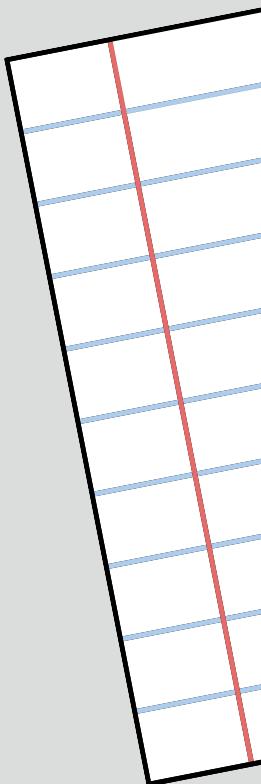


# **TÁC ĐỘNG CỦA CHẠTGPT ĐẾN VIỆC TỰ HỌC CỦA SINH VIÊN ĐẠI HỌC TẠI THÀNH PHỐ HỒ CHÍ MINH**



Lâm Tấn Duy - 2186400009  
Ngô Thị Hồng Uyên - 2186400213  
Nguyễn Văn Đạt - 2186400229

le

# Nội dung

Giới thiệu đề tài

Chuẩn bị dữ liệu

Tiền xử lý dữ liệu

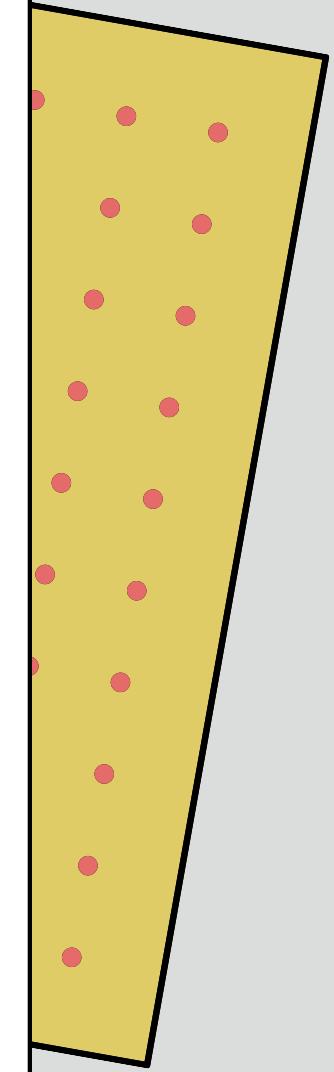
Áp dụng mô hình

Kết luận

# Giới thiệu đề tài



"Learning is a journey, not a destination we reach."



# LÝ DO CHỌN ĐỀ TÀI

**Khám phá các yếu tố ảnh hưởng đến quyết định và tần suất sử dụng ChatGPT:** Tác động của nó đến kết quả học tập và kỳ vọng của sinh viên, phân tích vai trò của nó trong việc tự học.

---

**So sánh các thuật toán phân loại (Classification):** XG-Boost, Random Forest, ANN, SVM, Logistic Regression để trả lời cho 2 câu hỏi:

- Yếu tố nào ảnh hưởng tới quyết định sử dụng ChatGPT của sinh viên?
  - ChatGPT ảnh hưởng đến kết quả học tập của sinh viên như thế nào?
- 

**Áp dụng thuật toán TF-IDF và RAKE:** Trích xuất từ khóa xem sinh viên cảm nghĩ như thế nào về ChatGPT?

# Thu thập dữ liệu



## Cách thức thu thập:

- Lập bảng câu hỏi và thực hiện khảo sát
- Thu thập dữ liệu thực tế từ sinh viên
- Địa điểm thu thập dữ liệu tại các trường đại học tại thành phố Hồ Chí Minh
- Bảng dữ liệu có tổng cộng 35 đặc trưng và 858 mẫu dữ liệu

# Tiền xử lý dữ liệu



"Learning is a journey, not a destination we reach."

# LÀM SẠCH DỮ LIỆU

- Dữ liệu từ các khảo sát trực tuyến (Google Forms) được tải về ở định dạng CSV.
- Loại bỏ những dữ liệu không phù hợp, bị thiếu hoặc trùng lặp.



