**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**ĐỒ ÁN HỌC PHẦN**

TÊN HỌC PHẦN: **Lập trình Python (PYPRO)**

MÃ SỐ LỚP HP: **IPPA233277\_05CLC**

Tên chủ đề EDA: **EDA PHÂN TÍCH ELECTRIC VEHICLE POPULATION TRONG QUÝ 2 NĂM 2023**

Chủ đề CV: **LẬP TRÌNH PYTHON NHẬN DIỆN KHUÔN MẶT TRONG VIDEO**

Chủ đề Game: **LẬP TRÌNH PYTHON GAME CAR RACING**

**Họ tên sinh viên: ĐINH ĐẠI HẢI ĐĂNG**

**Mã số sinh viên: 21110164**

**Lớp: 21110CL3**

**Ngày nộp: 28/11/2023**

**Ký tên:**

**TP.HCM, ngày 28 Tháng 11 năm 2023**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**KHOA [CÔNG NGHỆ THÔNG TIN|CHẤT LƯỢNG CAO]**

**ĐỀ TÀI ĐỒ ÁN HỌC PHẦN**

TÊN HỌC PHẦN: **Lập trình Python (PYPRO)**

MÃ SỐ LỚP HP: **IPPA233277\_05CLC**

Tên chủ đề EDA: PHÂN TÍCH ELECTRIC VEHICLE POPULATION TRONG QUÝ 2 NĂM 2023

Chủ đề CV: LẬP TRÌNH PYTHON NHẬN DIỆN KHUÔN MẶT TRONG VIDEO

Chủ đề Game: LẬP TRÌNH PYTHON GAME CAR RACING

**Giảng viên giảng dạy: VÕ XUÂN THỂ**

**Họ tên sinh viên: ĐINH ĐẠI HẢI ĐĂNG**

**Mã số sinh viên: 21110164**

**Lớp: 21110CL3**

**TÊN CÁC FILE SẢN PHẨM ĐỀ TÀI:**

Thư mục bài làm: **G5C12DinhDaiHaiDang\_Topic**

Tập tin báo cáo bài làm: **2\_ G5C12DinhDaiHaiDang \_Topic.docx**

Tập tin mã nguồn: **G5C12DinhDaiHaiDang \_Topic.py**

Thư mục bài làm phần EDA: **G5C12DinhDaiHaiDang \_EDA.Topic**

Tập dữ liệu thực nghiệm EDA: 12DinhDaiHaiDang\_ElectricVehiclePopulationData (nên sử dụng Keysearch)

Thư mục bài làm phần CV: **G5C12DinhDaiHaiDang \_CV.Topic**

File Video thực nghiệm CV: 1.mp4 (nên sử dụng Keysearch)

Thư mục bài làm phần Game: **G5C12DinhDaiHaiDang n\_Game.Topic**

Giới thiệu thư mục bài làm: **1\_ G5C12DinhDaiHaiDang \_Intro.docx**

**Tập các thư viện sử dụng:** # Speech: speech\_recognition, gtts, playsound

# GUI: tkinter

# EDA: pandas, numpy, scipy, sklearn

**NHẬN XÉT CỦA GIẢNG VIÊN**

**………………………………………………………..**

**………………………………………………………..**

**Ngày……./……../202…..**

**Giảng viên: Ký tên**

**TP.HCM, ngày 28 tháng 11 năm 2023**

**LỜI CẢM ƠN**

H

# DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |
| --- | --- |
| CSDL hoặc DB | Cơ sở dữ liệu: DataBase |
| NSD = Users = Account | Người Sử Dụng = là quyền làm việc trên WebApp được đảm bảo bởi tối thiểu là 2 yếu tố: tên đăng nhập (UserName) và mật khẩu (Password) |
|  |  |
| AI | Artificial intelligence: Trí tuệ nhân tạo |
| EDA | Exploratory Data Analysis: Phân tích Khám phá Dữ liệu (Còn gọi là: phân tích dữ liệu thăm dò) |
| GUI | Graphical User Interface: Giao diện đồ hoạ |
| PYPRO | Lập trình Python |

**DANH MỤC CÁC THUẬT NGỮ**

|  |  |
| --- | --- |
| Giao diện người dùng:  User-Interface | Là hệ thống các màn hình giao tiếp cho phép người sử dụng tương tác với các thành phần phần mềm trong HTTT, thường là 1 trong các dong: Win-form, Web-form, Mobile-Form. |
| Tài khoản (Account) | Là một quyền làm việc trên HTTT được cấp phát cho một cá nhân thông qua tên tài khoản (username) và mật khẩu (password). |
| Trợ lý ảo: Voice Assistant | Là các hệ thống có khả năng “nghe” và “nói” với con người, nhờ đó hỗ trợ con người trong một số chức năng như một “trợ lý”. |

**MỤC LỤC**

[**LỜI CẢM ƠN** 3](#_Toc151453978)

[DANH MỤC CÁC TỪ VIẾT TẮT 4](#_Toc151453979)

[**DANH MỤC CÁC THUẬT NGỮ** 4](#_Toc151453980)

[Chương 1: GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN 8](#_Toc151453981)

[1.1. Tổng quan về Đồ án học phần 8](#_Toc151453982)

[1.2. Nội dung chuyên môn chính của Đồ án học phần 8](#_Toc151453983)

[1.3. Công cụ và nền tảng kỹ thuật thực hiện Đồ án học phần 9](#_Toc151453984)

[1.4. Bố cục của báo cáo 9](#_Toc151453985)

[Chương 2: CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN 11](#_Toc151453986)

[2.1. GIỚI THIỆU VỀ LẬP TRÌNH PYTHON 11](#_Toc151453987)

[2.1.1. Giới thiệu chung 11](#_Toc151453988)

