

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**



**KHAI KHOÁNG DỮ LIỆU**

**Đề tài:**

**XÂY DỰNG WEBSITE  
PHÁT HIỆN SỚM BỆNH TIỂU ĐƯỜNG**

**Các sinh viên thực hiện:**

<b>Tên:</b>	<b>Đỗ Thành Công</b>	<b>Nguyễn Hưng</b>
<b>Mã số:</b>	<b>B1709526</b>	<b>B1709536</b>
<b>Khóa:</b>	<b>43</b>	<b>43</b>

**Cần Thơ, 11/2020**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**



**KHAI KHOÁNG DỮ LIỆU**

**Đề tài:**

**XÂY DỰNG WEBSITE  
PHÁT HIỆN SỚM BỆNH TIỂU ĐƯỜNG**

**Giáo viên hướng dẫn:  
Thầy Lưu Tiến Đạo**

**Cần Thơ, 11/2020**

[illegible]

## LỜI CẢM ƠN

Để có được bài Đồ án này, chúng em xin được bày tỏ lòng biết ơn chân thành và sâu sắc đến thầy Lưu Tiến Đạo, người đã trực tiếp tận tình hướng dẫn, giúp đỡ chúng em trong suốt quá trình thực hiện đồ án, nhờ những sự chỉ bảo và hướng dẫn quý giá đó mà bài đồ án này được hoàn thành một cách tốt nhất.

Cũng xin gửi lời cảm ơn chân thành đến các Thầy Cô Giảng viên Đại học Cần Thơ, đặc biệt là các Thầy Cô ở Khoa CNTT & TT, những người đã truyền đạt cho chúng em những kiến thức quý báu trong thời gian qua.

Tuy có nhiều cố gắng trong quá trình thực hiện đồ án, nhưng không thể tránh khỏi những sai sót. Rất mong nhận được sự đóng góp ý kiến quý báu của quý Thầy Cô và các bạn để bài niên luận hoàn thiện hơn.

Cần Thơ, ngày    tháng 11 năm 2020  
Người viết

Đỗ Thành Công, Nguyễn Hưng

## MỤC LỤC

<b>PHẦN GIỚI THIỆU .....</b>	<b>7</b>
1. Đặt vấn đề .....	7
2. Lịch sử giải quyết vấn đề .....	7
3. Mục tiêu đề tài .....	8
4. Đối tượng và phạm vi nghiên cứu .....	8
5. Phương pháp nghiên cứu .....	8
6. Kết quả đạt được .....	8
7. Bố cục luận văn.....	8
<b>PHẦN NỘI DUNG.....</b>	<b>10</b>
<b>CHƯƠNG 1 .....</b>	<b>10</b>
<b>MÔ TẢ BÀI TOÁN .....</b>	<b>10</b>
1. Mô tả tập dữ liệu.....	10
2. Mô tả chi tiết hệ thống .....	11
<b>2.1 Đối với cá nhân.</b> ....	11
<b>2.2 Đối với nhóm người hay tập thể.</b> .....	11
3 Vấn đề và giải pháp liên quan đến bài toán. ....	12
3.1 Vấn đề tạo file model .....	12
3.2 Vấn đề tiền xử lý dữ liệu. ....	12
3.3 Vấn đề upload file và xuất file kết quả cho một nhóm người. ....	12
3.4 Vấn đề xây dựng API .....	12
<b>CHƯƠNG 2 .....</b>	<b>13</b>
<b>THIẾT KẾ VÀ CÀI ĐẶT .....</b>	<b>13</b>
1. Thiết kế hệ thống .....	13
2. Sơ đồ hệ thống:.....	13
3. Sơ đồ hoạt động.....	13
4. Xây dựng giải thuật.....	15
5. Cài đặt giải thuật.....	15
5.1 Tạo giao diện web bằng Flask .....	15
5.2 Xây dựng model .....	18
5.3 Xây dựng API.....	19
<b>CHƯƠNG 3 .....</b>	<b>21</b>
Kiểm thử và đánh giá.....	21

<b>PHẦN KẾT LUẬN .....</b>	<b>25</b>
1.    Kết quả đạt được .....	25
2.    Hướng phát triển .....	25
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>26</b>

## DANH MỤC HÌNH

Hình 1: Sơ đồ bố cục trang web.....	13
Hình 2: Sơ đồ hoạt động dành đối với personal .....	14
Hình 3: Sơ đồ hoạt động đối với group .....	14
Hình 4: Sơ đồ hoạt động API.....	15
Hình 5: Giao diện trang web.....	16
Hình 6: Nhập thử thông tin Upload file .....	16
Hình 7: Giao diện Upload file.....	17
Hình 8: Upload thử một file.....	17
Hình 9: Lệnh JSON gửi về web service.....	20
Hình 10: Demo trang chủ.....	21
Hình 11: Demo upload file .....	21
Hình 12: Sau khi điền đầy đủ thông tin .....	22
Hình 13: Hiển thị kết quả.....	22
Hình 14: Giao diện upload file mẫu.....	23
Hình 15: Giao diện hoàn thành upload file và download file kết quả .....	23
Hình 16: File csv dự đoán.....	24
Hình 17: Sử dụng API dưới dạng câu lệnh JSON .....	24

## **DANH MỤC BẢNG**

Bảng 1: Mô tả cấu trúc web service API .....	19
Bảng 2: Thông điệp yêu cầu gửi về web service .....	20



## **DANH MỤC BIỂU ĐỒ**

Biểu đồ 1: Tỷ lệ mắc bệnh giữa nam và nữ .....	10
Biểu đồ 2: Tỷ lệ nhãn giữa positive và negative.....	11
Biểu đồ 3: Độ chính xác cho từng chỉ số .....	18

## **TÓM TẮT**

Website phát hiện sớm bệnh tiểu đường được xây dựng dựa trên tập dữ liệu có sẵn từ tập dữ liệu của các bệnh nhân bị tiểu đường được thu thập bởi Bệnh viện Sylhet, Bangladesh. Nhằm phát triển và đánh giá các giải thuật đã học được từ môn Data Mining, nhóm đã xây dựng một website đơn giản nhằm trả về được kết quả dự đoán bệnh từ dữ liệu được nhập vào thông qua các mục đánh giá có trên website.

Đồ án gồm 2 phần: xây dựng model và xây dựng website.

