**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**



**ĐỒ ÁN**

**HỌC PHẦN MÁY HỌC ỨNG DỤNG**

**Đề tài**

**Xây dựng mô hình máy học để dự đoán số lượng xe đạp được thuê**

**Nhóm sinh viên thực hiện:**

**1. Nguyễn Quốc Việt B2111908**

**2. Trường Huỳnh Tú Như B2111893**

**3. Nguyễn Hồng Tuấn Phát B2111894**

**4. Lê Thị Tiến B2105647**

#### Cần Thơ, 4/2024

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**



**ĐỒ ÁN**

**HỌC PHẦN MÁY HỌC ỨNG DỤNG**

**Đề tài**

**Xây dựng mô hình máy học để dự đoán số lượng xe đạp được thuê**

**Giảng viên hướng dẫn: Nhóm sinh viên thực hiện:**

**1. Nguyễn Q Việt B2111908**

**2. Trương H.T Như B2111893**

**3. Nguyễn H.T Phát B2111894**

**4. Lê Thị Tiến B2105647**

#### Cần Thơ, 4/2024

**NHẬN XÉT CỦA GIẢNG VIÊN**

---------------------------------------------------------------------------------

---------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

-----------------------------------------------------------------------------------------------------------------------------------------------------------------

---------------------------------------------------------------------------------

-----------------------------------------------------------------------------------------------------------------------------------------------------------------

---------------------------------------------------------------------------------

*Cần Thơ, ngày tháng năm*

(Ký và ghi rõ họ tên)

**MỤC LỤC**

**PHÂN CÔNG CÔNG VIỆC**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **STT** | **MSSV** | **Họ tên SV** | **Công việc** | **Thời gian** | **Ký tên nhận việc** | **Ghi chú** |
| 1 | B2111908 | Nguyễn Quốc Việt | Trưởng nhóm, phân tích, xây dựng giải thuật, Làm Slide, Thuyết trình | 31/3 – 10/4 | A close up of a piece of paper  Description automatically generated |  |
| 2 | B2111893 | Trương Huỳnh Tú Như | Thư ký, phân tích, xây dựng giải thuật, Làm Slide, Thuyết trình | 31/3 – 10/4 | A close up of a signature  Description automatically generated |  |
| 3 | B2111893 | Nguyễn Hồng Tuấn Phát | phân tích, xây dựng giải thuật, Làm Slide, Thuyết trình | 31/3 – 10/4 | A piece of paper with blue writing  Description automatically generated |  |
| 4 | B21105647 | Lê Thị Tiến | phân tích, xây dựng giải thuật, Làm Slide, Thuyết trình | 31/3 – 10/4 | A close up of a paper  Description automatically generated |  |

**PHẦN NỘI DUNG**

## Mô tả dữ liệu

Tập dữ liệu Bike Sharing Demand được thu thập từ Kaggle, tập dữ liệu Bike Sharing Demand là một trong những cuộc thi của Kaggle diễn ra vào năm 2015 với câu hỏi được đặt ra là: Dựa vào data của một hãng cung cấp cho thuê xe đạp ở Washington D.C, người tham gia cần phải dự báo được số lượng xe đạp sẽ được thuê.

## 2. Ý nghĩa của dữ liệu

a

## 3. Phân tích dữ liệu và lựa chọn mô hình

### 3.1 Phân tích dữ liệu

#### 3.1.1 Ý nghĩa của các đặc trưng

datetime: cho biết thời gian theo dịnh dạng "%m/%d/%Y %H:%M:%hS

season: cho biết mùa

1: mùa xuân

2: mùa hè

3: mùa thu

4: mùa đông

workingday: cho biết có phải là ngày làm việc không bao gồm cả cuối tuần và ngày lễ

1: ngày làm việc

0: ngày không làm việc

weather: Cho biết thời tiết

1: Trời trong, có mây, có mây rải rác

2: Sương mù + Đầy mây, sương mù + ít mây, sương mù

3: Tuyết nhẹ, mưa nhẹ + sấm sét + mây rải rác, mưa nhẹ + mây rải rác

4: Mưa lớn + mảnh băng + sấm sét + sương mù, tuyết + sương mù

temp: Nhiệt độ thực tế tính theo độ Celsius (độ °C)

atemp: Nhiệt độ mà con người cảm thấy tính theo độ Celsius (độ °C)

humidity: Độ ẩm tương đối

windspeed: Tốc độ gió

casual: Số lượng xe thuê mà không có đăng ký dịch vụ

registered: Số lượng xe thuê mà có đăng ký dịch vụ

count: Tổng số lượng xe được thuê

month: tháng (1-12)

year: năm (2011, 2012)

hour: giờ trong ngày (0-23 giờ)

weekday: ngày trong tuần

0: thứ 2

1: thứ 3

2: thứ 4

3: thứ 5

4: thứ 6

5: thứ 7

6: chủ nhật

#### 3.1.2 Tiền xử lý và phân tích dũ liệu

A screenshot of a computer screen

Description automatically generated

Hình 1: Bản đồ nhiệt của các đặc trưng trước khi xử lý

A screenshot of a computer screen

Description automatically generated

Hình 2: Kiểu dữ liệu của các đặc trưng trước khi xử lý

Do biến datetime có kiểu dữ liệu là object nên sẽ chuyển biến datetime sang các biến year, month, weekday, hour để thực hiện tính toán.

Do đề tài là dự đoán số lượng xe đạp được thuê mà biến casual + registered = count nên sẽ xoá biến casual + registered để xây dựng mô hình.

A screenshot of a computer

Description automatically generated

Hình 3: Bản đồ nhiệt của các đặc trưng sau khi xử lý

A screenshot of a computer

Description automatically generated

Hình 4:Kiểu dữ liệu của các đặc trưng sau khi xử lý

Vậy tập thuộc tính để xây dựng mô hình là: season, holiday, workingday, holiday, weather, temp, atemp, humidity, windspeed, hour, month, weekday, year, count. Trong đó count là thuộc tính mục tiêu cần dự đoán. Sau đây là biểu đồ phân phối của các thuộc tính.

A graph showing the different seasons

Description automatically generated with medium confidence

Hình 5: Biểu đồ phân phối giá trị của thuộc tính season

Dựa vào biểu đồ trên ta thấy dữ liệu season được phân phối đều nhau.

A white rectangular object with black text

Description automatically generated

Hình 6: Biểu đồ phân phối giá trị của thuộc tính holiday

A white rectangular object with black text

Description automatically generated

Hình 7: Biểu đồ phân phối giá trị của thuộc tính workingday

A graph showing the different weather conditions

Description automatically generated with medium confidence

Hình 8: Biểu đồ phân phối giá trị của thuộc tính weather

Biểu đồ phân phối giá trị của thuộc tính season A graph showing the distribution of temperature

Description automatically generated

Hình 9: Biểu đồ phân phối giá trị của thuộc tính temp

A graph of a bar graph

Description automatically generated with medium confidence

Hình 10: Biểu đồ phân phối giá trị của thuộc tính atemp

A graph showing the amount of humidity

Description automatically generated

Hình 11: Biểu đồ phân phối giá trị của thuộc tính humidity

A graph of a distribution of windspeed

Description automatically generated

Hình 12: Biểu đồ phân phối giá trị của thuộc tính windspeed

A diagram of a distribution of hours

Description automatically generated

Hình 13: Biểu đồ phân phối giá trị của thuộc tính hour

A graph of blue bars with white text

Description automatically generated

Hình 14: Biểu đồ phân phối giá trị của thuộc tính month

A graph showing the number of days and months

Description automatically generated with medium confidence

Hình 15: Biểu đồ phân phối giá trị của thuộc tính weekday

A white rectangular object with text

Description automatically generated

Hình 16: Biểu đồ phân phối giá trị của thuộc tính year

A graph of a number of bars

Description automatically generated

Hình 17: Biểu đồ phân phối giá trị của thuộc tính count

### 3.2 Lựa chọn mô hình

## 4. Cấu hình máy tính

## 5. Huấn luyện và Kết quả thực nghiệm

## 6. Đánh giá mô hình

### 6.1 Đánh giá mô hình phân lớp / Regression / Cluster

### 6.2. Nhận xét kết quả thực nghiệm

## PHẦN KẾT LUẬN

### 1. Kết quả đạt được

### 2. Hướng phát triển

# TÀI LIỆU THAM KHẢO

[1].

[2].