[2.1.2. Cài đặt “bộ thảo chương” (soạn thảo chương trình IDE) và “bộ dịch” 11](#_Toc151453989)

[2.1.3. Giới thiệu các bộ tương tác lập trình 11](#_Toc151453990)

[2.1.4. Cấu hình mô trường tương tác lập trình 11](#_Toc151453991)

[2.2. TỔNG QUAN VỀ NGÔN NGỮ LẬP TRÌNH PYTHON 11](#_Toc151453992)

[2.2.1. Danh hiệu (Identifier) 11](#_Toc151453993)

[2.2.2. Về viết mã lệnh lập trình (Code) 11](#_Toc151453994)

[2.2.3. Về chú giải (Comment) 11](#_Toc151453995)

[2.2.4. Các phép toán 11](#_Toc151453996)

[2.2.5. Chuỗi 11](#_Toc151453997)

[2.2.6. Lệnh 11](#_Toc151453998)

[2.2.7. Định nghĩa hàm 11](#_Toc151453999)

[2.2.8. Danh sách 11](#_Toc151454000)

[2.2.9. Bộ dữ liệu: Tuple 11](#_Toc151454001)

[2.2.10. Tập hợp: Set 11](#_Toc151454002)

[2.2.11. Từ điển: Dictionary 11](#_Toc151454003)

[2.2.12. Module 11](#_Toc151454004)

[2.2.13. Package 11](#_Toc151454005)

[2.2.14. Lớp 11](#_Toc151454006)

[2.2.15. Xử lý ngoại lệ 11](#_Toc151454007)

[2.3. GIỚI THIỆU CÁC THƯ VIỆN VÀ CÔNG CỤ CĂN BẢN 11](#_Toc151454008)

[2.3.1. Nạp thư viện 11](#_Toc151454009)

[2.3.2. speech\_recognition 11](#_Toc151454010)

[2.3.3. gtts 11](#_Toc151454011)

[2.3.4. playsound 11](#_Toc151454012)

[2.3.5. tkinter 11](#_Toc151454013)

[2.3.6. os 11](#_Toc151454014)

[2.3.7. Pandas 11](#_Toc151454015)

[2.3.8. Tensorflow 11](#_Toc151454016)

[2.3.9. NumPy 12](#_Toc151454017)

[2.3.10. SCIPY 12](#_Toc151454018)

[2.3.11. Theano 12](#_Toc151454019)

[2.3.12. Matplotlib 12](#_Toc151454020)

[2.3.13. scikit-learn = sklearn 12](#_Toc151454021)

[2.3.14. Keras 12](#_Toc151454022)

[2.3.15. NLTK 12](#_Toc151454023)

[2.3.16. PyTorch 12](#_Toc151454024)

[2.3.17. LightGBM 12](#_Toc151454025)

[2.3.18. Eli5 12](#_Toc151454026)

[2.4. KỸ THUẬT LẬP TRÌNH ĐỆ QUY: THÁP HÀ NỘI (HANOI TOWER) 12](#_Toc151454027)

[2.5. LẬP TRÌNH PYTHON XỬ LÝ GIỌNG NÓI (Trợ lý ảo: Voice Assistant) 12](#_Toc151454028)

[2.5.1. Giới thiệu chủ đề 12](#_Toc151454029)

[2.5.2. Tài liệu và nguồn thực nghiệm 12](#_Toc151454030)

[2.5.3. Giới thiệu Các thư viện Python liên quan 12](#_Toc151454031)

[2.5.4. Giới thiệu bài toán 12](#_Toc151454032)

[2.5.5. Giới thiệu code tham khảo 12](#_Toc151454033)

[2.6. LẬP TRÌNH GIAO DIỆN ĐỒ HỌA (GUI) 12](#_Toc151454034)

[2.7. LẬP TRÌNH PHÂN TÍCH KHÁM PHÁ (EDA) [Thăm dò] 12](#_Toc151454035)

[2.8. LẬP TRÌNH ĐỒ THỊ & BIỂU ĐỒ TRỰC QUANG HÓA DỮ LIỆU (VISUALIZAION: PLOT) 12](#_Toc151454036)

[2.9. LẬP TRÌNH THỊ GIÁC MÁY TÍNH (CV) 12](#_Toc151454037)

[2.10. LẬP TRÌNH GAMEs CĂN BẢN 12](#_Toc151454038)

[Chương 3: PHÂN TÍCH VÀ XÁC ĐINH CÁC CƠ SỞ KỸ THUẬT 13](#_Toc151454039)

[3.1. GIỚI THIỆU CHUNG VỀ CÁC CHỦ ĐỀ CỦA ĐỒ ÁN 13](#_Toc151454040)

[3.2. LẬP TRÌNH XỬ GIỌNG NÓI (TRỢ LÝ ẢO: VOICE ASSISTANT) 13](#_Toc151454041)

[3.2.1. Giới thiệu chung về chủ đề 13](#_Toc151454042)

[3.2.1.1. Mô tả yêu cầu của chủ đề 13](#_Toc151454043)

[3.2.1.2. Phân tích yều cầu của chủ đề 13](#_Toc151454044)

[3.2.2. Các nền tảng kỹ thuật 13](#_Toc151454045)

[3.2.2.1. Nền tảng kỹ thuật cơ bản 13](#_Toc151454046)

[3.2.2.2. Hệ thống thư việc sử dụng 13](#_Toc151454047)

[3.2.3. Thiết kế giao diện 14](#_Toc151454048)

[3.2.4. Mã lệnh lập trình (python) 14](#_Toc151454049)

[3.3. LẬP TRÌNH GIAO DIỆN ĐỒ HỌA (GUI) 15](#_Toc151454050)

[3.4. LẬP TRÌNH TRỰC QUANG HÓA DỮ LIỆU (PLOT) 15](#_Toc151454051)