Cách so sánh và phát hiện bệnh từ tập dữ liệu trên là đánh giá mang tính khách quan từ tập dữ liệu, mục đích của website nhằm cung cấp cho môn học Data Mining nên còn nhiều vấn đề chưa được giải quyết.

## PHẦN GIỚI THIỆU

### 1. Đặt vấn đề

Với sự phát triển vượt bậc của xã hội ngày nay việc một cá nhân mắc bệnh tiểu đường không còn là sự xa lạ với ngành y học. Có nhiều nguyên nhân dẫn tới việc mắc bệnh này như : chế độ ăn uống chứa lượng đường cao, một số người béo phì cũng có lượng đường cao trong máu, những người làm việc văn phòng ít vận động cũng là nguyên nhân gián tiếp dẫn đến việc mắc bệnh và còn nhiều nguyên nhân khác.

Nhằm có sự đánh giá khách quan nguy cơ mắc bệnh của cá nhân theo các độ tuổi khác nhau cùng với các yếu tố ảnh hưởng khác nên website dự đoán bệnh tiểu đường được xây dựng để nhằm một mục đích dự đoán cá nhân nào đó có mắc bệnh hay không.

Việc xây dựng website dự đoán bệnh tiểu đường cho ra kết quả là khách quan về hiện trạng cơ thể cá nhân có mắc bệnh hay không, website cũng góp một phần vào ngành y học dự đoán sớm cá nhân có mắc bệnh thông qua form thông tin từ website, kết quả cho ra nhằm giúp cá nhân có thể theo dõi mà đến các trung tâm y tế hay bệnh viện xét nghiệm một cách chính xác hơn để xác định cơ thể có mắc bệnh tiểu đường hay không và hơn hết là nếu có mắc bệnh thì sẽ được các bác sĩ định hướng cho việc cải thiện sức khỏe trong tương lai.

### 2. Lịch sử giải quyết vấn đề

Việc xây dựng website đã không còn xa lạ gì với mọi người, có thể áp dụng rất nhiều cách để xây dựng website, không chỉ trong một mà rất nhiều lĩnh vực khác nhau.

Từ những kiến thức đã có về lập trình web và áp dụng sự hiểu biết vào việc xây dựng website bằng Flask. Flask là một framework của Python hỗ trợ việc xây dựng web nhanh chóng và có thể kết hợp với model xử lý của Machine Learning.

Việc cài đặt và thiết kế yêu cầu người có kiến thức về ngôn ngữ lập trình Python, kiến thức về Machine Learning để xây dựng mô hình có tính hiệu quả và độ chính xác cao áp dụng vào việc xây dựng website.

Hiện tại website đã hoàn thành, có thể hoạt động ổn định nhưng còn các yêu cầu cao hơn từ giáo viên chưa được thực hiện nhưng cũng phần nào cũng đã đạt được những yêu cầu của đề án đề ra.

### 3. Mục tiêu đề tài

Xây dựng được website dự đoán từ các dữ liệu được trích xuất qua form dữ liệu trên giao diện web, áp dụng được các kiến thức học được từ Data Mining và các giải thuật của Machine Learning để giải quyết vấn đề dự đoán bệnh tiểu đường thông qua tập dữ liệu cho trước.

Một phần là website có thể đọc vào file .csv từ người dùng theo chuẩn của dữ liệu dự đoán bệnh tiểu đường (16 trường) và cho ra kết quả dự đoán dựa trên file csv mà người dùng đã cung cấp.

### 4. Đối tượng và phạm vi nghiên cứu

- Nghiên cứu về cách thức hoạt động của trang web.
- Nghiên cứu về thiết kế giao diện trang web.
- Phân tích tập dữ liệu, lựa chọn giải thuật phù hợp.
- Xây dựng website.

### 5. Phương pháp nghiên cứu

- Vẽ sơ đồ mô tả bố cục trang web và sự chuyển giao dữ liệu.
- Sử dụng Flask thiết kế, cài đặt và cấu hình trang web.
- Điền thông tin và hiển thị kết quả.
- Hoàn thiện đồ án.

### 6. Kết quả đạt được

- Website dự đoán bệnh tiểu đường cho ra kết quả chính xác dựa trên tập dữ liệu có sẵn
- Chức năng tải thêm tập tin .csv giúp cơ quan hay tập thể có thể dự đoán nhóm một cách dễ dàng.
- Xây dựng được API cho các lập trình viên trên các nền tảng khác sử dụng.
- Qua đó, có thể thấy website đã tương đối hoàn thành, giúp cho việc dự đoán các bệnh nhân có mắc bệnh tiểu đường hay không cũng nhanh chóng và hiệu quả hơn.

### 7. Bố cục luận văn

#### Phần giới thiệu

Giới thiệu tổng quát về đề tài.

#### Phần nội dung

**Chương 1:** Mô tả bài.

**Chương 2:** Thiết kế, cài đặt giải thuật, biểu diễn cơ sở dữ liệu, trình bày các bước xây dựng hệ thống.

**Chương 3:** Kiểm thử hệ thống và đánh giá độ chính xác, tốc độ của hệ thống.

**Phần kết luận**

Trình bày kết quả đạt được và hướng phát triển hệ thống.

