

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN CUỐI KÌ
NHẬP MÔN KHOA HỌC DỮ LIỆU

Người thực hiện: **TRẦN VĨNH CƠ – 23640391**

TRẦN NGUYỄN TOÀN PHÁT – 23643121

Lớp : **ĐHKHDL19A - 420301410401**

Khoá : **19**

Người hướng dẫn: **TS BÙI THANH HÙNG**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN CUỐI KÌ
NHẬP MÔN KHOA HỌC DỮ LIỆU

Người thực hiện: **TRẦN VĨNH CƠ - 23640391**

TRẦN NGUYỄN TOÀN PHÁT- 23643121

Lớp : **ĐHKHDL19A - 420301410401**

Khoá : **19**

Người hướng dẫn: **TS. BÙI THANH HÙNG**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025

LỜI CẢM ƠN

Trong quá trình chúng tôi thực hiện đồ án này, chúng tôi đã nhận được nhiều sự hỗ trợ từ các cá nhân, bạn bè, thầy cô. Trước tiên, chúng tôi xin gửi lời cảm ơn chân thành tới những cá nhân đã hỗ trợ, giúp đỡ chúng tôi trong suốt quá trình này. chúng tôi xin gửi lời cảm ơn đặc biệt đến TS. Bùi Thanh Hùng, giảng viên phụ trách hướng dẫn chúng tôi. Nhờ vào sự hướng dẫn của thầy mà đồ án của chúng tôi đã có thêm được nhiều bài học, nhiều kinh nghiệm hơn qua quá trình thực hiện đồ án.

Xin gửi lời cảm ơn đến các bạn sinh viên tham gia khảo sát. Những ý kiến, đánh giá mà các bạn cung cấp là nguồn thông tin quan trọng cho đồ án này của chúng tôi. Rất biết ơn sự hợp tác này của các bạn.

Cuối cùng, chúng tôi xin gửi lời chúc sức khỏe đến TS. Bùi Thanh Hùng, kính chúc thầy luôn mạnh khỏe để truyền tải tri thức, cảm hứng cho nhiều thế hệ sinh viên tiếp theo; để tiếp tục con đường nghiên cứu của mình. Chúc thầy có thêm nhiều thành công hơn trong cuộc sống và trong lĩnh vực mà thầy đang hoạt động.

ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH

Tôi xin cam đoan đây là sản phẩm đồ án của riêng chúng tôi và được sự hướng dẫn của TS. Bùi Thanh Hùng. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình. Trường đại học Công nghiệp TP Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày tháng năm

Tác giả

(ký tên và ghi rõ họ tên)

Trần Vĩnh Cơ

Trần Nguyễn Toàn Phát

PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

Tp. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

TÓM TẮT

Đối với bài 1, nhóm chúng tôi thống nhất chọn chủ đề: “Ảnh hưởng của mạng xã hội đối với đời sống sinh viên”. Nhóm chúng tôi nhận thấy đây là một chủ đề phổ biến do sự phát triển và lan rộng nhanh chóng của các nền tảng mạng xã hội như hiện nay. Để giải quyết bài toán này, nhóm chúng tôi đã tạo ra danh sách gồm hơn 100 câu hỏi/trả lời bằng tiếng Anh có liên quan tới chủ đề mà nhóm đã chọn. Sau khi tạo được bảng câu hỏi/trả lời bằng tiếng Anh, nhóm chúng tôi đã dịch danh sách này để thu được dữ liệu tiếng Việt tương ứng; tiếp theo nhóm dùng ChatGPT để áp dụng 5 luật tương ứng cho mỗi câu hỏi trong danh sách. Kết quả thu được và phát hiện: danh sách hơn 100 câu hỏi/ trả lời bằng tiếng Anh, danh sách hơn 100 câu hỏi/trả lời bằng tiếng Việt; danh sách hơn 500 câu hỏi/trả lời bằng tiếng Việt sau khi dùng ChatGPT; các bảng đánh giá chất lượng các câu dịch, các dữ liệu do ChatGPT sinh thêm.

Đối với bài 2, nhóm thống nhất chọn chủ đề: “Đánh giá mức độ hài lòng của sinh viên đối với chất lượng dịch vụ của thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh”. Đây là một chủ đề khá sát với thực tế do ngày càng có nhiều sinh viên có nhu cầu sử dụng thư viện trường. Để giải quyết được vấn đề trên, chúng tôi đặt ra 2 câu hỏi cho nghiên cứu của mình là: Mức độ hài lòng của sinh viên trường Đại học Công nghiệp Thành phố Hồ Chí Minh (IUH) đối với dịch vụ thư viện hiện tại như thế nào? Những yếu tố nào ảnh hưởng đến sự hài lòng của sinh viên trường IUH đối với thư viện? Từ đó chúng tôi xây dựng 13 câu hỏi cho phiếu khảo sát với 10 câu hỏi dạng lựa chọn phương án từ 1 đến 5, 3 câu hỏi mở với định dạng câu trả lời là text.

Sau khi thiết kế xong phiếu khảo sát, nhóm chúng tôi thực hiện khảo sát 60 bạn sinh viên trường IUH bằng những phiếu được in ra để thu thập dữ liệu phục vụ cho bước phân tích bằng Python. Quá trình thực hiện phân tích dữ liệu bằng Python, chúng tôi sử dụng các thư viện phù hợp như Pandas, Numpy, Matplotlib, Seaborn.. để thực hiện. Mỗi phân tích đều gồm các mô tả về dữ liệu, thống kê cơ bản, mối tương quan giữa câu

hỏi và kết quả, trực quan hóa kết quả phân tích. Chúng tôi thấy rằng mức độ hài lòng của sinh viên đối với dịch vụ thư viện IUH bị ảnh hưởng bởi các yếu tố liên quan đến thiết bị điện tử như máy tính, hệ thống máy lạnh..

Đối với bài tập 3, chúng tôi đã sử dụng thư viện requests và BeautifulSoup để cào dữ liệu từ trang web <https://quotes.toscrape.com/> và lưu kết quả trả về vào file kq.txt. chúng tôi đã trích xuất các đặc trưng của tác giả như: tên, ngày sinh, đường link của tác giả và những câu nói nổi tiếng tương ứng của mỗi tác giả.

Sau đó, ở bước khai phá dữ liệu nhóm chúng tôi sử dụng thư viện Pandas để thực hiện do làm việc với DataFrame giúp cho chúng tôi dễ dàng truy xuất và thao tác, thực hiện các phép toán thống kê. chúng tôi đã thực hiện các phân tích về tác giả; một số trường ngày sinh của vài tác giả còn trống, chúng tôi đã tìm thông tin của tác giả đó để lấy thông tin ngày sinh; phân tích về các câu nói nổi tiếng: câu nói dài nhất, câu nói ngắn nhất, độ dài trung bình, từ xuất hiện nhiều nhất,...; phân tích phong cách của tác giả thông qua câu nói nổi tiếng của người đó; phân tích những tác giả có phong cách tương đồng...

Tóm lại, sau khi thực hiện bài 3, nhóm chúng tôi đã có thêm nhiều kinh nghiệm hơn trong việc cào dữ liệu, xử lý, lưu trữ, phân tích và trực quan hóa dữ liệu; có thêm nhiều kiến thức mới, nhiều kỹ năng mới cho việc lập trình. Hơn nữa, chúng tôi có thêm nhiều hiểu biết hơn nhờ vào việc tiếp xúc với các câu nói của những người nổi tiếng, thành công trong cuộc sống.

MỤC LỤC

LỜI CẢM ƠN	I
PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN	III
TÓM TẮT	IV
MỤC LỤC	1
DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT	3
DANH MỤC CÁC HÌNH VẼ	4
DANH MỤC CÁC BẢNG	6
CHƯƠNG 1	7
1.1. Yêu cầu thu thập dữ liệu	7
1.2. Dịch và đánh giá câu dịch	20
1.3. Sinh thêm dữ liệu	55
1.4. Đánh giá dữ liệu sinh thêm	57
1.5. Lưu bộ dữ liệu	58
CHƯƠNG 2	59
2.1. Câu hỏi đặt ra	59
2.2. Xây dựng câu hỏi khảo sát	61
2.3. Phân tích	65
2.3.1. Kết quả khảo sát	65
2.3.1.2. Đối tượng khảo sát	66
2.3.1.3. Số lượng đối tượng tham gia khảo sát	66
2.3.1.4. Thời gian khảo sát	66
2.3.1.5. Kết quả thu được	67
2.3.2. Phân tích dữ liệu	67
2.4. Kết luận	90
2.4.1. Kết luận và đề xuất giải pháp	90
2.4.2. Hạn chế	92
2.4.3. Hướng phát triển trong tương lai	93

CHƯƠNG 3	94
3.1. Thu thập dữ liệu	94
3.2. Khai phá dữ liệu	103
3.2.1. Xử lý dữ liệu - Data Imputation	103
3.2.2. Khám phá dữ liệu - Data Exploration	109
3.2.3. Trích xuất đặc trưng- Feature Extraction	128
3.2.4. Suy luận	129
LÀM VIỆC NHÓM	134
TÀI LIỆU THAM KHẢO	136
PHỤ LỤC	138
TỰ ĐÁNH GIÁ	141

DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT

CÁC CHỮ VIẾT TẮT

IUH : Trường Đại học Công nghiệp Thành phố Hồ Chí Minh.