[Chương 4: LẬP TRÌNH PYTHON ELECTRIC VEHICLE POPULATION TRONG QUÝ 2 NĂM 2023 15](#_Toc151454052)

[4.1. Giới thiệu chủ đề EDA ……. 15](#_Toc151454053)

[3.4.1. Mô tả yêu câu của chủ đề 15](#_Toc151454054)

[3.4.2. Phân tích yều cầu của chủ đề 16](#_Toc151454055)

[3.5. CÁC NỀN TẢNG KỸ THUẬT 16](#_Toc151454056)

[3.5.1. Hệ thống thư viện sử dụng 16](#_Toc151454057)

[3.5.2. Tập dữ liệu thực nghiệm 16](#_Toc151454058)

[3.5.3. Mô tả giải thuật 17](#_Toc151454059)

[4.2. Thiết kế giao diện 19](#_Toc151454060)

[4.3. Mã lệnh lập trình (python) 19](#_Toc151454061)

[Chương 5: LẬP TRÌNH PYTHON NHẬN DIỆN KHUÔN MẶT TRONG VIDEO. 23](#_Toc151454062)

[Chương 6: LẬP TRÌNH PYTHON CAR RACING 23](#_Toc151454063)

[Chương 7: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN 23](#_Toc151454064)

[7.1. Kết luận 23](#_Toc151454065)

[7.1.1. Những kết quả đạt được 23](#_Toc151454066)

[7.1.2. Hạn chế 23](#_Toc151454067)

[7.2. Hướng phát triển 23](#_Toc151454068)

[7.2.1. Hướng khắc phục các hạn chế 23](#_Toc151454069)

[7.2.2. Hướng mở rộng ĐỒ ÁN HỌC PHẦN 23](#_Toc151454070)

[DANH MỤC TÀI LIỆU THAM KHẢO 24](#_Toc151454071)

1. **GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN**

## Tổng quan về Đồ án học phần

Đồ án Học phần vận dụng kiến thức và kỹ năng về Lập trình Python để xây dựng giao diện đồ hoạ (GUI) cho phép: tương tác với người dùng bằng giọng nói (Trợ lý ảo: Voice Assistant) với các chủ đề:

+ Phân tích thăm dò (EDA: Epolore Data Analysing) ELECTRIC VEHICLE POPULATION tại Hoa Kỳ trong Quý 2 Năm 2023 đối với tập dữ liệu thực nghiệm theo link: https://www.kaggle.com/datasets/willianoliveiragibin/electric-vehicle-population; có trực quang hóa dữ liệu dùng các kỹ thuật biểu đồ và đồ thị (Plot)

+ CV Nhận Diện Khuôn Mặt Trong Video

+ Games Car Racing

## Nội dung chuyên môn chính của Đồ án học phần

Bao gồm 6 nội dung lập trình chính dùng ngôn ngữ lập trình Python:

+ Trợ lý ảo: . thông tin nhập có 2 phương án (lời nói dạng “chọn” OR nhập ký tự)

. tin xuất “kết quả thăm dò” ở cả 2 dạng: text và lời nói

+ Phân tích dữ liệu khám phá || thăm dò [EDA] ELECTRIC VEHICLE POPULATION tại Hoa Kỳ trong Quý 2 Năm 2023

+ Trực quang hóa dữ liệu thăm dò ELECTRIC VEHICLE POPULATION bằng biểu đồ | đồ thi.

+ Giao diện đồ họa dạng [WinForm or Webform or Mobileform]

+ CV: Thị giác máy tính cho phép nhận diện khuôn mặt trong video

+ Game căn bản Car Racing (“Đua xe”).

Nội dung chuyên môn của Đồ án này là nền tảng cho các lĩnh vực:

1. Lập trình AI (TTNT), ML (HM), Deep Learning (HS)
2. Data Science: Lập trình Khoa học dữ liệu, đặc biệt EDA (…..): Phân tích dữ liệu thăm dò
3. CV: Lập trình Thị giác máy tính
4. Games: Lập trình Games
5. Data Visualization = Plot: Lập trình trực quang hóa dữ liệu, dùng biểu đồ & đồ thi
6. GUI (….): Lập trình giao diện đồ họa

## Công cụ và nền tảng kỹ thuật thực hiện Đồ án học phần

+ Ngôn ngữ lập trình **Python**

+ Công cụ lập trình và biên dịch: **PyCharm**

+ Các thư viện chính:

# B1: NẠP THƯ VIỆN

# Speech

import speech\_recognition as sr

from gtts import gTTS

import playsound

# tkinter

import tkinter as tk

from tkinter import messagebox

# thư viện OS (lập thư mục, files)

import os # THƯ VIỆN OS MS. WINDOWS = Lập thư mục & lưu file

# Nạp các thư viện cần thiết EDA

import numpy as np #Numeric Python: Thư viện về Đại số tuyến tính tính

import pandas as pd #Python Analytic on Data System: For data processing(Thư viện xử lý dữ liệu)

from scipy import stats # thư viện cung cấp các công cụ thống kê [statistics]   
 sub-lib của science python [các công cụ khoa học]

from sklearn import preprocessing # Thư viện tiền xử lý DL (XL ngoại lệ: Isolated)

from sklearn.feature\_selection import SelectKBest, chi2   
 # Nạp hàm Thư viện phân tích dữ liệu thăm dò

## Bố cục của báo cáo

Báo cáo gồm những nội dung như sau:

Chương 1: Giới thiệu đồ án của Học phần

Chương 2: Các cơ sở lý thuyết của Học phần phục việc thực hiện đề tài

Chương 3: Phân tích và xác đinh các cơ sở kỹ thuật thực hiện đề tài

Chương 4: Lập Trình Phân Tích Electric Vehicle Population Trong Quý 2 Năm 2023: Electric\_Vehicle\_Population\_Data

Chương 5: Lập Trình Python Nhận Diện Khuôn Mặt Trong Video

Chương 6: Lập trình Game Car Racing

Chương 7: Tổng kết các kết quản đạt được và còn hạn chế của đồ án, đồng thời đề xuất hướng khắc phục hạn chế và phát triển Đồ án.