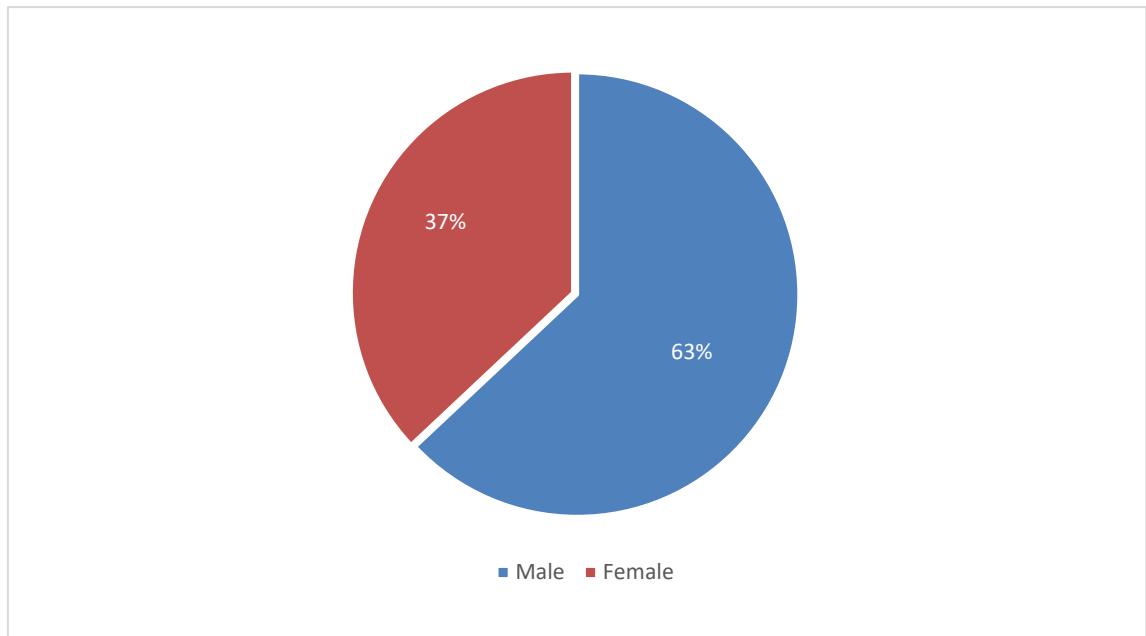
## PHẦN NỘI DUNG

### CHƯƠNG 1

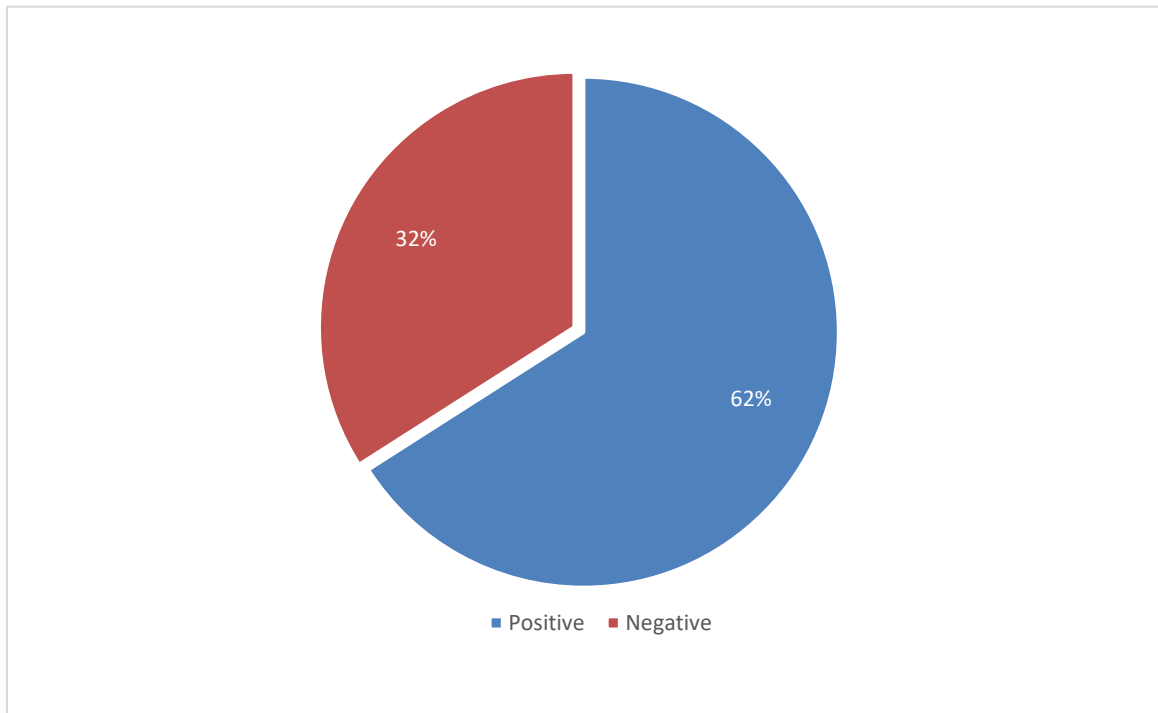
#### MÔ TẢ BÀI TOÁN

##### 1. Mô tả tập dữ liệu

- Tập dữ liệu của các bệnh nhân bị tiểu đường được thu thập bởi Bệnh viện Sylhet, Bangladesh
- Được thu thập bằng cách sử dụng bảng câu hỏi trực tiếp từ các bệnh nhân.
- Tập dữ liệu gồm 16 thuộc tính để dự đoán bệnh nhân có nguy cơ mắc bệnh hay không.
- Sau khi đọc tập dữ liệu ta nhận thấy được sự chênh lệch giữa các nhân



*Biểu đồ 1: Tỷ lệ mắc bệnh giữa nam và nữ*



Biểu đồ 2: Tỷ lệ nhận giữa positive và negative

## 2. Mô tả chi tiết hệ thống

Với mức phát triển vượt bậc của internet hiện nay và nhu cầu sử dụng web của người dân càng tăng cao. Việc xây dựng một trang web dự đoán bệnh có khả năng được truy cập và sử dụng tỉ lệ cao, web dự đoán bệnh cũng đáp ứng được phân dự đoán bệnh cho người truy cập hoặc nhóm người.

Các cá nhân hay người đại diện tập thể có thể kiểm tra tình hình về bệnh tiểu đường của mình hay tập thể thông qua website. Đối với cá nhân thì đơn giản hơn chỉ việc điền form thông tin có sẵn trên giao diện và nhấn Predict để cho ra kết quả của mình. Với tập thể nhóm thì việc này cũng khá phức tạp vì người thu thập hay người đại diện phải trích xuất thông tin thực và điền vào file csv theo tiêu chuẩn 16 trường đánh giá của cơ sở dữ liệu bệnh tiểu đường.

### 2.1 Đối với cá nhân.

- Truy cập trang web.
- Hoàn thành form dữ liệu có sẵn.
- Xác nhận.
- Nhận kết quả trả về ngay trên giao diện.

### 2.2 Đối với nhóm người hay tập thể.

- Thu thập thông tin thực.
- Điền vào file csv.
- Upload lên web thông qua chức năng loadfile.
- Xác nhận.
- Nhận kết quả thông qua file csv tương ứng.

### 3. Vấn đề và giải pháp liên quan đến bài toán.

#### 3.1 Vấn đề tạo file model

- Đọc dữ liệu từ database file csv.
- Chia tập dữ liệu để huấn luyện.
- Chọn tham số criterion hoặc là gini hoặc entropy.
- Chạy vòng lặp cho từng tham số.
- Tính độ chính xác cho giải thuật K láng giềng và Decision Tree.
- So sánh các kết quả tham số và giải thuật chọn giải thuật tốt nhất.