NLTK: Natural Language Toolkit.

TF-IDF: Term frequency–inverse document frequency.

DANH MỤC CÁC HÌNH VẼ

Hình 2.3.2.1. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 1.	69
Hình 2.3.2.2. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 2.	71
Hình 2.3.2.3. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 3.	73
Hình 2.3.2.4. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 4.	75
Hình 2.3.2.5. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 5.	77
Hình 2.3.2.6. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 6.	79
Hình 2.3.2.7. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 7.	81
Hình 2.3.2.8. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 8.	83
Hình 2.3.2.9. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 9.	85
Hình 2.3.2.10. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 10.	87
Hình 2.3.2.11. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 11.	88
Hình 2.3.2.12. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 12.	89
Hình 2.3.2.13. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 13.	90
Hình 3.1.1. Kết quả của việc lưu file thành công.	95
Hình 3.1.2. Dữ liệu trong file kq.txt.	95
Hình 3.1.3. Sơ đồ mô tả về Cấu trúc trang Web: http://quotes.toscrape.com/	96
Hình 3.1.4. Kết quả đọc tất cả các thẻ html(div) với lớp là 'quote'.	97
Hình 3.1.5. Kết quả danh sách các tác giả có chứa nhãn small.	98
Hình 3.1.6. Kết quả danh sách thông tin các tác giả.	103
Hình 3.2.1.1. Kết quả sau khi thêm cột STT và lưu vào file mới.	104
Hình 3.2.1.2. Kết quả bộ dữ liệu sau khi thêm cột Năm mất.	107
Hình 3.2.1.3. Kết quả bộ dữ liệu sau khi thêm Tuổi.	109
Hình 3.2.2.1. Kết quả thống kê về tác giả và câu nói nổi tiếng có trong bộ dữ liệu.	110
Hình 3.2.2.2. Biểu đồ thể hiện thống kê các tác giả và câu nói nổi tiếng theo phân trăm.	111

Hình 3.2.2.3. Kết quả Thống kê về năm sinh và độ tuổi của các tác giả.	114
Hình 3.2.2.4. Biểu đồ thống kê năm sinh của các tác giả.	115
Hình 3.2.2.5. Biểu đồ thống kê độ tuổi của các tác giả.	115
Hình 3.2.2.6. Kết quả thống kê về các câu nói nổi tiếng: câu dài nhất, ngắn nhất, số từ.	117
Hình 3.2.2.7. Kết quả thống kê tổng quan về số từ trong các câu nói.	117
Hình 3.2.2.8. Kết quả thống kê về các từ và số lần trong các câu nói.	119
Hình 3.2.2.9. Biểu đồ thể hiện 10 từ xuất hiện phổ biến nhất trong các câu nói nổi tiếng.	121
Hình 3.2.2.10. Hình ảnh thể hiện những từ được xuất hiện nhiều nhất trong các câu nói nổi tiếng theo kích thước của từ.	122
Hình 3.2.2.11. Kết quả số lượng câu nói theo từng tác giả.	123
Hình 3.2.2.12. Kết quả thống kê tổng quan về số lượng câu nói của từng tác giả.	124
Hình 3.2.2.13. Biểu đồ phân phối của các câu nói nổi tiếng theo tác giả.	125
Hình 3.2.2.14. Biểu đồ mối quan hệ giữa tác giả dựa trên số lần trùng câu nói nổi tiếng.	127
Hình 3.2.4.1. Kết quả báo cáo đánh giá, dự đoán sau khi huấn luyện mô hình và kết quả của độ tương quan cao nhất.	133

DANH MỤC CÁC BẢNG

Bảng 1.1. Các câu hỏi/trả lời tiếng Anh.	7
Bảng 1.2. Đánh giá các câu dịch bộ dữ liệu câu hỏi/trả lời tiếng Anh.	21
Bảng 1.3. Ví dụ về áp dụng 5 luật để sinh thêm câu hỏi.	56
Bảng 1.4. Bảng đánh giá bộ dữ liệu sinh thêm.	57
Bảng 2.2.1. Câu hỏi thuộc khía cạnh Chất lượng các thiết bị điện tử và kết nối và lý giải kèm theo.	63
Bảng 2.2.2. Câu hỏi thuộc khía cạnh Không gian thu viện và lý giải kèm theo. ..	64
Bảng 2.2.3. Câu hỏi thuộc khía cạnh Chất lượng phục vụ của cán bộ, nhân viên và lý giải kèm theo.	65
Bảng 2.2.4. Câu hỏi thuộc khía cạnh Mong muốn của sinh viên và lý giải kèm theo.	65

CHƯƠNG 1

THU THẬP DỮ LIỆU

1.1. Yêu cầu thu thập dữ liệu:

Hiện nay, với sự phát triển mạnh mẽ của công nghệ và truyền thông các bạn trẻ, đặc biệt là sinh viên đã tham gia sử dụng mạng xã hội để phục vụ cho nhiều nhu cầu cá nhân của mình. Ngày càng có nhiều hoạt động trong đời sống của sinh viên có sự hiện diện của mạng xã hội, điều đó tạo ra những mặt ảnh hưởng nhất định đến đời sống sinh viên. Nhận thấy điều đó, nhóm quyết định chọn đề tài: “Ảnh hưởng của mạng xã hội đến đời sống sinh viên”.

Từ tên đề tài đã lựa chọn, các thành viên đã dùng công cụ Google Scholar để tìm kiếm các bài nghiên cứu khoa học, bài báo khoa học, sách....có liên quan đến đề tài của nhóm để tham khảo nội dung, dựa vào nội dung tham khảo để chọn ra các câu hỏi/trả lời tiếng Anh phù hợp. Sau quá trình làm việc, kết quả thu được là 120 câu hỏi/trả lời như trong Bảng 1.1 bên dưới.

Bảng 1.1. Các câu hỏi/trả lời tiếng Anh.

Câu hỏi	Câu trả lời
How does Facebook positively affect your study timings in routine life?	It helps me stay updated with study schedules shared in groups.
How do groups and pages related to studies created by scholars on Facebook help students in their studies?	They provide useful resources and expert advice.
Does Facebook have a positive impact on students' academic performance?	Yes, if used to join academic groups and follow educational pages.
How does logging into Facebook help	I use it to discuss assignments and

you focus on work and stay in contact with classmates?	plan group projects.
How does using Twitter in your daily life affect your academic performance?	It helps me follow academic accounts and learn from experts.
How does sharing study material on Twitter help students who are weak in their studies?	It gives them access to simplified content and discussions.
Why do you prefer to express your ideas and feelings on Twitter?	It's quick and I can connect with people who share my interests.
Why is your membership on the Twitter website important to you?	I follow educational accounts and stay informed.
How are social networking sites affecting the way students speak or write?	They influence informal language and grammar use.
How can social networking sites be an effective tool for E-learning for students?	They allow resource sharing and online discussions.
Do you think social networking sites change students' behavior in an effective manner?	Yes, they improve collaboration and awareness.
How are social networking sites more effective in communicating with friends?	They offer instant messaging and group chats.