1. **CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN**

## GIỚI THIỆU VỀ LẬP TRÌNH PYTHON

### Giới thiệu chung

### Cài đặt “bộ thảo chương” (soạn thảo chương trình IDE[[1]](#footnote-1)) và “bộ dịch”

### Giới thiệu các bộ tương tác lập trình

### Cấu hình mô trường tương tác lập trình

## TỔNG QUAN VỀ NGÔN NGỮ LẬP TRÌNH PYTHON

### Danh hiệu (Identifier)

### Về viết mã lệnh lập trình (Code)

### Về chú giải (Comment)

### Các phép toán

### Chuỗi

### Lệnh

### Định nghĩa hàm

### Danh sách

### Bộ dữ liệu: Tuple

### Tập hợp: Set

### Từ điển: Dictionary

### Module

### Package

### Lớp

### Xử lý ngoại lệ

## GIỚI THIỆU CÁC THƯ VIỆN VÀ CÔNG CỤ CĂN BẢN

### Nạp thư viện

### speech\_recognition

### gtts

### playsound

### tkinter

### os

### Pandas

### Tensorflow

### NumPy

### SCIPY

### Theano

### Matplotlib

### scikit-learn = sklearn

### Keras

### NLTK

### PyTorch

### LightGBM

### Eli5

## KỸ THUẬT LẬP TRÌNH ĐỆ QUY: THÁP HÀ NỘI (HANOI TOWER)

## LẬP TRÌNH PYTHON XỬ LÝ GIỌNG NÓI (Trợ lý ảo: Voice Assistant)

### Giới thiệu chủ đề

### Tài liệu và nguồn thực nghiệm

### Giới thiệu Các thư viện Python liên quan

### Giới thiệu bài toán

### Giới thiệu code tham khảo

**[1] Nghe tiếng Việt => Text**

**[2] Text => đọc tiếng Việt**

## LẬP TRÌNH GIAO DIỆN ĐỒ HỌA (GUI)

## LẬP TRÌNH PHÂN TÍCH KHÁM PHÁ (EDA) [Thăm dò]

## LẬP TRÌNH ĐỒ THỊ & BIỂU ĐỒ TRỰC QUANG HÓA DỮ LIỆU (VISUALIZAION: PLOT)

## LẬP TRÌNH THỊ GIÁC MÁY TÍNH (CV)

## LẬP TRÌNH GAMEs CĂN BẢN

1. **PHÂN TÍCH VÀ XÁC ĐINH CÁC CƠ SỞ KỸ THUẬT**

## GIỚI THIỆU CHUNG VỀ CÁC CHỦ ĐỀ CỦA ĐỒ ÁN

Đồ án này gồm 3 chủ đề chính:

+ Lập trình EDA thăm dò Phân Tích Electric Vehicle Population Trong Quý 2 Năm 2023 [Chương 4:]; có trực quang hóa dữ liệu dùng Plot vẽ đồ thị và biểu đồ [3.4].

+ Lập trình CV hỗ trợ nhận diện Nhận Diện Khuôn Mặt Trong Video [Chương 5:]

+ Lập trình Game căn bản: Car Racing [Chương 6:]

Được thực hiện trên hệ thống giao diện GUI [3.3] với cơ chế nhập thông tin vào (input) và xuất kết quả (output) theo hình thức Text và Voice nhờ xử lý “Trợ lý ảo” [3.2].

## LẬP TRÌNH XỬ GIỌNG NÓI (TRỢ LÝ ẢO: VOICE ASSISTANT)

## Giới thiệu chung về chủ đề

### Mô tả yêu cầu của chủ đề

+ Lập trình App xử lý "lời nới" (Speech): Ứng dụng “Trợ lý ảo” (Voice Assistant)

+ Gồm 2 phần (Part) cơ bản sau

1. Nghe tiếng Việt => Text: alo alo

2. Trả lời bằng tiếng Việt : Text => Nói tiếng Việt: alo alo

### Phân tích yều cầu của chủ đề

+ Chủ đề lập trình python này là cơ sở của các ứng dụng về AI, ML, Deep ML.

+ Các hệ thống này thường có 3 phần :

Part1: Speech….: chính là part1 của chủ đề này

Part2: Các kỹ thuật AI

Part2: Voice….: chính là part2 của chủ đề này

## Các nền tảng kỹ thuật

### Nền tảng kỹ thuật cơ bản

### Hệ thống thư việc sử dụng

B1: Cài đặt thư viện

conda install …. Các thư viện sau:

------------------gốc hướng dẫn / dùng python trong visual (pip3)

pip3 install SpeechRecognition

*pip3 install SpeechRecognition pydub # Long audio source*

pip3 install pyaudio

B2 : Nạp thư viện vào App

import speech\_recognition as sr

**import** speech\_recognition **as** sr

**from** gtts **import** gTTS

**import** os

**import** time

**import** playsound

## Thiết kế giao diện

+ Xử lý trong cửa sổ Console Output của Spyder

## Mã lệnh lập trình (python)

**FULL CODE**

# -\*- coding: utf-8 -\*-

"""

Created on Sat Jul 16 19:59:48 2023

@author: VOXUAN

"""

# B1: NẠP THƯ VIỆN

import speech\_recognition as sr

from gtts import gTTS

import playsound

# B2: CHỌN PHƯƠNG ÁN NHẬP ÂM THANH TỪ MICROPHONE

r = sr.Recognizer()

# B3: XỬ LÝ NHẬN DIỆN MIC

#PHẦN 1: IN RA TEXT THEO LỜI NÓI = Nghe tiếng Việt => Text

with sr.Microphone() as Source:

#hiệu chỉnh mic để chuẩn bị nói

print("Hieu chinh nhieu trươc khi noi!")

r.adjust\_for\_ambient\_noise(Source, duration=1)

#nhận lời nói của người dùng từ MIc mặc định lưu dữ liệu âm thanh vào audio\_data

print("Nói tiếng Việt đi, sau 5s sẽ in ra văn bản!")

audio\_data = r.record(Source, duration = 5)

#In ra văn bản text

print("KẾT QUẢ NHẬN DIỆN ..................")