#### 3.2 Vấn đề tiền xử lý dữ liệu.

- Dữ liệu ban đầu có các trường mang theo kiểu Boolean (Yes/No) và Gender là Male và Female
- Chuyển đổi Yes = 1, No = 0, Male = 1, Female = 0 thông qua hàm **replace** được cung cấp bởi thư viện **pandas**.

#### 3.3 Vấn đề upload file và xuất file kết quả cho một nhóm người.

- Lấy file upload từ người dùng thông qua **request.files['filename']** được cung cấp bởi thư viện **flask**.
- Chuyển đổi dữ liệu về chuẩn UTF8-sig.
- Dự đoán nhãn cho từng dòng dữ liệu.
- Sau khi xử lý toàn bộ thì xuất file csv kết quả tương ứng thông qua hàm **to\_csv** và gửi trả file cho người dùng thông qua hàm **make\_response**

#### 3.4 Vấn đề xây dựng API

- Tạo link liên kết để nhận và trả kết quả bằng phương thức POST.
- Web service nhận thông tin dự đoán phải là câu lệnh dưới dạng JSON
- Xử lý dự đoán và trả về kết quả.



## CHƯƠNG 2

### THIẾT KẾ VÀ CÀI ĐẶT

#### 1. Thiết kế hệ thống

Website được tạo nên bằng ngôn ngữ Python thông qua Flask framework. Flask hỗ trợ các hàm có sẵn tương tự các ngôn ngữ lập trình web khác để tạo nên trang web dễ dàng và thuận tiện.

Hệ thống trang web gồm: Điền form thông tin cho cá nhân, upload file cho nhóm người dùng hoặc tập thể và hỗ trợ API cho lập trình viên khác sử dụng.

Đối với cá nhân cần hoàn thành đầy đủ các mục trên form của website và xác nhận để nhận được kết quả, đối với người đại diện cho nhóm người hay tập thể cần tải lên file csv tương ứng sau đó xác nhận và nhận được file kết quả. Đối với lập trình viên khác cần gửi câu lệnh dạng JSON lên sever để nhận về kết quả.

#### 2. Sơ đồ hệ thống:



*Hình 1: Sơ đồ bố cục trang web*

#### 3. Sơ đồ hoạt động

##### Personal:



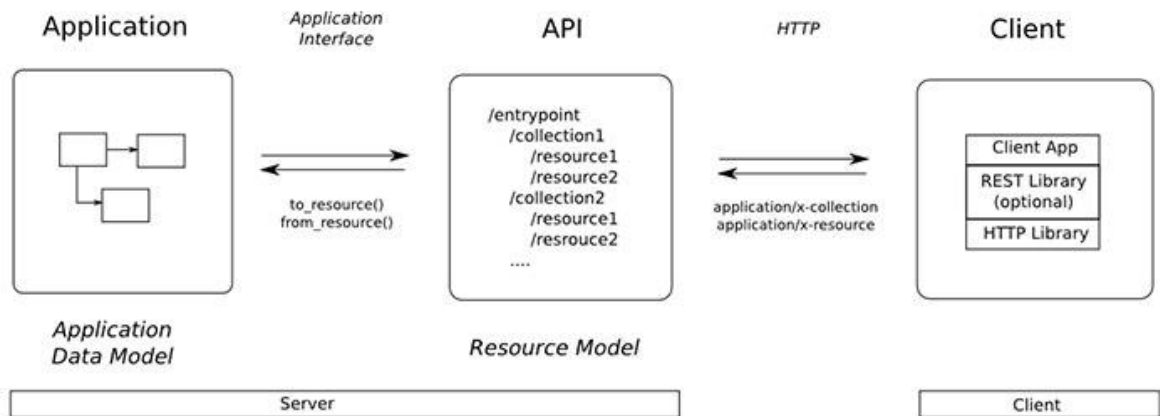
*Hình 2: Sơ đồ hoạt động dành đối với personal*



**Group:**

*Hình 3: Sơ đồ hoạt động đối với group*

## API:



Hình 4: Sơ đồ hoạt động API

## 4. Xây dựng giải thuật

- Personal (cá nhân): Hoàn thành form dữ liệu của web xác nhận và nhận kết quả.
- Group: Upload file csv thu thập được từ thực tế lên hệ thống xác nhận và nhận file csv kết quả tương ứng.
- API: Gửi dữ liệu kiểu JSON lên sever và nhận về kết quả

## 5. Cài đặt giải thuật

### 5.1 Tạo giao diện web bằng Flask

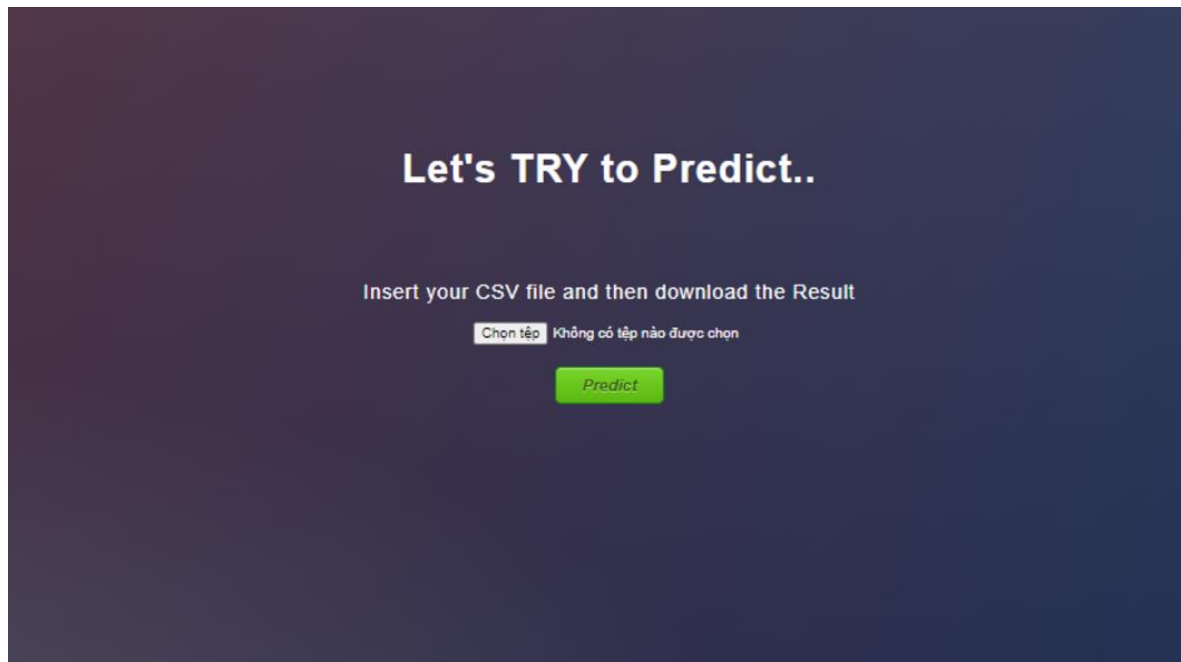
- Dùng ngôn ngữ python và html để tạo giao diện chính của trang, gồm 16 trường dữ liệu cần nhập gồm có:
  - Age: Tuổi tác.
  - Gender: Giới tính.
  - Polyuria: Chu kỳ đi tiểu không ổn định.
  - Polydipsia: Khát quá mức hay uống quá nhiều nước.
  - Sudden weight loss: Giảm cân đột ngột.
  - Weakness: Có không khỏe trong người không.
  - Polyphagia: Mức lượng đường trong máu cao.
  - Genital thrush: Tưa miệng.
  - Visual Blurring: Mờ mắt.
  - Itching: Ngứa.
  - Irritability: Cáu gắt.
  - Delayed healing: Hồi phục chậm sau khi bệnh.