Do social networking sites differentiate students based on their demography?	Sometimes, based on content preference and language.
How has using materials from SMS like Wikipedia for research helped improve your grades?	It provides quick background knowledge and references.
How do materials obtained from social networking sites complement what you have been taught in class?	They offer real-life examples and updates.
How does engaging in academic forums on social media increase your understanding of topics discussed in class?	I get different perspectives and detailed explanations.
How do group discussions on social media contribute to good academic results?	They clarify doubts and promote peer learning.
How does your addiction to social networks affect your academic life?	It reduces my focus and wastes study time.
In what ways do online social networks distract you from your studies?	Notifications and scrolling reduce my concentration.
Why can time spent on social media never be compared to time spent on studies?	Studying gives long-term value, unlike social media.
Why has there been no improvement in	I often prioritize entertainment over

your grades since you became engaged in social networking sites?	studying.
How frequently do you check social media in the 15-minute window before bedtime?	Almost every night.
How often do you browse social media while having breakfast in the morning?	Most mornings.
How many times do you engage with social media during your evening meal?	Usually once or twice.
How often do you access social media within the first 15 minutes after waking up?	Nearly every day.
How frequently do you use social media while having lunch?	Often, especially when eating alone.
Do you compare yourself to others on social media? How does it make you feel?	Sometimes I do, and it makes me feel insecure about my own achievements.
How do you balance social media usage and academic responsibilities?	I try to set limits, like turning off notifications during study time.
Has social media helped you in building professional networks?	Not much, but I follow some professionals in my field for inspiration.
Do you think social media addiction is	Yes, many of my friends struggle

a problem among students?	with spending too much time on it.
How do you feel when you take a break from social media?	I feel more productive and less stressed, but I also miss updates from friends.
What measures do you think could reduce the negative impact of social media on students?	Setting time limits and using apps that track usage could help.
Do you think social media affects your mental health? If yes, how?	Yes, it sometimes makes me feel anxious or lonely when I see others' highlight reels.
How do you think teachers or institutions could use social media to support learning?	They could create interactive groups or post useful resources for students.
Do you believe social media can enhance collaborative learning among students?	Yes, group chats and shared documents make teamwork easier.
What advice would you give to other students about using social media responsibly?	Prioritize academic goals and limit social media use during study hours.
Do you think social media usage affects your self-esteem? If yes, how?	Yes, sometimes I feel insecure when comparing myself to others on social media, which lowers my self-esteem.
Have you ever created a fake account	No, I prefer to use my real identity

on social media? If yes, why?	online.
How do you feel when you receive fewer likes or comments than expected on your posts?	I feel disappointed and sometimes question my self-worth.
Do you secretly check someone else's social media profile (e.g., crush, ex-partner)? How often?	Yes, I sometimes check my ex-partner's profile, about once a week.
Do you believe social media addiction is a real issue among students? Why or why not?	Yes, because many students spend excessive time online, neglecting studies and real-life interactions.
How does social media influence your mood on a daily basis?	It boosts my mood when I see positive content but makes me anxious if I encounter negative posts.
Have you ever tried to reduce your social media usage? Were you successful?	Yes, but I relapsed after a few days due to fear of missing out (FOMO).
Do you think the number of followers/friends you have reflects your social status?	Partially, but I know it's not a true measure of real-life relationships.
How do you react when you see others posting about their achievements or happy moments?	I feel happy for them but sometimes envious if I'm not doing as well.
Do you use social media more when	Yes, it's a way to distract myself and

you feel lonely or stressed?	feel connected to others.
How does social media usage affect your attendance in physical classes or lectures?	Sometimes I skip lectures thinking I can catch up later through shared notes online, but it's not as effective.
Do you think social media helps you stay updated with the latest medical advancements?	Yes, I follow medical pages and influencers who post about breakthroughs and conferences.
What percentage of your social media time is spent on academic vs. non-academic activities?	About 40% is academic (e.g., research, discussions), and 60% is for entertainment or socializing.
Have you ever felt distracted from studying due to social media notifications?	Frequently. I often pause studying to check messages or updates, which breaks my focus.
Do you use social media to connect with medical professionals or mentors?	Occasionally, I join LinkedIn groups or Twitter threads where professionals share insights.
How has social media influenced your preparation for exams?	It's helpful for quick doubt-solving but sometimes leads to procrastination.
Do you think social media platforms should be formally integrated into medical education?	Yes, but with guidelines to ensure professionalism and avoid misinformation.
Have you ever experienced anxiety or	Yes, seeing others post about their

stress due to social media comparisons with peers?	study progress makes me feel behind sometimes.
Do you participate in online medical quizzes or challenges on social media?	Yes, they're fun and help reinforce knowledge, but I don't rely on them solely.
How do you verify the accuracy of medical information shared on social media?	I cross-check with textbooks or trusted sites like PubMed before accepting any claims.
Has social media usage affected your sleep schedule during exam periods?	Yes, late-night scrolling reduces my sleep, impacting next-day productivity.
Have you ever abstained from social media to improve academic performance? What was the outcome?	Yes, during midterms. My grades improved due to better focus and time management.
How do you verify the credibility of academic information shared on social media?	I cross-check with textbooks or reputable academic sites like Google Scholar.
What social media features distract you the most from studying?	Endless scrolling on Instagram Reels and Facebook notifications disrupt my concentration.
How much time do you spend each day on social media on your mobile phone?	About 2-3 hours, mainly for surfing news and communication.
Do you feel that using social media	Yes, sometimes it reduces study

affects your grades?	time, resulting in lower scores.
Do you often check social networks during class?	Occasionally, about 1-2 times per session.
Does using social media distract you from studying?	Yes, social media can be distracting when you need to concentrate.
Do you use social media to support your studies?	Yes, join study groups and find resources on social media.
Do you feel anxious when you can't access social media?	Sometimes mild anxiety, especially when waiting for important messages.
Do you compare yourself to others on social media?	Yes, sometimes compare, feel inferior when seeing others successful.
Does using social media make you feel lonely?	Sometimes, when I see my friends seem happier on social media.
Have you ever experienced bullying on social media?	No, never experienced cyberbullying.
Do you feel pressured when using social media?	Yes, pressure from comparison or needing to maintain an image.
Do you plan your social media usage time?	Tried to plan, but not followed through.
Do you feel out of control of your time	Yes, often surf social media longer

when using social networks?	than intended.
Does using social media affect your study time?	Yes, reduce study time due to too much social media surfing.
Do you often procrastinate on work due to social media?	Yes, social media often makes me procrastinate on my homework.
Do you use tools to limit your time on social media?	Tried apps like Forest, but haven't stuck with it.
How do you use social media to keep in touch with friends and family?	Use text messages and group chats for regular communication.
Do you feel social media helps you connect with others?	Yes, it makes it easier to stay in touch with friends and family.
Have you ever had a conflict with someone on social media?	Rare, just a few small arguments on the post.
Do you feel like your relationship is being affected by social media?	Sometimes, due to lack of face-to-face interaction with friends.
Do you use social media to participate in groups or communities?	Yes, join study and interest groups on social media.
Do you feel like you are addicted to social media?	Feels dependent at times, especially when surfing mindlessly.
Have you ever tried to cut down on your social media usage?	Yes, tried time limit but hard to maintain.
Do you feel frustrated when you can't	Sometimes annoying, but can get

access social networks?	used to.
Do you check social media regularly even when there are no new notifications?	Yes, check regularly about 5-6 times/day.
Do you feel like social media use affects your sleep?	Yes, surfing social media before bed makes it harder to fall asleep.
Do you use social media to search for career opportunities?	Yes, use LinkedIn to find jobs and make professional connections.
Do you follow experts in your field on social media?	Yes, follow to learn knowledge and experience.
Do you feel social media helps you develop your professional skills?	Yes, through posts and courses shared on social media.
Do you share your academic or professional achievements on social media?	Occasionally, mainly share milestones.
Have you received positive feedback from employers via social media?	No, but received an interview invitation via LinkedIn.
Are you taking steps to protect your privacy on social media?	Yes, keep your account private and limit sharing of sensitive information.
Do you share sensitive personal information on social media?	No, I avoid sharing sensitive information.

Have you ever had a privacy issue on social media?	No, no serious problems yet.
Do you feel safe using social media?	Relatively safe, but still be careful with what you share.
Do you know how to set privacy on social media platforms?	Yes, I know how to set basic privacy.
Do you feel social media makes you more aware of social issues?	Yes, social media provides information about social issues.
Do you participate in social or charity campaigns via social media?	Occasionally, mostly share or donate small amounts.
Do you feel social media promotes your community involvement?	Yes, it makes it easier for me to participate in community events.
Do you share your personal views on social issues on social media?	Rare, only share when necessary.
Do you feel social media influences your political views?	A little bit, due to exposure to different opinions.
Do you use social media to discuss homework with friends?	Yes, regular chat group exchange.
Do you feel social media helps you understand lessons better?	Sometimes, through documents and group discussions on social networks.
Are you often distracted by social	Yes, social media easily distracts me

media while studying?	when studying.
Do you use social media to search for study information?	Yes, find documentation and video tutorials on social media.
Do you feel that using social media affects your academic performance?	Yes, it reduces performance due to time loss.
Do you feel that using social media affects your sleep?	Yes, surfing social media late at night reduces sleep quality.
Do you feel tired after using social networks for a long time?	Yes, feel tired and lose energy.
Do you feel eye pain or headache after using social media?	Yes, especially when using the phone for too long.
Do you take time to rest after using social media?	Occasionally, but often forget to rest.
Do you do exercises to minimize the negative effects of social media?	Rarely, but sometimes light exercise is used to relieve stress.
What do you mainly use social media for?	Communicate, entertain, and find educational information.
Do you use social media for entertainment?	Yes, watch funny videos and posts.
Do you use social media to stay updated with news?	Yes, follow the news through social networking sites.
Do you use social media to connect	Yes, mainly to keep in touch with

with friends?	friends.
Do you use social media for studying?	Yes, join study groups and find materials.
How do you rate the overall impact of social media on your life?	Both benefits (connection, learning) and disadvantages (distraction, stress).
Do you feel that social media brings more benefits than harm?	It depends on the time, but the effects are often more prominent if left unchecked.
Do you have any suggestions to help students use social media more effectively?	Set time limits, prioritize studying, and use social media for positive purposes.
Have you ever received learning support from communities/groups on social networks?	Yes, get materials and advice from the study group.
Do you feel guilty about the time you spend on social media?	Yes, especially when procrastinating on school work.

1.2. Dịch và đánh giá câu dịch:

Sau khi có được bộ câu hỏi/trả lời tiếng Anh, nhóm đã nhập dữ liệu vào công cụ Google Sheets và dùng tiện ích Translate my sheet để thực hiện dịch các câu hỏi/trả lời từ tiếng Anh sang tiếng Việt. Việc dịch các câu hỏi/trả lời này phải thực hiện 2 lần vì tiện ích Translate my sheet phiên bản miễn phí giới hạn chỉ dịch 100 cells với mỗi lần dịch, như vậy ở lần đầu nhóm thực hiện dịch cho 50 câu hỏi/trả lời đầu tiên, lần thứ hai thực hiện dịch cho 10 câu hỏi/trả lời còn

lại. Kết quả sau khi dịch được nhóm thực hiện đánh giá như trong Bảng 1.2 bên dưới.

Bảng 1.2. Đánh giá các câu dịch bộ dữ liệu câu hỏi/trả lời tiếng Anh.

Câu hỏi tiếng Anh	Câu hỏi tiếng Việt	Đánh giá	Câu trả lời tiếng Anh	Câu trả lời tiếng Việt	Đánh giá
How does Facebook positively affect your study timings in routine life?	It helps me stay updated with study schedules shared in groups.	Good	Facebook ảnh hưởng tích cực đến thời gian học tập trong cuộc sống thường ngày của bạn như thế nào?	Nó giúp tôi cập nhật lịch học được chia sẻ trong nhóm.	Tốt
How do groups and pages related to studies created by scholars on Facebook help students in their studies?	They provide useful resources and expert advice.	Good	Các nhóm và trang liên quan đến nghiên cứu do học giả tạo ra trên Facebook giúp ích gì cho sinh viên trong quá trình học tập?	Họ cung cấp các nguồn tài nguyên hữu ích và lời khuyên của chuyên gia.	Tốt
Does Facebook have a	Yes, if used to join academic groups and	Good	Facebook có tác động tích cực đến kết quả học	Có, nếu dùng để tham gia	Tốt

positive impact on students' academic performance?	follow educational pages.		tập của học sinh không?	các nhóm học thuật và theo dõi các trang giáo dục.	
How does logging into Facebook help you focus on work and stay in contact with classmates?	I use it to discuss assignments and plan group projects.	Good	Việc đăng nhập vào Facebook giúp bạn tập trung vào công việc và giữ liên lạc với bạn học như thế nào?	Tôi sử dụng nó để thảo luận bài tập và lập kế hoạch dự án nhóm.	Tốt
How does using Twitter in your daily life affect your academic performance?	It helps me follow academic accounts and learn from experts.	Good	Việc sử dụng Twitter trong cuộc sống hàng ngày ảnh hưởng thế nào đến kết quả học tập của bạn?	Nó giúp tôi theo dõi các tài khoản học thuật và học hỏi từ các chuyên gia.	Tốt
How does sharing study material on Twitter help students who are weak in	It gives them access to simplified content and discussions.	Good	Việc chia sẻ tài liệu học tập trên Twitter có giúp ích gì cho những học sinh học yếu?	Nó cho phép họ tiếp cận nội dung và thảo luận được đơn	Tốt

their studies?				giản hóa.	
Why do you prefer to express your ideas and feelings on Twitter?	It's quick and I can connect with people who share my interests.	Good	Tại sao bạn thích thể hiện ý tưởng và cảm xúc của mình trên Twitter?	Thật nhanh chóng và tôi có thể kết nối với những người có cùng sở thích với tôi.	Tốt
Why is your membership on the Twitter website important to you?	I follow educational accounts and stay informed.	Good	Tại sao tư cách thành viên trên trang web Twitter lại quan trọng với bạn?	Tôi theo dõi các tài khoản giáo dục và cập nhật thông tin.	Tốt
How are social networking sites affecting the way students speak or write?	They influence informal language and grammar use.	Good	Các trang mạng xã hội ảnh hưởng thế nào đến cách nói hoặc viết của học sinh?	Chúng ảnh hưởng đến cách sử dụng ngôn ngữ không chính thức và ngữ pháp.	Tốt
How can	They allow	Good	Làm thế nào	Chúng cho	Tốt

social networking sites be an effective tool for E-learning for students?	resource sharing and online discussions.		các trang mạng xã hội có thể trở thành công cụ học tập trực tuyến hiệu quả cho sinh viên?	phép chia sẻ tài nguyên và thảo luận trực tuyến.	
Do you think social networking sites change students' behavior in an effective manner?	Yes, they improve collaboration and awareness.	Good	Bạn có nghĩ rằng các trang mạng xã hội có thể thay đổi hành vi của học sinh một cách hiệu quả không?	Có, chúng cải thiện khả năng cộng tác và nhận thức.	Tốt
How are social networking sites more effective in communicating with friends?	They offer instant messaging and group chats.	Good	Làm thế nào các trang mạng xã hội có thể hiệu quả hơn trong việc giao tiếp với bạn bè?	Họ cung cấp dịch vụ nhắn tin tức thời và trò chuyện nhóm.	Tốt
Do social networking sites differentiate	Sometimes, based on content preference	Good	Các trang mạng xã hội có phân biệt học sinh dựa trên đặc	Đôi khi, dựa trên sở thích về nội dung và	Tốt

students based on their demography?	and language.		điểm nhân khẩu học không?	ngôn ngữ.	
How has using materials from SMS like Wikipedia for research helped improve your grades?	It provides quick background knowledge and references.	Good	Việc sử dụng tài liệu từ SMS như Wikipedia để nghiên cứu có giúp cải thiện điểm số của bạn như thế nào?	Cung cấp kiến thức cơ bản và tài liệu tham khảo nhanh chóng.	Tốt
How do materials obtained from social networking sites complement what you have been taught in class?	They offer real-life examples and updates.	Good	Những tài liệu thu thập được từ các trang mạng xã hội bổ sung cho những gì bạn đã được học trên lớp như thế nào?	Họ cung cấp những ví dụ thực tế và thông tin cập nhật.	Tốt
How does engaging in	I get different perspectives	Good	Việc tham gia các diễn đàn	Tôi nhận được nhiều	Tốt

academic forums on social media increase your understanding of topics discussed in class?	and detailed explanations.		học thuật trên mạng xã hội giúp bạn hiểu rõ hơn về các chủ đề được thảo luận trên lớp như thế nào?	góc nhìn khác nhau và lời giải thích chi tiết.	
How do group discussions on social media contribute to good academic results?	They clarify doubts and promote peer learning.	Good	Thảo luận nhóm trên mạng xã hội đóng góp như thế nào vào kết quả học tập tốt?	Họ giải đáp những thắc mắc và thúc đẩy việc học tập lẫn nhau.	Tốt
How does your addiction to social networks affect your academic life?	It reduces my focus and wastes study time.	Good	Việc nghiện mạng xã hội ảnh hưởng đến cuộc sống học tập của bạn như thế nào?	Nó làm giảm sự tập trung của tôi và lãng phí thời gian học.	Tốt
In what ways	Notifications	Good	Mạng xã hội	Thông báo	Tốt

do online social networks distract you from your studies?	and scrolling reduce my concentration.		trực tuyến làm bạn mất tập trung vào việc học như thế nào?	và việc cuộn trang làm giảm khả năng tập trung của tôi.	
Why can time spent on social media never be compared to time spent on studies?	Studying gives long-term value, unlike social media.	Good	Tại sao thời gian dành cho mạng xã hội không bao giờ có thể so sánh với thời gian dành cho việc học?	Học tập mang lại giá trị lâu dài, không giống như mạng xã hội.	Tốt
Why has there been no improvement in your grades since you became engaged in social networking sites?	I often prioritize entertainment over studying.	Good	Tại sao điểm số của bạn không được cải thiện kể từ khi bạn tham gia vào các trang mạng xã hội?	Tôi thường ưu tiên giải trí hơn học tập.	Tốt
How frequently do	Almost every night.	Good	Bạn kiểm tra mạng xã hội	Hầu như đêm nào	Tốt

you check social media in the 15-minute window before bedtime?			bao nhiêu lần trong khoảng thời gian 15 phút trước khi đi ngủ?	cũng vậy.	
How often do you browse social media while having breakfast in the morning?	Most mornings.	Good	Bạn thường xuyên lướt mạng xã hội trong khi ăn sáng vào buổi sáng như thế nào?	Hầu hết các buổi sáng.	Tốt
How many times do you engage with social media during your evening meal?	Usually once or twice.	Good	Bạn sử dụng mạng xã hội bao nhiêu lần trong bữa tối?	Thông thường là một hoặc hai lần.	Tốt
How often do you access social media within the first 15	Nearly every day.	Good	Bạn thường xuyên truy cập mạng xã hội trong vòng 15 phút đầu tiên	Gần như mỗi ngày.	Tốt

minutes after waking up?			sau khi thức dậy như thế nào?		
How frequently do you use social media while having lunch?	Often, especially when eating alone.	Good	Bạn sử dụng mạng xã hội thường xuyên như thế nào trong khi ăn trưa?	Thường xuyên, đặc biệt là khi ăn một mình.	Tốt
Do you compare yourself to others on social media? How does it make you feel?	Sometimes I do, and it makes me feel insecure about my own achievements.	Good	Bạn có so sánh mình với người khác trên mạng xã hội không? Bạn cảm thấy thế nào?	Đôi khi tôi làm vậy, và điều đó khiến tôi cảm thấy bất an về những thành tựu của mình.	Tốt
How do you balance social media usage and academic responsibilities?	I try to set limits, like turning off notifications during study time.	Good	Bạn cân bằng việc sử dụng mạng xã hội và trách nhiệm học tập như thế nào?	Tôi cố gắng đặt ra giới hạn, chẳng hạn như tắt thông báo trong thời gian học.	Tốt

Has social media helped you in building professional networks?	Not much, but I follow some professionals in my field for inspiration.	Good	Mạng xã hội có giúp bạn xây dựng mạng lưới chuyên nghiệp không?	Không nhiều lắm, nhưng tôi theo dõi một số chuyên gia trong lĩnh vực của mình để lấy cảm hứng.	Tốt
Do you think social media addiction is a problem among students?	Yes, many of my friends struggle with spending too much time on it.	Good	Bạn có nghĩ rằng nghiện mạng xã hội là một vấn đề đối với học sinh không?	Đúng vậy, nhiều bạn bè của tôi gặp khó khăn vì dành quá nhiều thời gian vào việc đó.	Tốt
How do you feel when you take a break from social media?	I feel more productive and less stressed, but I also miss updates from friends.	Good	Bạn cảm thấy thế nào khi tạm ngừng sử dụng mạng xã hội?	Tôi cảm thấy năng suất hơn và ít căng thẳng hơn, nhưng tôi cũng nhớ	Tốt

				tin tức cập nhật từ bạn bè.	
What measures do you think could reduce the negative impact of social media on students?	Setting time limits and using apps that track usage could help.	Good	Bạn nghĩ những biện pháp nào có thể làm giảm tác động tiêu cực của mạng xã hội đối với học sinh?	Việc đặt giới hạn thời gian và sử dụng ứng dụng theo dõi mức sử dụng có thể giúp ích.	Tốt
Do you think social media affects your mental health? If yes, how?	Yes, it sometimes makes me feel anxious or lonely when I see others' highlight reels.	Good	Bạn có nghĩ mạng xã hội ảnh hưởng đến sức khỏe tinh thần của bạn không? Nếu có, thì ảnh hưởng như thế nào?	Đúng vậy, đôi khi tôi cảm thấy lo lắng hoặc cô đơn khi xem những thước phim nổi bật của người khác.	Tốt
How do you think teachers or institutions could use social media	They could create interactive groups or post useful	Good	Bạn nghĩ giáo viên hoặc tổ chức có thể sử dụng mạng xã hội để hỗ trợ	Họ có thể tạo các nhóm tương tác hoặc đăng các tài	Tốt

to support learning?	resources for students.		việc học như thế nào?	nguyên hữu ích cho sinh viên.	
Do you believe social media can enhance collaborative learning among students?	Yes, group chats and shared documents make teamwork easier.	Good	Bạn có tin rằng mạng xã hội có thể tăng cường khả năng học tập cộng tác giữa học sinh không?	Có, trò chuyện nhóm và chia sẻ tài liệu giúp làm việc nhóm dễ dàng hơn.	Tốt
What advice would you give to other students about using social media responsibly?	Prioritize academic goals and limit social media use during study hours.	Good	Bạn có thể đưa ra lời khuyên gì cho các sinh viên khác về việc sử dụng mạng xã hội một cách có trách nhiệm?	Ưu tiên các mục tiêu học tập và hạn chế sử dụng mạng xã hội trong giờ học.	Tốt
Do you think social media usage affects your self-esteem? If yes, how?	Yes, sometimes I feel insecure when comparing myself to others on	Good	Bạn có nghĩ việc sử dụng mạng xã hội ảnh hưởng đến lòng tự trọng của bạn không? Nếu có, thì ảnh	Đúng vậy, đôi khi tôi cảm thấy bất an khi so sánh mình với người khác	Tốt

	social media, which lowers my self-esteem.		hưởng như thế nào?	trên mạng xã hội, điều này làm giảm lòng tự trọng của tôi.	
Have you ever created a fake account on social media? If yes, why?	No, I prefer to use my real identity online.	Good	Bạn đã bao giờ tạo tài khoản giả trên mạng xã hội chưa? Nếu có, tại sao?	Không, tôi thích sử dụng danh tính thật của mình trên mạng.	Tốt
How do you feel when you receive fewer likes or comments than expected on your posts?	I feel disappointed and sometimes question my self-worth.	Good	Bạn cảm thấy thế nào khi bài đăng của bạn nhận được ít lượt thích hoặc bình luận hơn mong đợi?	Tôi cảm thấy thất vọng và đôi khi nghi ngờ giá trị bản thân.	Tốt
Do you secretly check someone else's social media profile (e.g., crush,	Yes, I sometimes check my ex-partner's profile, about once a week.	Good	Bạn có bí mật kiểm tra trang cá nhân trên mạng xã hội của người khác (ví dụ: người	Có, đôi khi tôi kiểm tra trang cá nhân của người yêu cũ, khoảng	Tốt

ex-partner)? How often?			yêu, người yêu cũ) không? Tần suất như thế nào?	một lần một tuần.	
Do you believe social media addiction is a real issue among students? Why or why not?	Yes, because many students spend excessive time online, neglecting studies and real-life interactions.	Good	Bạn có tin rằng nghiện mạng xã hội là một vấn đề thực sự đối với học sinh không? Tại sao và tại sao không?	Có, vì nhiều sinh viên dành quá nhiều thời gian trực tuyến, bỏ bê việc học và tương tác ngoài đời thực.	Tốt
How does social media influence your mood on a daily basis?	It boosts my mood when I see positive content but makes me anxious if I encounter negative posts.	Good	Mạng xã hội ảnh hưởng đến tâm trạng của bạn hàng ngày như thế nào?	Tôi cảm thấy tâm trạng tốt hơn khi thấy nội dung tích cực nhưng lại lo lắng nếu gặp phải bài đăng tiêu cực.	Tốt

Have you ever tried to reduce your social media usage? Were you successful?	Yes, but I relapsed after a few days due to fear of missing out (FOMO).	Good	Bạn đã bao giờ thử giảm việc sử dụng mạng xã hội chưa? Bạn có thành công không?	Có, nhưng tôi đã tái nghiện sau vài ngày vì sợ bỏ lỡ (FOMO).	Tốt
Do you think the number of followers/friends you have reflects your social status?	Partially, but I know it's not a true measure of real-life relationships.	Good	Bạn có nghĩ số lượng người theo dõi/bạn bè phản ánh địa vị xã hội của bạn không?	Một phần thôi, nhưng tôi biết đó không phải là thước đo thực sự cho các mối quan hệ trong đời thực.	Tốt
How do you react when you see others posting about their achievements or happy moments?	I feel happy for them but sometimes envious if I'm not doing as well.	Good	Bạn phản ứng thế nào khi thấy người khác đăng về thành tích hoặc khoảnh khắc hạnh phúc của họ?	Tôi cảm thấy vui cho họ nhưng đôi khi lại ghen tị nếu mình không làm tốt như vậy.	Tốt
Do you use	Yes, it's a	Good	Bạn có sử dụng	Đúng, đó là	Tốt

social media more when you feel lonely or stressed?	way to distract myself and feel connected to others.		mạng xã hội nhiều hơn khi cảm thấy cô đơn hoặc căng thẳng không?	cách để tôi quên đi mọi thứ và cảm thấy kết nối với người khác.	
How does social media usage affect your attendance in physical classes or lectures?	Sometimes I skip lectures thinking I can catch up later through shared notes online, but it's not as effective.	Good	Việc sử dụng mạng xã hội ảnh hưởng như thế nào đến việc tham gia lớp học hoặc bài giảng trực tiếp của bạn?	Đôi khi tôi bỏ qua bài giảng vì nghĩ rằng mình có thể theo kịp sau thông qua các bài giảng chia sẻ trực tuyến, nhưng cách này không hiệu quả.	Tốt
Do you think social media helps you stay updated with the latest medical	Yes, I follow medical pages and influencers who post about	Good	Bạn có nghĩ rằng mạng xã hội giúp bạn cập nhật những tiến bộ y tế mới nhất không?	Có, tôi theo dõi các trang y khoa và những người có	Tốt

advancements ?	breakthroughs and conferences.			sức ảnh hưởng, những người đăng bài về các đột phá và hội nghị.	
What percentage of your social media time is spent on academic vs. non-academic activities?	About 40% is academic (e.g., research, discussions), and 60% is for entertainment or socializing.	Good	Bạn dành bao nhiêu phần trăm thời gian sử dụng mạng xã hội cho các hoạt động học thuật so với các hoạt động không liên quan đến học thuật?	Khoảng 40% là dành cho mục đích học thuật (ví dụ: nghiên cứu, thảo luận) và 60% dành cho mục đích giải trí hoặc giao lưu.	Tốt
Have you ever felt distracted from studying due to social media	Frequently. I often pause studying to check messages or updates,	Good	Bạn có bao giờ cảm thấy mất tập trung khi học vì thông báo trên mạng xã hội không?	Thường xuyên. Tôi thường dừng học để kiểm tra tin nhắn hoặc	Tốt

notifications?	which breaks my focus.			cập nhật, điều này làm mất tập trung.	
Do you use social media to connect with medical professionals or mentors?	Occasionally, I join LinkedIn groups or Twitter threads where professionals share insights.	Good	Bạn có sử dụng mạng xã hội để kết nối với chuyên gia y tế hoặc người cố vấn không?	Thỉnh thoảng, tôi tham gia các nhóm LinkedIn hoặc chủ đề Twitter nơi các chuyên gia chia sẻ hiểu biết sâu sắc.	Tốt
How has social media influenced your preparation for exams?	It's helpful for quick doubt-solving but sometimes leads to procrastination.	Good	Mạng xã hội đã ảnh hưởng đến việc chuẩn bị cho kỳ thi của bạn như thế nào?	Nó hữu ích để giải quyết nhanh chóng những nghi ngờ nhưng đôi khi lại dẫn đến sự trì hoãn.	Tốt
Do you think social media	Yes, but with guidelines to	Good	Bạn có nghĩ rằng các nền	Có, nhưng phải có	Tốt

platforms should be formally integrated into medical education?	ensure professionalism and avoid misinformation.		tăng truyền thông xã hội nên được tích hợp chính thức vào giáo dục y tế không?	hướng dẫn để đảm bảo tính chuyên nghiệp và tránh thông tin sai lệch.	
Have you ever experienced anxiety or stress due to social media comparisons with peers?	Yes, seeing others post about their study progress makes me feel behind sometimes.	Good	Bạn đã bao giờ cảm thấy lo lắng hoặc căng thẳng vì bị so sánh với bạn bè trên mạng xã hội chưa?	Đúng vậy, đôi khi tôi cảm thấy mình tụt hậu khi thấy người khác đăng về tiến độ học tập của mình.	Tốt
Do you participate in online medical quizzes or challenges on social media?	Yes, they're fun and help reinforce knowledge, but I don't rely on them solely.	Good	Bạn có tham gia các câu đố y tế trực tuyến hoặc thử thách trên mạng xã hội không?	Đúng, chúng rất thú vị và giúp củng cố kiến thức, nhưng tôi không chỉ dựa vào chúng.	Tốt
How do you	I cross-check	Good	Làm thế nào để	Tôi kiểm tra	Tốt

verify the accuracy of medical information shared on social media?	with textbooks or trusted sites like PubMed before accepting any claims.		xác minh tính chính xác của thông tin y tế được chia sẻ trên mạng xã hội?	lại với sách giáo khoa hoặc các trang web đáng tin cậy như PubMed trước khi chấp nhận bất kỳ khiêu nại nào.	
Has social media usage affected your sleep schedule during exam periods?	Yes, late-night scrolling reduces my sleep, impacting next-day productivity.	Good	Việc sử dụng mạng xã hội có ảnh hưởng đến giấc ngủ của bạn trong thời gian thi không?	Đúng vậy, việc lướt web vào đêm khuya làm giảm giấc ngủ của tôi, ảnh hưởng đến năng suất làm việc vào ngày hôm sau.	Tốt
Have you ever abstained from social	Yes, during midterms. My grades	Good	Bạn đã bao giờ kiêng sử dụng mạng xã hội để	Có, trong kỳ thi giữa kỳ. Điểm	Tốt

media to improve academic performance? What was the outcome?	improved due to better focus and time management.		cải thiện thành tích học tập chưa? Kết quả thế nào?	của tôi được cải thiện nhờ tập trung và quản lý thời gian tốt hơn.	
How do you verify the credibility of academic information shared on social media?	I cross-check with textbooks or reputable academic sites like Google Scholar.	Good	Làm thế nào để xác minh độ tin cậy của thông tin học thuật được chia sẻ trên mạng xã hội?	Tôi đối chiếu với sách giáo khoa hoặc các trang web học thuật có uy tín như Google Scholar.	Tốt
What social media features distract you the most from studying?	Endless scrolling on Instagram Reels and Facebook notifications disrupt my concentration.	Good	Những tính năng nào trên mạng xã hội khiến bạn mất tập trung nhất khi học?	Việc liên tục cuộn trên Instagram Reels và thông báo trên Facebook làm mất tập	Tốt

				trung của tôi.	
How much time do you spend each day on social media on your mobile phone?	About 2-3 hours, mainly for surfing news and communication.	Good	Bạn dành bao nhiêu thời gian mỗi ngày cho MXH trên điện thoại di động?	Khoảng 2-3 giờ, chủ yếu để lướt tin tức và liên lạc.	Tốt
Do you feel that using social media affects your grades?	Yes, sometimes it reduces study time, resulting in lower scores.	Good	Bạn có cảm thấy việc sử dụng MXH ảnh hưởng đến điểm số của bạn không?	Có, đôi khi làm giảm thời gian học, dẫn đến điểm số thấp hơn.	Tốt
Do you often check social networks during class?	Occasionally, about 1-2 times per session.	Good	Bạn có thường xuyên kiểm tra MXH trong giờ học không?	Thỉnh thoảng, khoảng 1-2 lần mỗi buổi học.	Tốt
Does using social media distract you from studying?	Yes, social media can be distracting when you need to	Good	Việc sử dụng MXH có làm bạn mất tập trung khi học không?	Có, MXH dễ gây phân tâm khi cần tập trung.	Tốt