#chuyển lời nói thành văn bản

try:

text = r.recognize\_google(audio\_data,language="vi")

except:

text = "Quý vị nói gì nghe không rõ...!"

#in kết quả ra

print("Quý vị đã nói là : ",format(text))

#PHẦN 2: XUẤT RA LỜI NÓI THEO VĂN BẢN ĐÃ NHẬP = Trả lời bằng tiếng Việt : Text => Nói tiếng Việt

# ĐỌC CHỮ VIỆT => NÓI RA TIẾNG VIỆT

def VXT(t):

vx = gTTS(text=t, lang = 'vi')

xuan = '06VXT.mp3'

vx.save(xuan)

playsound.playsound(xuan)

VXT(text)

#VXT("số thứ tự 06, họ tên: VÕ XUÂN THỂ là giảng viên viên của TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH")

## LẬP TRÌNH GIAO DIỆN ĐỒ HỌA (GUI)

<kết cấu nội dung bài GUI vào đây>

## LẬP TRÌNH TRỰC QUANG HÓA DỮ LIỆU (PLOT)

<kết cấu nội dung bài PLOT vào đây>

1. **LẬP TRÌNH PYTHON PHÂN TÍCH ELECTRIC VEHICLE POPULATION TRONG QUÝ 2 NĂM 2023**

## Giới thiệu chủ đề EDA ELECTRIC VEHICLE POPULATION

Vận dụng kỹ thuật lập trình EDA kết hợp với GUI và Voice Assistant cho phép dùng tập dữ liệu thực nghiệm về dân số sử đụng để dự đoán và tư vấn các nhà quản lý hoạch định kế hoạch bán xe điện trong các năm tới.

Dữ liệu thực nghiệm gồm dân số sử dụng xe diện trong quá khứ của Hoa Kỳ

12DinhDaiHaiDang\_ElectricVehiclePopulationData.csv : dạng file Excel .csv

. Số records: 150482

. Số thuộc tính input: 17 để đoán thuộc tính thứ 17: target variable:

= “Census Tract”[Y|N]

## Phân tích yều cầu của chủ đề

+ Chủ đề lập trình python này là cơ sở của các ứng dụng về DataWarehouse, Bigdata, Data Analysis, Data Sciences, . . .

+ Đây là xử lý nền tảng cho các nội dung chuyên môn (nêu trên) chủ yếu là giai đoạn tiền xử lý dữ liệu đầu vào (Input Preporcessing), như: xử lý NULL, lọc dữ liệu, tích hợp dữ liệu (Integration),…

**Một số trường hợp tiền xử lý dữ liệu đầu vào (input) của các hệ thống EDA**

[1] Xử lý NULL là 150137

## Các nền tảng kỹ thuật liên quan

### Hệ thống thư viện sử dụng

**Bước 1: Nhập các thư viện cần thiết**

import numpy as np

import pandas as pd

from scipy import stats

from sklearn import preprocessing

from sklearn.feature\_selection import SelectKBest, f\_regression

import os

### Tập dữ liệu thực nghiệm

12DinhDaiHaiDang\_ElectricVehiclePopulationData

**Bước 2: Tải tập dữ liệu**

directory\_path = "./"

if os.path.exists(directory\_path):

all\_files = os.listdir(directory\_path)

csv\_files = [f for f in all\_files if f.endswith(".csv")]

if csv\_files:

print("Danh sách tệp CSV có sẵn:")

for i, csv\_file in enumerate(csv\_files):

print(f"{i + 1}. {csv\_file}")

try:

selected\_index = int(input("Nhập số tương ứng với tệp bạn muốn chọn: ")) - 1

if 0 <= selected\_index < len(csv\_files):

selected\_csv\_file = csv\_files[selected\_index]

file\_path = os.path.join(directory\_path, selected\_csv\_file)

print(f"Bạn đã chọn tệp: {selected\_csv\_file}")

print(f"Đường dẫn tệp: {file\_path}")

else:

print("Lựa chọn không hợp lệ. Vui lòng chọn số thứ tự hợp lệ.")

except ValueError:

print("Lựa chọn không hợp lệ. Vui lòng chọn số thứ tự hợp lệ.")

else:

print("Không có tệp CSV nào trong thư mục.")

else:

print(f"Thư mục '{directory\_path}' không tồn tại.")

# Đọc dữ liệu từ tệp CSV đã chọn

df = pd.read\_csv(file\_path)

print('Độ lớn của bảng [frame] dữ liệu thời tiết:', df.shape)

# Hiển thị số lượng dòng từ DataFrame

while True:

try:

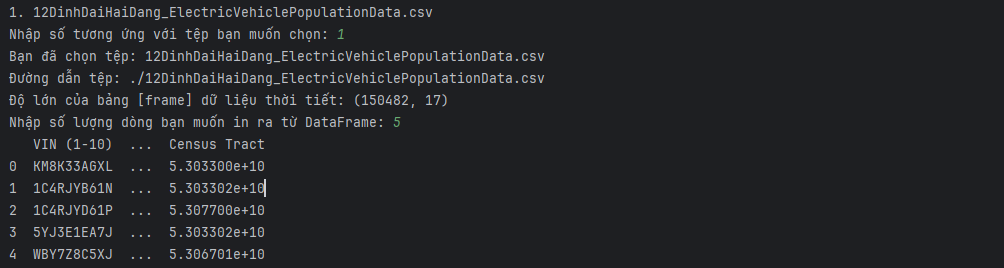
num\_rows\_to\_display = int(input("Nhập số lượng dòng bạn muốn in ra từ DataFrame: "))

print(df.head(num\_rows\_to\_display))

break

except ValueError:

print("Lựa chọn không hợp lệ. Vui lòng nhập một số nguyên.") #Display the shape of the Chạy python G5C12DinhDaiHaiDang \_Topic.py sẽ hiển thị kết quả:



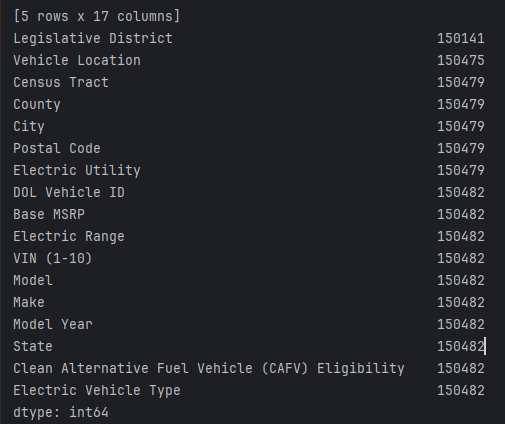
### Mô tả giải thuật

**Bước 3: Xử lý dữ liệu**

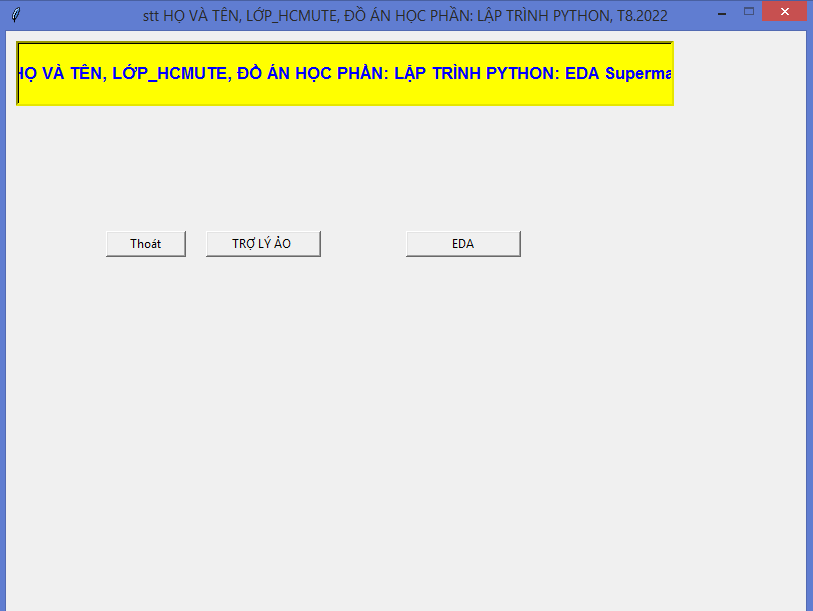
#Xử lý CỘT dữ liệu NULL quá nhiều OR không có giá trị phân tích

print(df.count().sort\_values())

**Kết quả:**



## Thiết kế giao diện



## Mã lệnh lập trình (python)

# -\*- coding: utf-8 -\*-

"""

Created on Sat Jul 16 19:59:48 2023

@author: VOXUAN

"""

# B1: NẠP THƯ VIỆN

# Speech

import speech\_recognition as sr02vtth

from gtts import gTTS

import playsound

# tkinter

import tkinter as tk02vtth

from tkinter import messagebox

# thư viện OS (lập thư mục, files)

import os # THƯ VIỆN OS MS. WINDOWS = Lập thư mục & lưu file

# Nạp các thư viện cần thiết EDA

import numpy as np #Numeric Python: Thư viện về Đại số tuyến tính tính

import pandas as pd #Python Analytic on Data System: For data processing (Thư viện xử lý dữ liệu)

from scipy import stats # thư viện cung cấp các công cụ thống kê [statistics] sub-lib của science python [các công cụ khoa học]

from sklearn import preprocessing # Thư viện tiền xử lý DL (XL ngoại lệ: Isolated)

from sklearn.feature\_selection import SelectKBest, chi2 # Nạp hàm Thư viện phân tích dữ liệu thăm dò

# B2: KHAI BÁO TÊN THƯ MỤC & FILE LƯU CÁC THÔNG TIN BÀI LÀM

vtth02\_FILE = "vtth020.mp3" # lưu tên file Input

vtth02\_DIR = 'vtth02' # Thư mục lưu các file [trên]

os.makedirs(vtth02\_DIR, exist\_ok=True) # TẠO THƯ MỤC LƯU (từ thư viện os - của OS MS. Windows))

"""

CÁC HÀM THỰC HIỆN CÁC CHỨC NĂNG: HỆ THỐNG

"""

def Thoat():

traloi = messagebox.askquestion("Xác nhận","Thiệt thoát không (Y/N)?")

if traloi == "yes": wn.destroy()

"""

CÁC HÀM THỰC HIỆN CÁC CHỨC NĂNG = SPEECH

"""

def Lenh(): # NHẬP ÂM THANH TỪ MICROPHONE

r = sr02vtth.Recognizer()

with sr02vtth.Microphone() as Source:

#hiệu chỉnh mic để chuẩn bị nói

messagebox.showinfo("Nhắc nhở", "Hieu chinh nhieu trươc khi noi!")

r.adjust\_for\_ambient\_noise(Source, duration=1)

#nhận lời nói ra lệnh từ người dùng thông qua MIc [mặc định] lưu dữ liệu âm thanh vào audio\_data

messagebox.showinfo("Cảnh báo", "Bấm OK để bắt đầu Chọn lệnh bằng tiếng Việt, trong 3s" )

audio\_data = r.record(Source, duration = 3)

try:

vlenh = r.recognize\_google(audio\_data,language="vi")

except:

vlenh = "Quý vị nói gì nghe không rõ...!"

# xuất kết quả ra

messagebox.showinfo("Quý vị đã nói là", format(vlenh))

vText = gTTS(text=vlenh, lang = 'vi')

#vFile = '06VXT.mp3'

vText.save(vtth02\_FILE)

playsound.playsound(vtth02\_FILE)

"""

CÁC HÀM THỰC HIỆN CÁC CHỨC NĂNG = EDA

"""

def EDA():

df = pd.read\_csv('./GxttHoTen\_Topic.csv')

# Display the shape of the data set (xem lượng dòng & cột dữ liệu của tập DL gốc)

messagebox.showinfo("Độ lớn của bảng [frame] dữ liệu thời tiết", df.shape)

""""""""""""""""""""""""""""""""""""""""""""

""" CÁC NỘI DUNG SAU CHƯA THỰC HIỆN

""""""""""""""""""""""""""""""""""""""""""

# Bước 3: Xử lý CỘT dữ liệu NULL quá nhiều OR không có giá trị phân tích

# Checking for null values (Kiểm tra giá trị null = đếm số dòng có dữ liệu ứng từng thuộc# tính)

messagebox.showinfo("CÁC CỘT DỬ LIỆU SẮP XẾP THEO THỨ TỰ CÓ DỮ LIỆU TỪ ÍT => NHIỀU",df.count().sort\_values()) #df.count(): đếm số lượng dòng có dữ liệu của df, .sort\_values() sx tăng dân

df = df.drop(columns=['Sunshine','Evaporation','Cloud3pm','Cloud9am','Location','Date','RISK\_MM'],axis=1)

#df = df.drop(columns=['Sunshine','Evaporation','Cloud3pm','Cloud9am','Pressure9am',# 'Pressure3pm','WindDir3pm', 'WindDir9am', 'WindGustDir',# 'WindGustSpeed','Location','Date','RISK\_MM'],axis=1)

messagebox.showinfo("Độ lớn của bảng [frame] dữ liệu SAU KHI XỬ LÝ CỘT NULL", df.shape) # kiểm tra lại số lượng cột & dòng của df sau khi XL NULL cột

# Bước 4: Xử lý DÒNG dữ liệu NULL

# Removing null values (Xóa tất cả các dòng có giá trị null trong tập FRAME dữ liệu.)

df = df.dropna(how='any')

messagebox.showinfo("Độ lớn của bảng [frame] dữ liệu SAU KHI XỬ LÝ DÒNG NULL", df.shape) # kiểm tra lại số lượng cột & dòng của df sau khi XL NULL các dòng DL

# Bước 5: Xử lý loại bỏ các giá trị ngoại lệ (cá biệt): isolated

#kiểm tra tập dữ liệu có bất kỳ ngoại lệ nào không

z = np.abs(stats.zscore(df.\_get\_numeric\_data())) # Dò tìm và lấy các giá trị cá biệt trong tập dữ liệu gốc thông qua điểm z (z\_score)

messagebox.showinfo("MA TRAN Z-SCORE", z) # in ra tập (ma trận) các giá trị z-score từ tập dữ liệu gốc

df= df[(z < 3).all(axis=1)] # kiểm tra và chỉ giữ lại trong df các giá trị số liệu tưng ứng với z-score < 3 # {loại các giá trị >= 3} vì các giá trị z-score >=3 tướng ứng với số liệu quá khác biệt so với các số liệu còn lại (“cá biệt” = “ngoại lệ” = isolated}

messagebox.showinfo("Độ lớn của bảng [frame] dữ liệu SAU KHI XỬ LÝ NGOẠI LỆ", df.shape)# xác định số dòng & cột dữ liệu sau khu xử lý các giá trị cá biệt

# Bước 6: Thay thế các vị trí giá trị 0 và 1 bởi CÓ (Yes) và KHÔNG (No).

#Thay thế yes (CÓ) and no (KO) vào vị trí giá trị 1 (Y) và 0 (N) tương ứng cột|biến RainToday và# RainTomorrow

df['RainToday'].replace({'KHONG': 'No', 'CO': 'Yes'},inplace = True)

df['RainTomorrow'].replace({'KHONG': 'No', 'CO': 'Yes'},inplace = True)

#Bước 7: Chuẩn hóa (Rời rạc hóa) tập dữ liệu Input dùng ..MaxMin

# # CHUẨN HÓA DL

# scaler = preprocessing.MinMaxScaler() #preprocessing là Sub-Library của thư viện sklearn=> hàm .MinMaxScaler() Rời rạc hóa tập dữ liệu Input

# scaler.fit(df)

# df = pd.DataFrame(scaler.transform(df), index=df.index, columns=df.columns) # Rời rạc hóa số liệu theo thang đo scaler

# df.iloc[4:10]

# messagebox.showinfo("Độ lớn của bảng [frame] dữ liệu SAU KHI CHUẨN HÓA DL", df.shape)# xác định số dòng & cột dữ liệu sau khu xử lý các giá trị cá biệt