- Partial paresis: Triệu chứng liệt một phần.
  - Muscle stiffness: Cứng cơ.
  - Alopecia: Rụng tóc.
  - Obesity: Béo phì.
- Dùng thẻ input tạo text nhập tuổi, còn lại vì dữ liệu là male, female, yes và no nên dùng select option để tạo sự lựa chọn (hình 4).

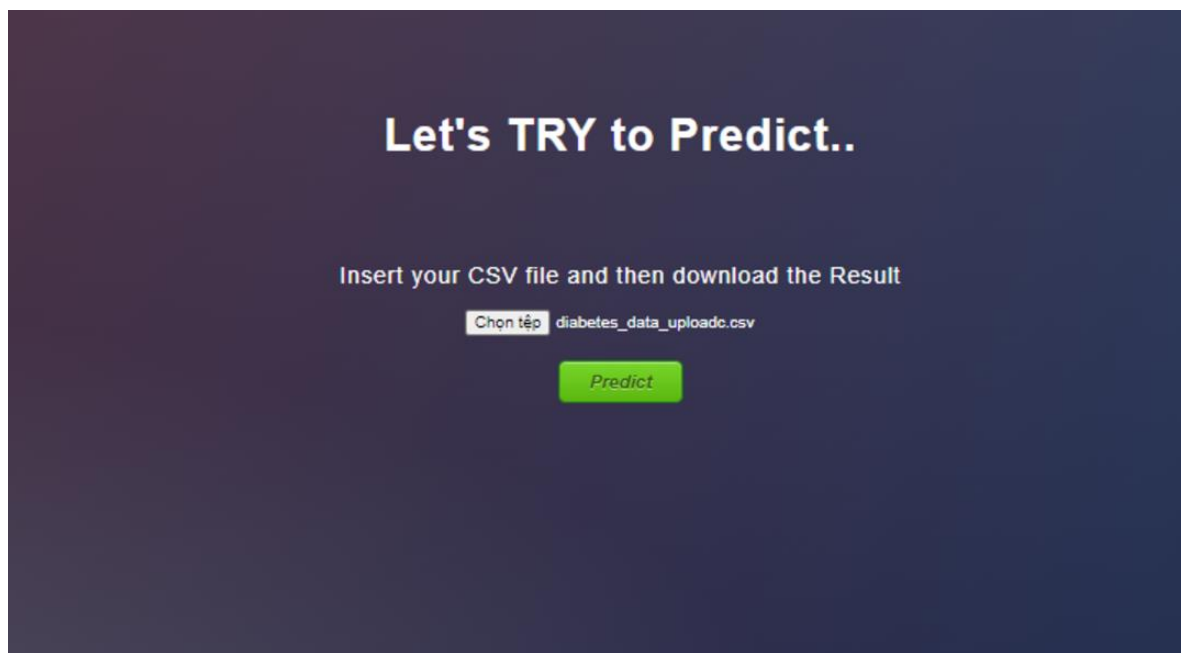
Hình 5: Giao diện trang web

Hình 6: Nhập thử thông tin Upload file

- Từ giao diện trang chủ có button liên kết đến trang upload file csv



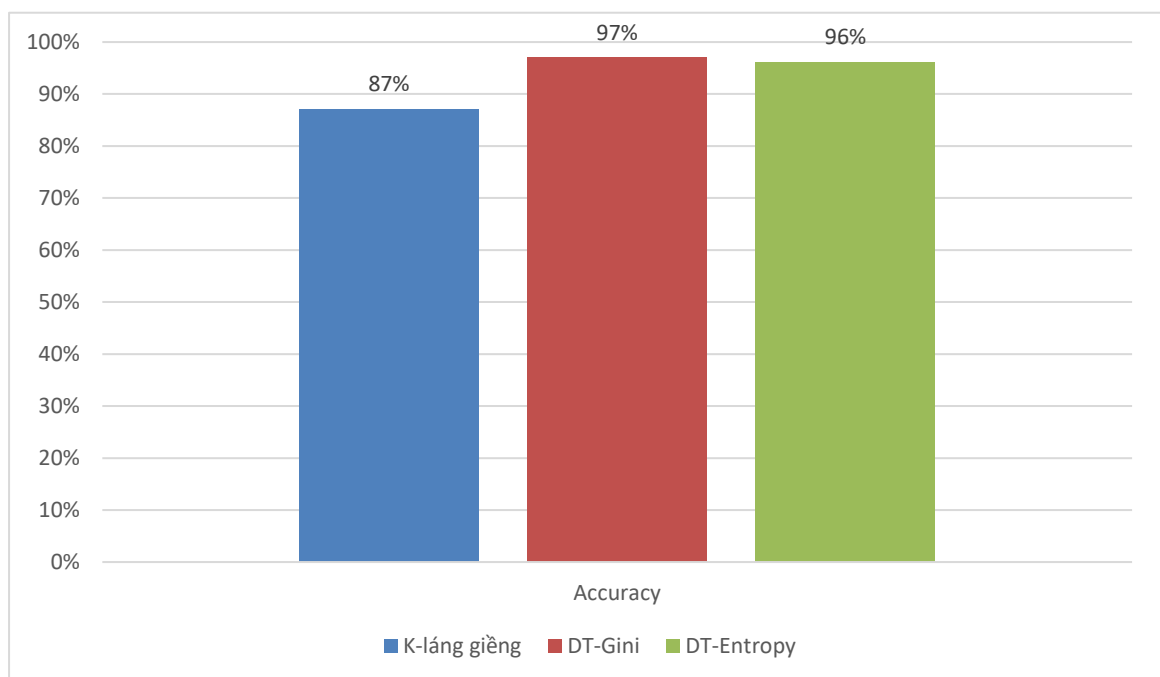
Hình 7: Giao diện Upload file



Hình 8: Upload thử một file

## 5.2 Xây dựng model

- Sau khi tiền xử lý dữ liệu, chọn giải thuật phù hợp cho vấn đề dự đoán để tạo model
- 2 giải thuật được đề xuất:
  - KNeighborsClassifier
  - Decision Tree
- Đối với giải thuật K láng giềng (K-Nearest Neighbor – KNN):
  - Độ chính xác của mô hình sau khi sử dụng K-Neighbors Classifier qua nghi thức kiểm tra k-fold với  $k=10$  là 87%
- Đối với giải thuật Decision Tree:
  - Sử dụng các chỉ số :
    - Độ lợi thông tin (Entropy)
    - Chỉ số Gini
  - Độ chính xác sau khi sử dụng mô hình cây quyết định với 2 tham số trên ta thu được:
    - Entropy: qua nghi thức kiểm tra k-fold là 96%
    - Gini: qua nghi thức kiểm tra k-fold là 97%



Biểu đồ 3: Độ chính xác cho từng chỉ số

- Chọn nFold = 10 vì số lượng phần tử trong tập dữ liệu nhiều hơn 300
  - Theo biểu đồ, sau khi chạy thử các giải thuật, chúng tôi nhận thấy Decision Tree với chỉ số gini cao hơn các giải thuật.
- => Chọn giải thuật Decision Tree với chỉ số gini để tạo Model cho bài toán.

- ❖ Sau khi chạy thuật toán Decision Tree nhận thấy 2 luật đặc trưng từ cây biểu đồ
  - Nếu tuổi nhỏ hơn 34 không mắc các triệu chứng như: Polyuria (Chu kỳ đi tiểu không ổn định), Muscle stiffness (Cứng cơ), Aplopecia: (Rụng tóc) kể cả là nam hay nữ thì kết quả là âm tính