# XỬ LÝ DỮ LIỆU

- Kiểm tra và xử lý đa cộng tuyến: Dùng hệ số VIF (Variance Inflation Factor) tính toán cho từng biến. VIF vượt quá ngưỡng 5 sẽ bị loại bỏ.
- Phân tích phân phối nhãn và xử lý mất cân bằng dữ liệu: Dùng phương pháp SMOTE tăng cường dữ liệu cho các lớp thiểu số.
- Chuẩn hóa và chia tách dữ liệu: Chia dữ liệu (Train: 70%; Test: 30%), chuẩn hóa bằng MinMaxScaler về khoảng [0, 1].



# Áp dụng mô hình



"Learning is a journey, not a destination we reach."

# MÔ HÌNH

- Nghiên cứu được thực hiện trên 5 mô hình:

SVM

LOGISTIC  
REGRESSION

RANDOM FOREST

XGBOOST

ARTIFICIAL NEURAL  
NETWORK (ANN)

- Dùng phương pháp Gridsearch để tìm kiếm siêu tham số (hyperparameters) cho từng mô hình giảm thiểu nguy cơ overfitting hoặc underfitting.

# KẾT QUẢ

- Đánh giá hiệu quả của các mô hình, các chỉ số như ma trận nhầm lẫn (Confusion Matrix), Precision, Recall và F1 Score.
- Kết quả sau khi áp dụng mô hình để trả lời cho các câu hỏi:
  1. Yếu tố nào ảnh hưởng tới quyết định sử dụng ChatGPT của sinh viên?
  2. ChatGPT ảnh hưởng đến kết quả học tập của sinh viên như thế nào?

# KẾT QUẢ

1. Yếu tố nào ảnh hưởng tới quyết định sử dụng ChatGPT của sinh viên?

Models	Accuracy	Precision (0/1)	Recall (0/1)	F1-score (0/1)
SVM	0.87	0.87 / 0.87	0.86 / 0.88	0.87 / 0.87
Logistic Regression	0.86	0.84 / 0.87	0.87 / 0.85	0.85 / 0.86
Random Forest	0.86	0.85 / 0.87	0.86 / 0.86	0.86 / 0.86
XG-Boost	0.88	0.86 / 0.89	0.89 / 0.86	0.87 / 0.88
ANN	0.87	0.85 / 0.89	0.89 / 0.85	0.87 / 0.87

Từ mô hình cho thấy XG-Boost cho kết quả cao nhất.

# KẾT QUẢ

2. ChatGPT ảnh hưởng đến kết quả học tập của sinh viên như thế nào?

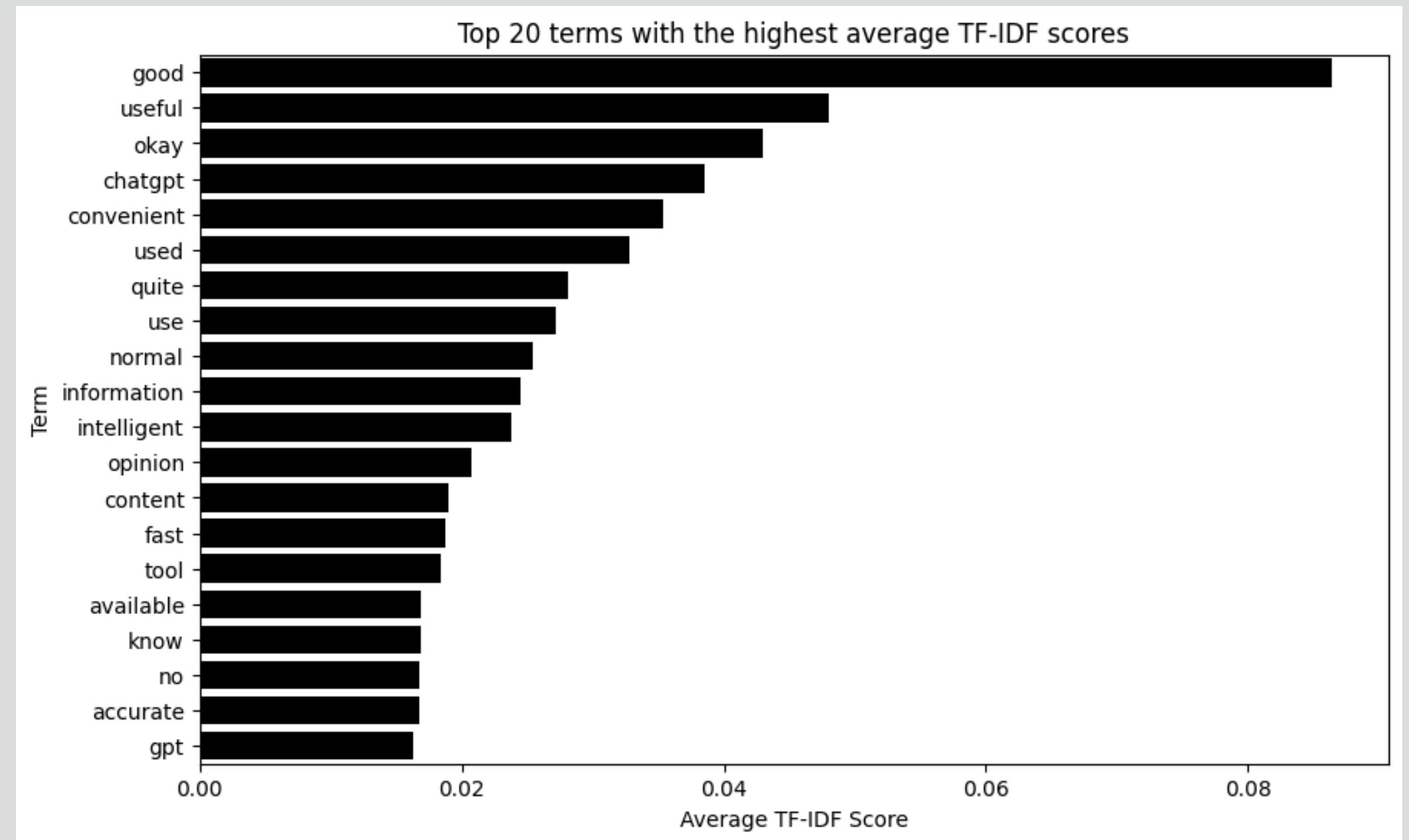
Models	Accuracy	Precision	Recall	F1-score
		(0/1/2)	(0/1/2)	(0/1/2)
SVM	0.85	0.90 / 0.85 / 0.82	0.87 / 0.85 / 0.84	0.88 / 0.85 / 0.83
Logistic Regression	0.61	0.59 / 0.57 / 0.68	0.57 / 0.55 / 0.74	0.58 / 0.56 / 0.74
Random Forest	0.81	0.78 / 0.86 / 0.79	0.86 / 0.77 / 0.79	0.81 / 0.81 / 0.79
XG-Boost	0.88	0.88 / 0.91 / 0.83	0.92 / 0.86 / 0.85	0.90 / 0.88 / 0.84
ANN	0.83	0.89 / 0.84 / 0.78	0.84 / 0.84 / 0.82	0.87 / 0.84 / 0.80

Từ mô hình cho thấy XG-Boost cho kết quả cao nhất.

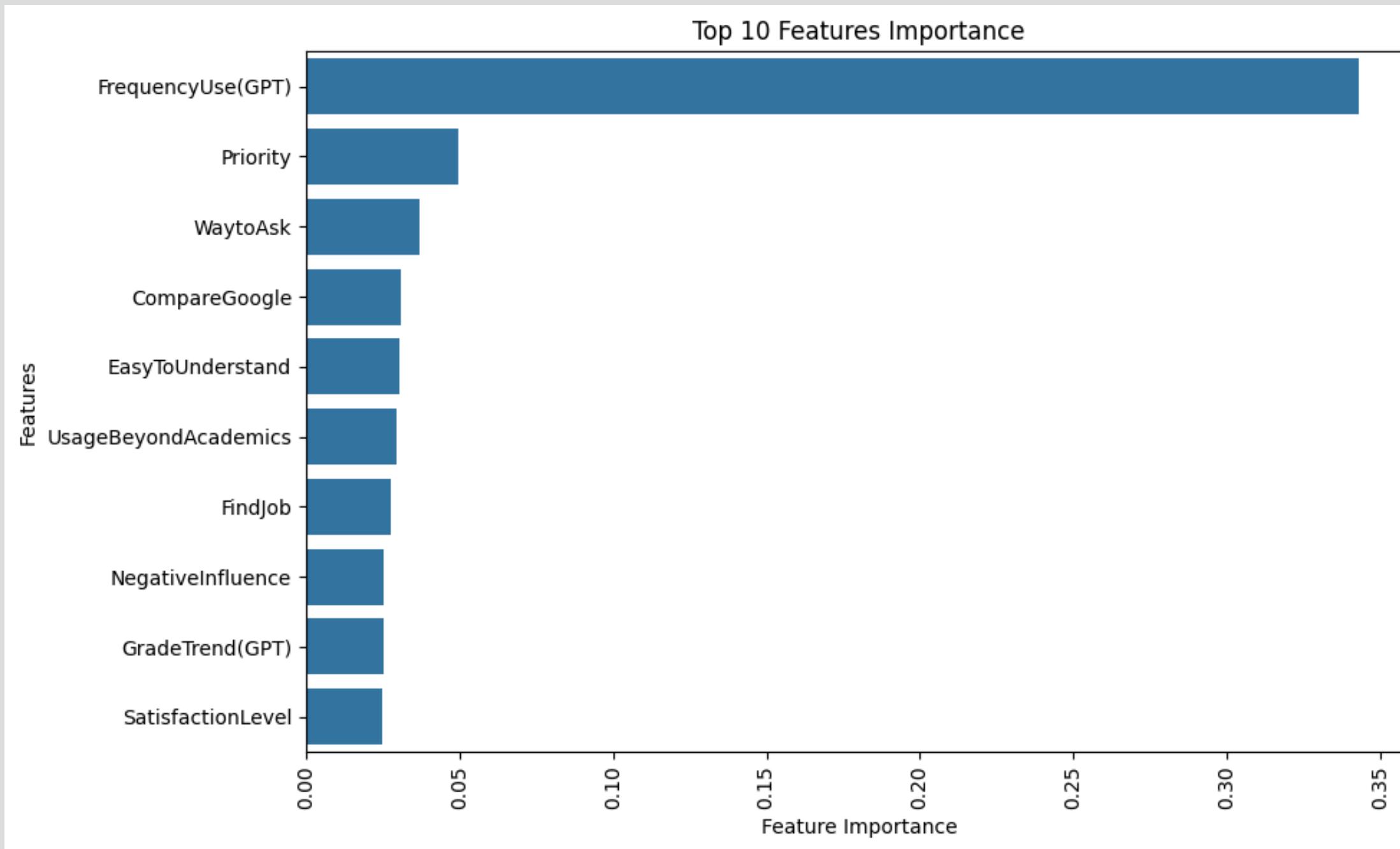
# KẾT QUẢ

3. Cảm nghĩ của các sinh viên về ChatGPT?

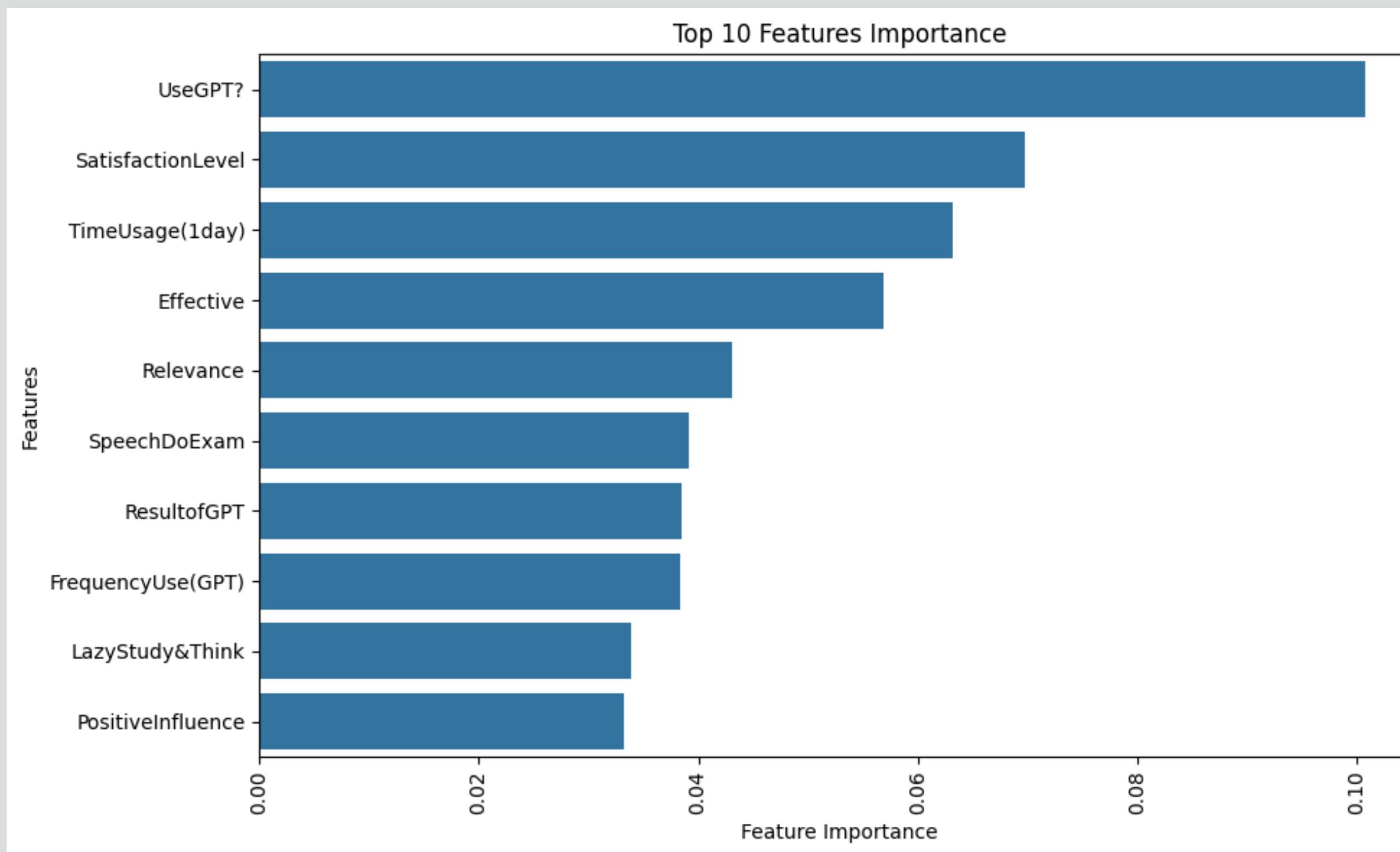
### 3. Cảm nghĩ của các sinh viên về ChatGPT?



# 1. Yếu tố nào ảnh hưởng tới quyết định sử dụng ChatGPT của sinh viên?



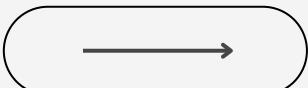
## 2. ChatGPT ảnh hưởng đến kết quả học tập của sinh viên như thế nào?





Presentation are communication tools that can be used as demonstrations, lectures, reports, and more. it is mostly presented before an audience.

# HELLO!



## Welcome to our company

Presentation are communication tools that can be used as demonstrations, lectures, reports, and more. it is mostly presented before an audience.