	concentrate.				
Do you use social media to support your studies?	Yes, join study groups and find resources on social media.	Good	Bạn có sử dụng MXH để hỗ trợ việc học không?	Có, tham gia nhóm học tập và tìm tài liệu trên MXH.	Tốt
Do you feel anxious when you can't access social media?	Sometimes mild anxiety, especially when waiting for important messages.	Good	Bạn có cảm thấy lo lắng khi không thể truy cập MXH không?	Đôi khi lo lắng nhẹ, đặc biệt khi chờ tin nhắn quan trọng.	Tốt
Do you compare yourself to others on social media?	Yes, sometimes compare, feel inferior when seeing others successful.	Good	Bạn có so sánh bản thân với người khác trên MXH không?	Có, thỉnh thoảng so sánh, cảm thấy tự ti khi thấy người khác thành công.	Tốt
Does using social media make you feel lonely?	Sometimes, when I see my friends seem happier on social media.	Good	Việc sử dụng MXH có khiến bạn cảm thấy cô đơn không?	Đôi khi, khi thấy bạn bè có vẻ hạnh phúc hơn trên MXH.	Tốt
Have you	No, never	Good	Bạn có từng trải	Không,	Tốt

ever experienced bullying on social media?	experienced cyberbullying.		qua bắt nạt trên MXH không?	chưa từng trải qua bắt nạt trực tuyến.	
Do you feel pressured when using social media?	Yes, pressure from comparison or needing to maintain an image.	Good	Bạn có cảm thấy áp lực khi sử dụng MXH không?	Có, áp lực từ việc so sánh hoặc cần duy trì hình ảnh.	Tốt
Do you plan your social media usage time?	Tried to plan, but not followed through.	Good	Bạn có lập kế hoạch thời gian sử dụng MXH không?	Thử lập kế hoạch, nhưng chưa thực hiện đều đặn.	Tốt
Do you feel out of control of your time when using social networks?	Yes, often surf social media longer than intended.	Good	Bạn có cảm thấy mất kiểm soát thời gian khi sử dụng MXH không?	Có, thường lướt MXH lâu hơn dự định.	Tốt
Does using social media affect your study time?	Yes, reduce study time due to too much social	Good	Việc sử dụng MXH có ảnh hưởng đến thời gian học tập	Có, giảm thời gian học do lướt MXH quá	Tốt

	media surfing.		của bạn không?	nhiều.	
Do you often procrastinate on work due to social media?	Yes, social media often makes me procrastinate on my homework.	Good	Bạn có thường xuyên trì hoãn công việc do MXH không?	Có, MXH thường khiến tôi trì hoãn bài tập.	Tốt
Do you use tools to limit your time on social media?	Tried apps like Forest, but haven't stuck with it.	Good	Bạn có sử dụng công cụ để giới hạn thời gian trên MXH không?	Đã thử ứng dụng như Forest, nhưng chưa kiên trì.	Tốt
How do you use social media to keep in touch with friends and family?	Use text messages and group chats for regular communication.	Good	Bạn sử dụng MXH để giữ liên lạc với bạn bè và gia đình như thế nào?	Dùng tin nhắn và nhóm chat để liên lạc thường xuyên.	Tốt
Do you feel social media helps you connect with others?	Yes, it makes it easier to stay in touch with friends and family.	Good	Bạn có cảm thấy MXH giúp bạn kết nối với người khác không?	Có, giúp giữ liên lạc với bạn bè và gia đình dễ dàng hơn.	Tốt
Have you	Rare, just a	Good	Bạn có từng	Hiếm, chỉ	Tốt

ever had a conflict with someone on social media?	few small arguments on the post.		gặp mâu thuẫn với ai đó trên MXH không?	vài lần tranh luận nhỏ trên bài đăng.	
Do you feel like your relationship is being affected by social media?	Sometimes, due to lack of face-to-face interaction with friends.	Good	Bạn có cảm thấy mối quan hệ của mình bị ảnh hưởng bởi MXH không?	Đôi khi, do thiếu tương tác trực tiếp với bạn bè.	Tốt
Do you use social media to participate in groups or communities?	Yes, join study and interest groups on social media.	Good	Bạn có sử dụng MXH để tham gia vào các nhóm hoặc cộng đồng không?	Có, tham gia nhóm học tập và sở thích trên MXH.	Tốt
Do you feel like you are addicted to social media?	Feels dependent at times, especially when surfing mindlessly.	Good	Bạn có cảm thấy mình nghiện MXH không?	Có lúc cảm thấy phụ thuộc, đặc biệt khi lướt vô thức.	Tốt
Have you ever tried to cut down on	Yes, tried time limit but hard to	Good	Bạn có từng cố gắng giảm thời gian sử dụng	Có, đã thử giới hạn thời gian	Tốt

your social media usage?	maintain.		MXH không?	nhưng khó duy trì.	
Do you feel frustrated when you can't access social networks?	Sometimes annoying, but can get used to.	Good	Bạn có cảm thấy khó chịu khi không thể truy cập MXH không?	Đôi khi khó chịu, nhưng có thể quen dần.	Tốt
Do you check social media regularly even when there are no new notifications?	Yes, check regularly about 5-6 times/day.	Good	Bạn có thường xuyên kiểm tra MXH ngay cả khi không có thông báo mới không?	Có, kiểm tra theo thói quen khoảng 5-6 lần/ngày.	Tốt
Do you feel like social media use affects your sleep?	Yes, surfing social media before bed makes it harder to fall asleep.	Good	Bạn có cảm thấy việc sử dụng MXH ảnh hưởng đến giấc ngủ của bạn không?	Có, lướt MXH trước khi ngủ làm khó ngủ hơn.	Tốt
Do you use social media to search for career opportunities?	Yes, use LinkedIn to find jobs and make professional	Good	Bạn có sử dụng MXH để tìm kiếm cơ hội nghề nghiệp không?	Có, qua LinkedIn để tìm việc làm và kết nối chuyên	Tốt

	connections.			nghịệp.	
Do you follow experts in your field on social media?	Yes, follow to learn knowledge and experience.	Good	Bạn có theo dõi các chuyên gia trong lĩnh vực của mình trên MXH không?	Có, theo dõi để học hỏi kiến thức và kinh nghiệm.	Tốt
Do you feel social media helps you develop your professional skills?	Yes, through posts and courses shared on social media.	Good	Bạn có cảm thấy MXH giúp bạn phát triển kỹ năng nghề nghiệp không?	Có, qua các bài đăng và khóa học chia sẻ trên MXH.	Tốt
Do you share your academic or professional achievements on social media?	Occasionally, mainly share milestones.	Good	Bạn có chia sẻ thành tựu học tập hoặc nghề nghiệp trên MXH không?	Thỉnh thoảng, chủ yếu chia sẻ cột mốc quan trọng.	Tốt
Have you received positive feedback from employers via	No, but received an interview invitation via LinkedIn.	Good	Bạn có nhận được phản hồi tích cực từ nhà tuyển dụng qua MXH không?	Chưa, nhưng đã nhận được lời mời phỏng vấn qua	Tốt

social media?				LinkedIn.	
Are you taking steps to protect your privacy on social media?	Yes, keep your account private and limit sharing of sensitive information.	Good	Bạn có thực hiện các bước để bảo vệ quyền riêng tư trên MXH không?	Có, đặt tài khoản riêng tư và hạn chế chia sẻ thông tin nhạy cảm.	Tốt
Do you share sensitive personal information on social media?	No, I avoid sharing sensitive information.	Good	Bạn có chia sẻ thông tin cá nhân nhạy cảm trên MXH không?	Không, tôi tránh chia sẻ thông tin nhạy cảm.	Tốt
Have you ever had a privacy issue on social media?	No, no serious problems yet.	Good	Bạn có từng gặp sự cố liên quan đến quyền riêng tư trên MXH không?	Không, chưa gặp sự cố nghiêm trọng.	Tốt
Do you feel safe using social media?	Relatively safe, but still be careful with what you share.	Good	Bạn có cảm thấy an toàn khi sử dụng MXH không?	Tương đối an toàn, nhưng vẫn cẩn thận với nội dung chia sẻ.	Tốt
Do you know	Yes, I know	Good	Bạn có biết	Có, tôi biết	Tốt

how to set privacy on social media platforms?	how to set basic privacy.		cách thiết lập quyền riêng tư trên các nền tảng MXH không?	cách cài đặt quyền riêng tư cơ bản.	
Do you feel social media makes you more aware of social issues?	Yes, social media provides information about social issues.	Good	Bạn có cảm thấy MXH giúp bạn nhận thức rõ hơn về các vấn đề xã hội không?	Có, MXH cung cấp thông tin về các vấn đề xã hội.	Tốt
Do you participate in social or charity campaigns via social media?	Occasionally, mostly share or donate small amounts.	Good	Bạn có tham gia vào các chiến dịch xã hội hoặc từ thiện qua MXH không?	Thỉnh thoảng, chủ yếu chia sẻ hoặc quyên góp nhỏ.	Tốt
Do you feel social media promotes your community involvement?	Yes, it makes it easier for me to participate in community events.	Good	Bạn có cảm thấy MXH thúc đẩy sự tham gia của bạn vào cộng đồng không?	Có, giúp tôi tham gia các sự kiện cộng đồng dễ hơn.	Tốt
Do you share your personal	Rare, only share when	Good	Bạn có chia sẻ quan điểm cá	Hiếm, chỉ chia sẻ khi	Tốt