  - Với độ tuổi bất kỳ và có mắc các triệu chứng như: Polyuria (Chu kỳ đi tiểu không ổn định), Polydipsia (Khát quá mức hay uống quá nhiều nước), Polyphagia (Mức lượng đường trong máu cao)

### 5.3 Xây dựng API

Mô tả:

Request URL	https://diabetes-risk-pred.herokuapp.com/
Request Method	POST
Form data	Đối tượng dự đoán bao gồm { "age": {giá trị}, "gender": {giá trị}, "polyuria": {giá trị}, "polydipsia": {giá trị}, "sudden weight loss": {giá trị}, "weakness": {giá trị}, "polyphagia": {giá trị}, "genital thrush": {giá trị}, "visual blurring": {giá trị}, "itching": {giá trị}, "irritability": {giá trị}, "delayed healing": {giá trị}, "partial paresis": {giá trị}, "muscle stiffness": {giá trị}, "aplopecia": {giá trị}, "obesity": {giá trị} }

Bảng 1: Mô tả cấu trúc web service API

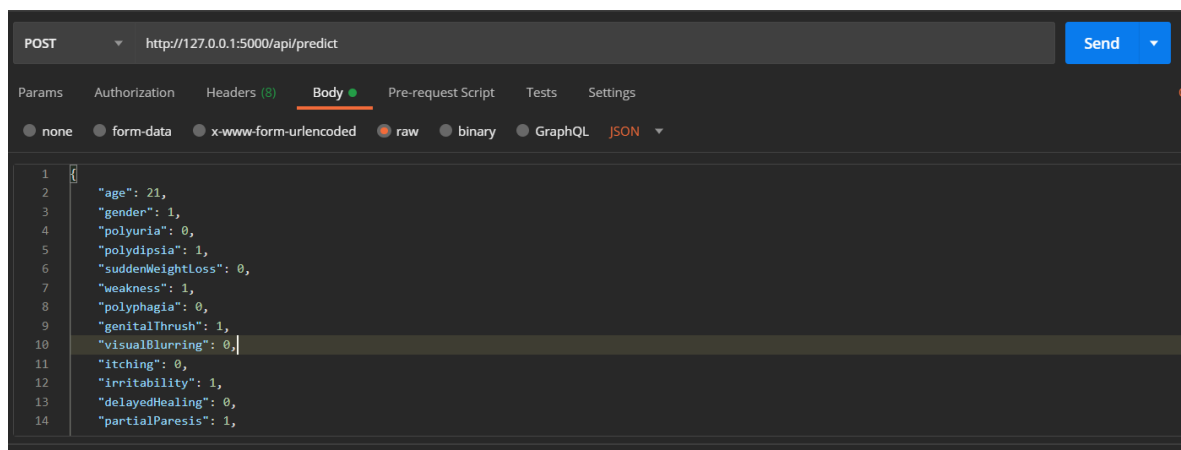
- Thông điệp yêu cầu gửi về Web Service

Tên thuộc tính	Giá trị
Age	Int
Gender	Male (1) or Female (0)
Polyuria	Yes (1) or No (0)
Polydipsia	Yes (1) or No (0)

Sudden weight loss	Yes (1) or No (0)
Weakness	Yes (1) or No (0)
Polyphagia	Yes (1) or No (0)
Genital thrush	Yes (1) or No (0)
Visual Blurring	Yes (1) or No (0)
Itching	Yes (1) or No (0)
Irritability	Yes (1) or No (0)
Delayed healing	Yes (1) or No (0)
Partial paresis	Yes (1) or No (0)
Muscle stiffness	Yes (1) or No (0)
Aplopecia	Yes (1) or No (0)
Obesity	Yes (1) or No (0)

Bảng 2: Thông điệp yêu cầu gửi về web service

- Kết nối Web Service thông qua cổng tiếp nhận: gửi câu lệnh yêu cầu dưới dạng JSON



Hình 9: Lệnh JSON gửi về web service

- Thông điệp trả về:
  - Cấu trúc:
    - Header status:
    - Header Content-Type: application/son; charset=UTF-8
    - Json data
- Trong đó có 1 trong 2 kết quả sẽ được trả về: Positive (dương tính) hoặc Negative (âm tính).