# # GIAI ĐOẠN 3: PHÂN TÍCH DỮ LIỆU THĂM DÒ : EDA

# # Bước 8: Nạp các thuộc tính quan trọng vào Dataset

# #The important features are put in a data frame

# df = df[['Humidity3pm','Rainfall','RainToday','RainTomorrow']]

# # Bước 9: thực hiện các tính toán trên mô hình phân tích

# #To simplify computations we will use only one feature (Humidity3pm) to build the model

# X = df

# X = df[['Humidity3pm']]

# y = df[['RainTomorrow']]

# X = df.loc[:,df.columns!='RainTomorrow']

# y = df[['RainTomorrow']]

# selector = SelectKBest(chi2, k=3)

# selector.fit(X, y)

# X\_new = selector.transform(X)

# df(['Rainfall', 'Humidity3pm', 'RainToday'], dtype='object')

# messagebox.showinfo(" KÊT QUẢ", X.columns[selector.get\_support(indices=True)])# xác định số dòng & cột dữ liệu sau khu xử lý các giá trị cá biệt

"""

"""

THỦ TỤC CHÍNH = GUI

"""

# B3: LẬP GUI (EX4)

# Tạo một cửa sổ mới

wn = tk02vtth.Tk()

#Thêm tiêu đề cho cửa sổ

wn.title("stt HỌ VÀ TÊN, LỚP\_HCMUTE, ĐỒ ÁN HỌC PHẦN: LẬP TRÌNH PYTHON, T8.2023")

#Đặt kích thước của cửa sổ

wn.geometry('800x600')

#Không cho thay đổi size

wn.resizable(tk02vtth.FALSE, tk02vtth.FALSE)

#Tiêu đề Form = tên đề tài

t = "stt HỌ VÀ TÊN, LỚP\_HCMUTE, ĐỒ ÁN HỌC PHẦN: LẬP TRÌNH PYTHON: EDA Supermarket"

lblDT = tk02vtth.Label(wn, text=t, background = "yellow", fg = "blue", relief = tk02vtth.SUNKEN, font=("Arial Bold", 13), borderwidth = 3, width = 65, height = 3)

lblDT.place(x = 10, y = 10)

# B4: CÁC NÚT LỆNH: EX1 = Speech

#Thoát

btnThoat = tk02vtth.Button(wn, text = "Thoát", width = 10, command = Thoat)

btnThoat.place(x =100, y = 200)# căn cứ vào kích thước form [wn.geometry("800x600")] => canh vị trí Button "thoát"

#Xử lý lời nói = speech = VOICE ASSISTANT

btnNoi = tk02vtth.Button(wn, text = "TRỢ LÝ ẢO ", width = 15, command = Lenh)

btnNoi.place(x = 200, y = 200)

# B5: CÁC NÚT LỆNH: EX3 = EDA = DỰ PHÂN TÍCH DỮ LIỆU KHÁM PHÁ

btnEDA = tk02vtth.Button(wn, text = "EDA", width = 15, command = EDA)

btnEDA.place(x = 400, y = 200)

#Lặp vô tận để hiển thị cửa sổ

wn.mainloop()

1. **LẬP TRÌNH PYTHON NHẬN DIỆN KHUÔN MẶT TRONG VIDEO**

<kết cấu nội dung bài CV vào đây>

## Giới thiệu chủ đề CV …….

## Phân tích chủ đề CV …….

## Các nền tảng kỹ thuật CV

## Các thư viện liên quan

## Giới thiệu Clip …….thực nghiệm

## Mô tả giải thuật CV nhận diện ………. từ Clip……

## Giao diện CV nhận diện ………. từ Clip……

## Mã lệnh lập trình CV nhận diện ………. từ Clip……

1. **LẬP TRÌNH PYTHON CAR RACING**

<kết cấu nội dung bài Game vào đây>

## Giới thiệu chủ đề Game …….

## Phân tích chủ đề Game …….

## Các nền tảng kỹ thuật Lập trình Game

## Các thư viện liên quan

## Giới thiệu Các mẫu media…….thực nghiệm (mẫu xe và đường đua)

## Mô tả giải thuật Game….

## Giao diện Game ….

## Mã lệnh lập trình Game……

1. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN

## Kết luận

## Những kết quả đạt được

ĐỒ ÁN HỌC PHẦN đã thực hiện việc Lập trình Python …………..; bao gồm:

+ Giao diện đồ họa: GUI

+ Xử lý giọng nói trợ lý ảo: Voice Assistant

+ Phân tích khám phá trên tập dữ liệu thực nghiệm: ……………………., có Plot

+ CV ……………

+ Game………..

## Hạn chế

Vì thời gian có hạn nên chưa

+ liên kết voice assistant để ra lệnh phân tích khám phá bằng lời nói.

+ ….

## Hướng phát triển

## Hướng khắc phục các hạn chế

Đồ án sẽ tiếp tục thực hiện

+ liên kết voice assistant để ra lệnh phân tích khám phá bằng lời nói

+ …

## Hướng mở rộng ĐỒ ÁN HỌC PHẦN

Sẽ tiếp tục mở rộng sản phẩm đề tài với các ứng dụng thực tiễn.

# DANH MỤC TÀI LIỆU THAM KHẢO

1. Võ Xuân Thể (2023), Tài liệu học tập Lập trình Python, Trường Đại học Sư phạm kỹ thuật Tp.HCM, Tài liệu lưu hành nội bộ được cung cấp theo từng buổi học
2. Võ Xuân Thể (2023), Bài tập thực hành Lập trình Python, Trường Đại học Sư phạm kỹ thuật Tp.HCM, Tài liệu lưu hành nội bộ được cung cấp theo từng buổi học
3. http://… : nguồn dataFrame được sử dụng trong đề tài này
4. ….

1. IDE (Integrated Development Environment) là môi trường tích hợp dùng để viết code để phát triển ứng dụng [↑](#footnote-ref-1)