

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
ĐẠI HỌC QUỐC GIA TP.HCM



TÊN ĐỒ ÁN

Môn học: Python cho Khoa học Dữ liệu

Giảng viên hướng dẫn: ThS. Hà Minh Tuấn

Nhóm sinh viên thực hiện:

Nguyễn Thị Ngọc Anh	23280037
Trương Thị Quỳnh Giang	23280052
Trần Trung Kiên	23280066

TP.HCM, 2025

# NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

1. Mục tiêu và nội dung của đồ án

2. Kết quả đạt được

3. Ý thức làm việc của sinh viên:

TP. Hồ Chí Minh, ngày ... tháng ... năm 2025

Giảng viên hướng dẫn

ThS. Hà Minh Tuấn

## LỜI CẢM ƠN

# Contents

Bảng ký hiệu và chữ viết tắt	4
Danh sách bảng	5
Danh sách hình	6
Giới thiệu	8
<b>Chương 1: CƠ SỞ LÝ THUYẾT</b>	<b>9</b>
1.1 Các phương pháp vectorize dữ liệu . . . . .	9
1.1.1 One Hot . . . . .	9
1.1.2 TF IDF . . . . .	9
1.1.3 PhoBert . . . . .	9
1.2 Một số phương pháp khác . . . . .	9
1.3 Các mô hình học máy . . . . .	9
1.3.1 Phân tích cảm xúc là gì? . . . . .	9
1.3.2 Phân tích cảm xúc hoạt động thế nào? . . . . .	9
1.4 Phân loại bài toán phân tích cảm xúc . . . . .	9
1.4.1 Logistic Regression . . . . .	9
1.4.2 Support Vector Machine . . . . .	9
1.4.3 Phân tích cảm xúc theo cấp độ . . . . .	9
1.4.4 Phân tích cảm xúc theo khía cạnh . . . . .	9

# Bảng ký hiệu và chữ viết tắt

# Danh sách bảng

# **Danh sách hình**

# Tóm tắt

# **Giới thiệu**

# Chương 1 CƠ SỞ LÝ THUYẾT

## 1.1 Các phương pháp vectorize dữ liệu

1.1.1 One Hot

1.1.2 TF IDF

1.1.3 PhoBert

## 1.2 Một số phương pháp khác

## 1.3 Các mô hình học máy

1.3.1 Phân tích cảm xúc là gì?

1.3.2 Phân tích cảm xúc hoạt động thế nào?

## 1.4 Phân loại bài toán phân tích cảm xúc

1.4.1 Logistic Regression

1.4.2 Support Vector Machine

Support Vector Machine (SVM) là mô hình học máy giám sát, mục tiêu của thuật toán là tìm ra một siêu phẳng (hyperplane) sao cho có thể phân tách tối ưu các điểm dữ liệu thuộc các lớp khác nhau. “Tối ưu” ở đây nghĩa là tìm ra siêu phẳng tạo ra khoảng cách lớn nhất (margin) giữa các lớp dữ liệu. Các điểm dữ liệu có khoảng cách nhỏ nhất đến siêu mặt phẳng (các điểm gần nhất) được gọi là các vector hỗ trợ (support vectors). Margin là độ rộng tối đa của dải không chứa bất kỳ điểm dữ liệu nào và song song với siêu phẳng.

1.4.3 Phân tích cảm xúc theo cấp độ

1.4.4 Phân tích cảm xúc theo khía cạnh

# Bibliography

[1] scikit-learn documentation: <https://scikit-learn.org/>