## CHƯƠNG 3

### Kiểm thử và đánh giá

Giao diện trang chủ:

Predict by CSV

### Early stage diabetes risk prediction

Age:	<input type="text" value="Age"/>	Visual Blurring	<input type="text" value="Yes"/>
Gender	<input type="text" value="Male"/>	Itching	<input type="text" value="Yes"/>
Polyuria	<input type="text" value="Yes"/>	Irritability	<input type="text" value="Yes"/>
Polydipsia	<input type="text" value="Yes"/>	Delayed healing	<input type="text" value="Yes"/>
Sudden weight loss	<input type="text" value="Yes"/>	Partial paresis	<input type="text" value="Yes"/>
Weakness	<input type="text" value="Yes"/>	Muscle stiffness	<input type="text" value="Yes"/>
Polyphagia	<input type="text" value="Yes"/>	Alopecia	<input type="text" value="Yes"/>
Genital thrush	<input type="text" value="Yes"/>	Obesity	<input type="text" value="Yes"/>

Predict

Hình 10: Demo trang chủ

### Let's TRY to Predict..

Insert your CSV file and then download the Result

Chọn tệp Không có tệp nào được chọn

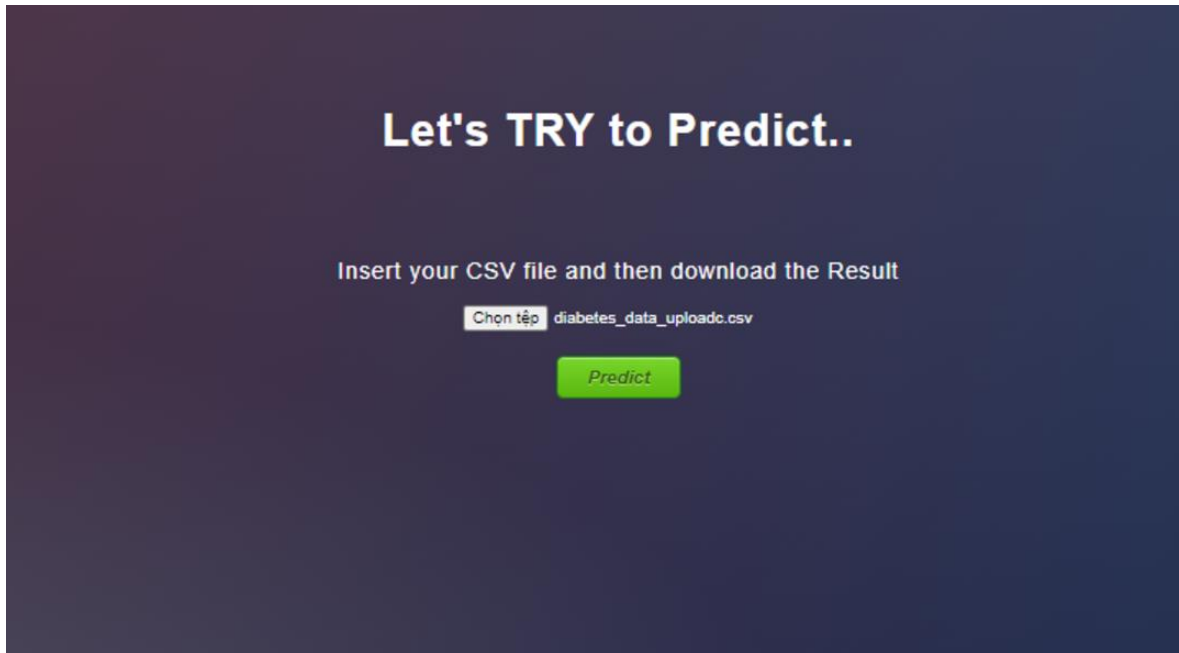
Predict

Hình 11: Demo upload file

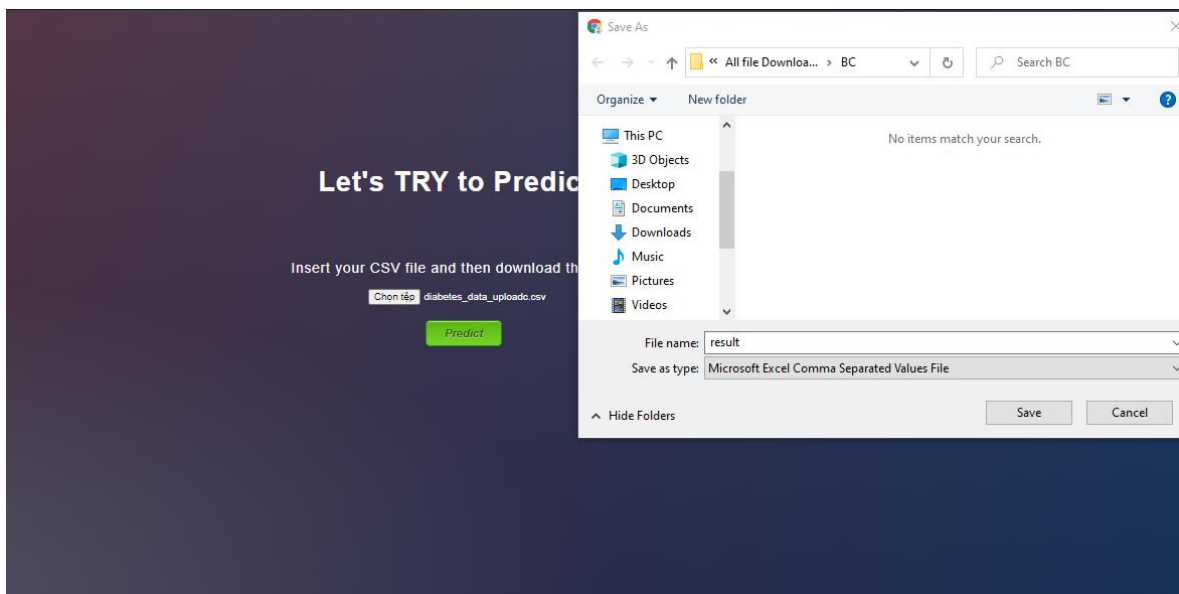
Giao diện của trang web sau khi cung cấp thông tin đầy đủ của cá nhân:

Hình 12: Sau khi điền đầy đủ thông tin

Hình 13: Hiện thị kết quả



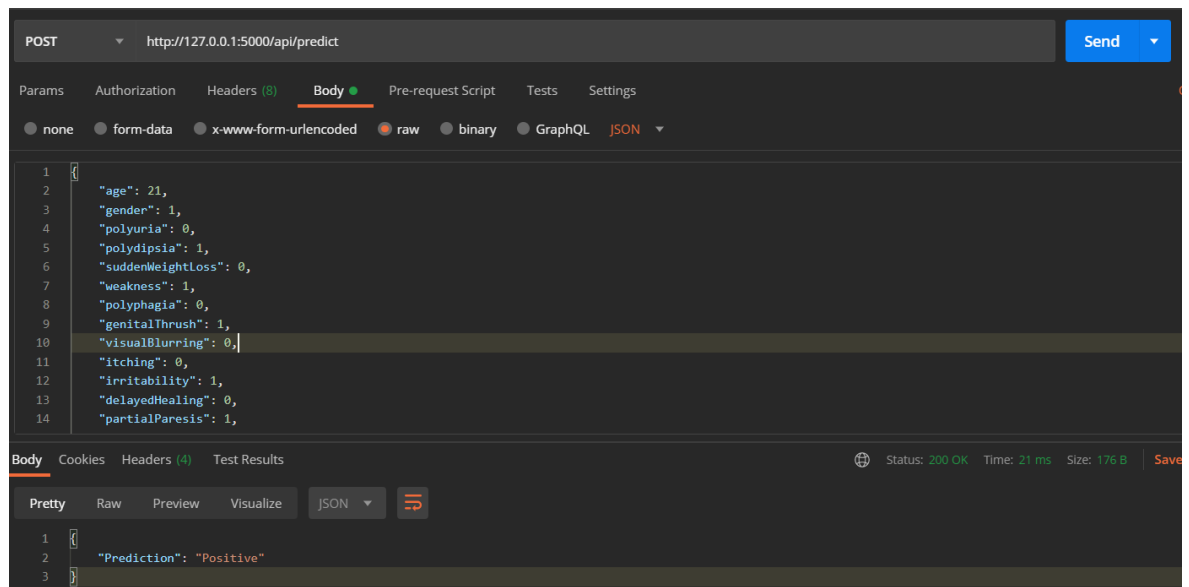
Hình 14: Giao diện upload file mẫu



Hình 15: Giao diện hoàn thành upload file và download file kết quả

41	26	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	['Negative']
42	60	1	0	0	0	0	1	1	0	1	1	0	1	0	1	0	1	1	1	0	['Negative']
43	65	1	0	0	0	0	1	0	0	1	1	0	1	1	1	1	1	1	1	0	['Negative']
44	72	1	0	0	0	0	0	1	0	1	1	0	1	1	0	1	1	1	1	0	['Negative']
45	30	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	['Negative']
46	45	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	['Negative']
47	65	1	1	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	['Negative']
48	70	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	1	1	1	0	['Negative']
49	35	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	['Negative']
50	54	1	0	0	0	0	1	0	0	0	1	0	1	0	1	0	1	1	1	0	['Negative']
51	30	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	['Negative']
52	46	1	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	1	0	['Negative']
53	53	1	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1	1	1	0	['Negative']
54	42	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	['Negative']
55	55	0	1	1	1	1	1	0	0	1	0	0	0	0	0	1	0	0	0	1	['Positive']
56	48	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	['Positive']
57	55	0	1	1	1	1	0	1	0	0	1	0	1	0	1	1	0	0	0	0	['Positive']
58	39	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	0	0	['Positive']
59	43	0	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	1	0	['Positive']
60	35	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	['Positive']
61	47	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	['Positive']
62	50	0	1	1	0	1	1	1	0	0	0	0	1	1	1	0	0	0	0	0	['Positive']
63	48	0	1	1	0	1	0	0	0	1	1	0	1	1	1	0	0	0	0	0	['Positive']
64	35	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	0	['Positive']

Hình 16: File csv dự đoán



Hình 17: Sử dụng API dưới dạng câu lệnh JSON

## PHẦN KẾT LUẬN

### 1. Kết quả đạt được

Website dự đoán bệnh tiểu đường sau khi đã hoàn thành đã được những kết quả sau:

- Cách thức cài đặt và sử dụng tương đối đơn giản, dễ dàng thông qua FLask.
- Có thể dự đoán được hầu hết các dữ liệu nhập vào từ form dữ liệu trên giao diện web.
- Có thể load file.csv tùy nhu cầu của cá nhân hay tập thể.
- Xây dựng được API cho các lập trình viên trên các nền tảng khác sử dụng

### 2. Hướng phát triển

Về sự phát triển của website sẽ ngày càng hoàn thiện, có thêm API có thể cung cấp cho người dùng trực tiếp sử dụng trên web thông qua máy chủ có sẵn như :

- Không ngừng tìm kiếm, đổi mới các phương thức đánh giá, tiện lợi hơn.
- Tìm kiếm khắc phục nhược điểm xử lý khi người dùng bỏ trống 1 thuộc tính nào đó của website. Đưa ra hướng giải quyết và khắc phục hiệu quả nhất
- Website được tối ưu hóa sao cho có thể nâng cấp và mở rộng trong tương lai với nhiều chức năng mới, phức tạp. Đảm bảo được tính tiết kiệm và hiệu quả cao.

## TÀI LIỆU THAM KHẢO

1. Giáo trình Khai mở dữ liệu (Data Mining) – PGS.TS. Đỗ Thanh Nghị.
2. Giới thiệu *Flask*  
<https://flask.palletsprojects.com/en/1.1.x/>
3. Tập dữ liệu về dự đoán bệnh tiểu đường  
[https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset.](https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset)