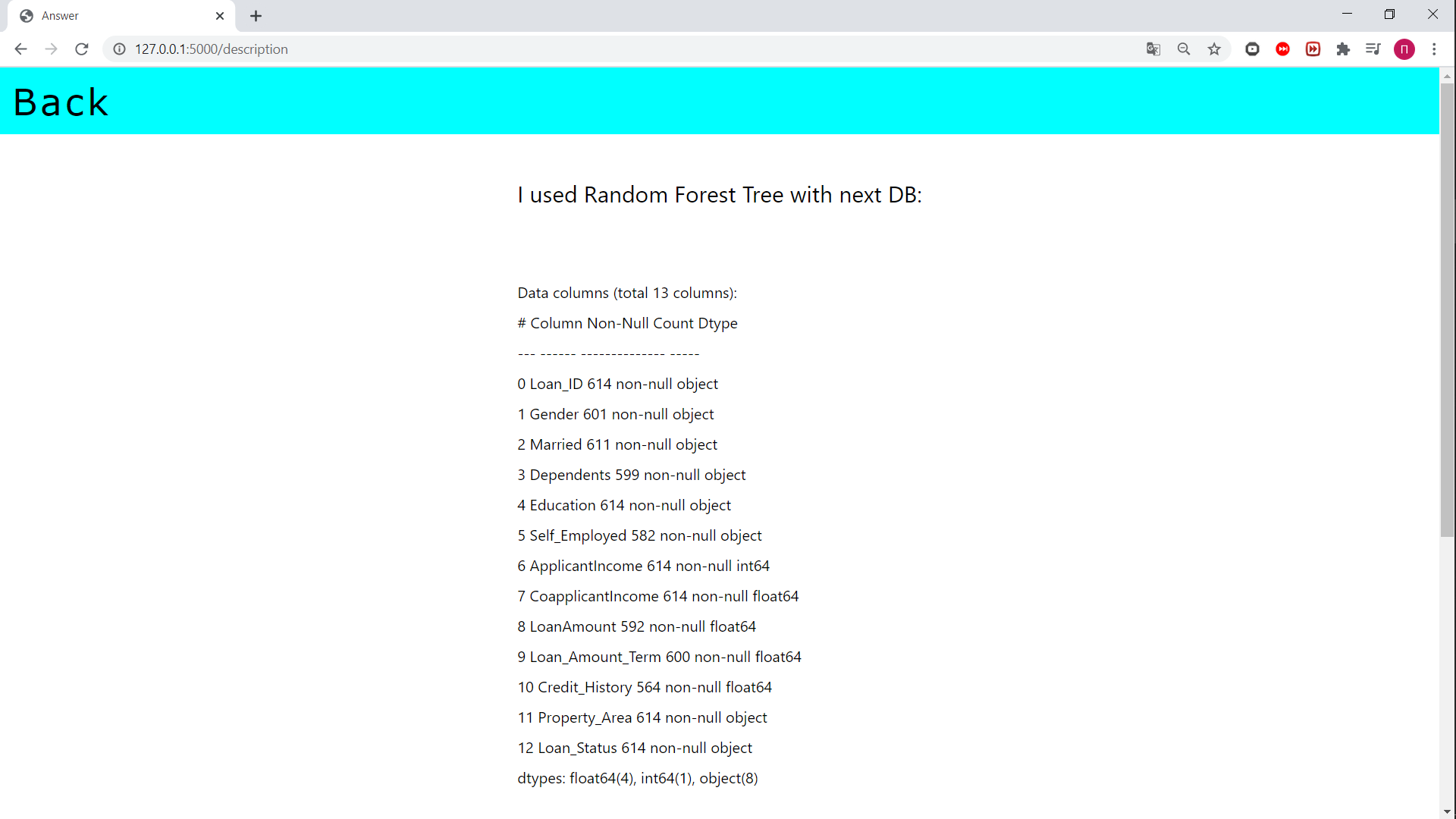


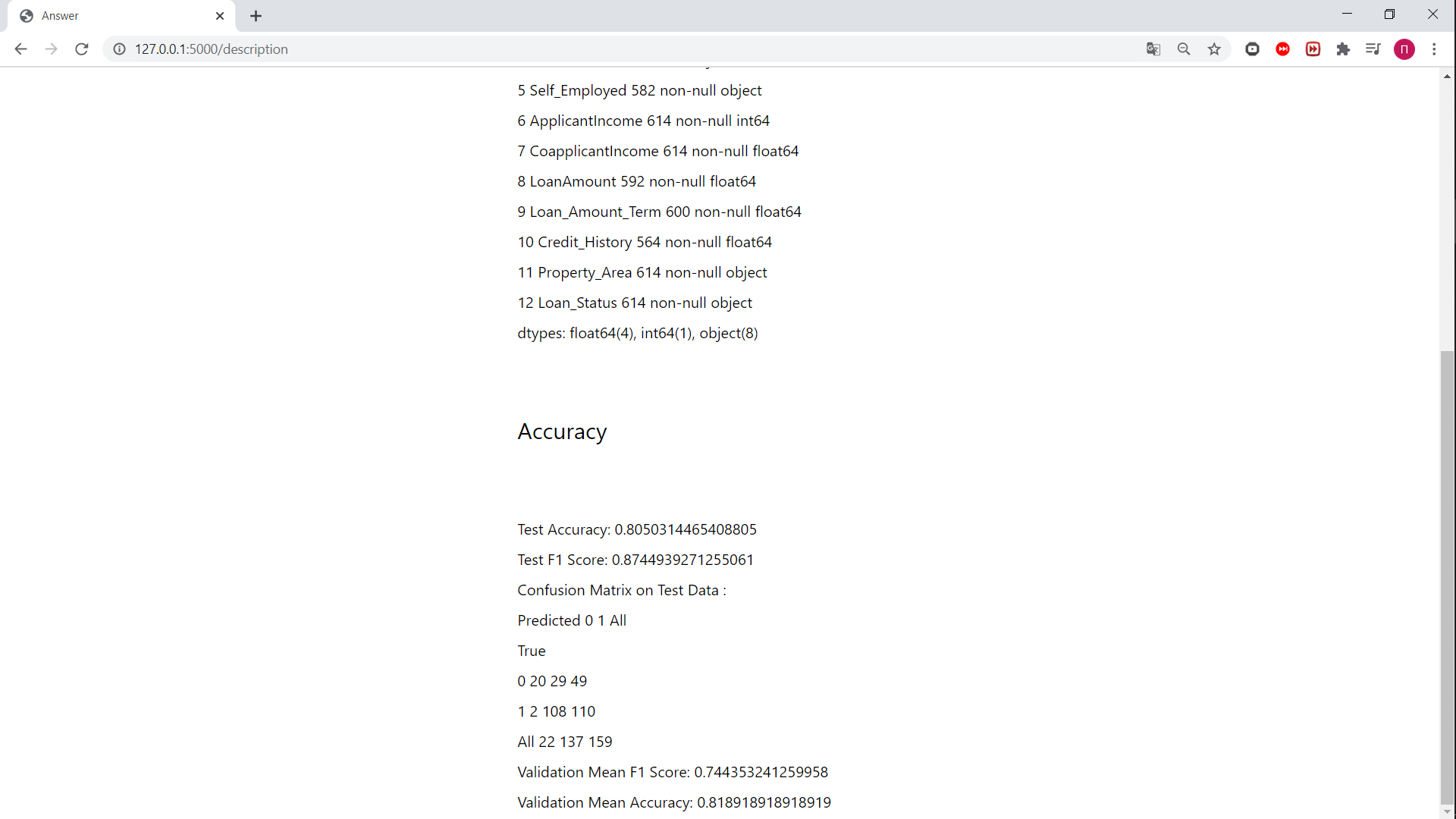


# ВІЗУАЛЬНИЙ ІНТЕРФЕЙС









# ВИСНОВКИ

Для виконання даної розрахунково-графічної роботи було обрано тему

“Створення системи для автоматизації видачі кредиту ”. Для реалізації даної теми було підібрано метод логістичної регресії. Програма представлена у вигляді web-application. Програма написана на мові програмування Python +HTML+CSS. Додаткові технології – flask.

Сервіс має сторінку привітання, сторінку введення інформації, сторінку відповіді про кредитоспроможність та сторінку пояснення результатів. При натисканні на кнопку «документація» користувач перейде на сторінку з документацією, презентацією та основним кодом на гітхабі(https://github.com/PashaLysyi321/credit-scoring-of-the-bank-s-clients).

# СПИСОК ЛІТЕРАТУРИ

1. Васильев, Ф. П. Линейное программирование / Ф.П. Васильев, А.Ю. Иваницкий. - М.: Факториал Пресс, 2016. - 352 c.

2. http://statistica.ru/theory/logisticheskaya-regressiya/

3.https://ru.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F\_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F

4. https://habr.com/ru/company/io/blog/265007/

5.http://www.machinelearning.ru/wiki/index.php?title=%D0%9B%D0%BE%D0%B3%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F\_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F