views on social issues on social media?	necessary.		nhân về các vấn đề xã hội trên MXH không?	cảm thấy cần thiết.	
Do you feel social media influences your political views?	A little bit, due to exposure to different opinions.	Good	Bạn có cảm thấy MXH ảnh hưởng đến quan điểm chính trị của bạn không?	Có một chút, do tiếp xúc với nhiều ý kiến khác nhau.	Tốt
Do you use social media to discuss homework with friends?	Yes, regular chat group exchange.	Good	Bạn có sử dụng MXH để trao đổi bài tập với bạn bè không?	Có, thường xuyên trao đổi qua nhóm chat.	Tốt
Do you feel social media helps you understand lessons better?	Sometimes, through documents and group discussions on social networks.	Good	Bạn có cảm thấy MXH giúp bạn hiểu bài học tốt hơn không?	Đôi khi, qua tài liệu và thảo luận nhóm trên MXH.	Tốt
Are you often distracted by social media while	Yes, social media easily distracts me when	Good	Bạn có thường xuyên bị xao nhãng bởi MXH khi học	Có, MXH dễ làm tôi mất tập trung khi	Tốt

studying?	studying.		tập không?	học.	
Do you use social media to search for study information?	Yes, find documentation and video tutorials on social media.	Good	Bạn có sử dụng MXH để tìm kiếm thông tin học tập không?	Có, tìm tài liệu và video hướng dẫn trên MXH.	Tốt
Do you feel that using social media affects your academic performance?	Yes, it reduces performance due to time loss.	Good	Bạn có cảm thấy việc sử dụng MXH ảnh hưởng đến hiệu suất học tập không?	Có, làm giảm hiệu suất do mất thời gian.	Tốt
Do you feel that using social media affects your sleep?	Yes, surfing social media late at night reduces sleep quality.	Good	Bạn có cảm thấy việc sử dụng MXH ảnh hưởng đến giấc ngủ không?	Có, lướt MXH khuya làm giảm chất lượng giấc ngủ.	Tốt
Do you feel tired after using social networks for a long time?	Yes, feel tired and lose energy.	Good	Bạn có cảm thấy mệt mỏi sau khi sử dụng MXH trong thời gian dài không?	Có, cảm thấy mệt và mất năng lượng.	Tốt
Do you feel	Yes,	Good	Bạn có cảm	Có, đặc biệt	Tốt

eye pain or headache after using social media?	especially when using the phone for too long.		thấy đau mắt hoặc đau đầu sau khi sử dụng MXH không?	khi dùng điện thoại quá lâu.	
Do you take time to rest after using social media?	Occasionally, but often forget to rest.	Good	Bạn có dành thời gian để nghỉ ngơi sau khi sử dụng MXH không?	Thỉnh thoảng, nhưng thường quên nghỉ ngơi.	Tốt
Do you do exercises to minimize the negative effects of social media?	Rarely, but sometimes light exercise is used to relieve stress.	Good	Bạn có thực hiện các bài tập thể dục để giảm thiểu tác động tiêu cực của MXH không?	Hiếm, nhưng đôi khi tập nhẹ để giảm căng thẳng.	Tốt
What do you mainly use social media for?	Communicate, entertain, and find educational information.	Good	Bạn chủ yếu sử dụng MXH để làm gì?	Liên lạc, giải trí, và tìm thông tin học tập.	Tốt
Do you use social media for entertainment	Yes, watch funny videos and posts.	Good	Bạn có sử dụng MXH để giải trí không?	Có, xem video và bài đăng hài hước.	Tốt

?					
Do you use social media to stay updated with news?	Yes, follow the news through social networking sites.	Good	Bạn có sử dụng MXH để cập nhật tin tức không?	Có, theo dõi tin tức qua các trang MXH.	Tốt
Do you use social media to connect with friends?	Yes, mainly to keep in touch with friends.	Good	Bạn có sử dụng MXH để kết nối với bạn bè không?	Có, chủ yếu để giữ liên lạc với bạn bè.	Tốt
Do you use social media for studying?	Yes, join study groups and find materials.	Good	Bạn có sử dụng MXH để học tập không?	Có, tham gia nhóm học và tìm tài liệu.	Tốt
How do you rate the overall impact of social media on your life?	Both benefits (connection, learning) and disadvantages (distraction, stress).	Good	Bạn đánh giá tổng thể tác động của MXH đến cuộc sống của bạn như thế nào?	Cả lợi ích (kết nối, học tập) và tác hại (phân tâm, áp lực).	Tốt
Do you feel that social media brings more benefits than harm?	It depends on the time, but the effects are often more prominent if	Good	Bạn có cảm thấy MXH mang lại nhiều lợi ích hơn là tác hại không?	Tùy lúc, nhưng tác hại thường nổi bật hơn nếu không	Tốt

	left unchecked.			kiểm soát.	
Do you have any suggestions to help students use social media more effectively?	Set time limits, prioritize studying, and use social media for positive purposes.	Good	Bạn có đề xuất nào để giúp sinh viên sử dụng MXH hiệu quả hơn không?	Đặt giới hạn thời gian, ưu tiên học tập, sử dụng MXH cho mục đích tích cực.	Tốt
Have you ever received learning support from communities/groups on social networks?	Yes, get materials and advice from the study group.	Good	Bạn có từng nhận được sự hỗ trợ học tập từ các cộng đồng/nhóm trên MXH không?	Có, nhận được tài liệu và lời khuyên từ nhóm học tập.	Tốt
Do you feel guilty about the time you spend on social media?	Yes, especially when procrastinating on school work.	Good	Bạn có cảm thấy tội lỗi về thời gian bạn dành cho MXH không?	Có, đặc biệt khi trì hoãn công việc học tập.	Tốt

1.3. Sinh thêm dữ liệu:

Nhóm lựa chọn 5 luật sau để áp dụng vào từng câu hỏi trong 120 câu hỏi gốc và sinh ra 600 câu hỏi mới. Những luật này giúp biến đổi câu hỏi ban đầu thành những câu mới có nội dung phong phú hơn, cụ thể 5 luật là:

- Luật đồng nghĩa: thay thế từ/cụm từ trong câu hỏi bằng từ/cụm từ đồng nghĩa.
- Luật thay đổi góc nhìn: biến đổi câu hỏi bằng cách thay đổi chủ thể, đối tượng hoặc cách tiếp cận sự việc.
- Luật cụ thể hóa: thay thế từ/cụm từ chung chung trong câu hỏi bằng từ/cụm từ chi tiết cụ thể hơn.
- Luật tổng quát hóa: thay thế từ/cụm từ chi tiết bằng khái niệm rộng hơn.
- Luật đảo ngược: thay đổi trật tự từ/cấu trúc trong câu hỏi để đảo ngược các quan hệ giữa các thành phần.

Ví dụ:

Câu hỏi gốc: “Facebook ảnh hưởng tích cực đến thời gian học tập trong cuộc sống thường ngày của bạn như thế nào?” sau khi áp dụng các luật thì kết quả thu được như trong Bảng 1.3 bên dưới.

Bảng 1.3. Ví dụ về áp dụng 5 luật để sinh thêm câu hỏi.

Luật	Câu hỏi được sinh ra
Luật đồng nghĩa	Facebook giúp quản lý thời gian học tập hiệu quả như thế nào?
Luật thay đổi góc nhìn	Giáo viên đánh giá thế nào về việc dùng Facebook để quản lý thời gian học?
Luật cụ thể hóa	Cụ thể tính năng Nhóm trên Facebook giúp bạn học nhóm môn Toán ra sao?
Luật tổng quát hóa	Các nền tảng mạng xã hội ảnh hưởng thế nào đến quản lý thời gian?

Luật đảo ngược	Những mặt tiêu cực của Facebook đối với việc học là gì?
-----------------------	---

1.4. Đánh giá dữ liệu sinh thêm:

Để đánh giá bộ dữ liệu được ChatGPT sinh thêm một cách tổng quan và chính xác, chúng tôi đã đặt ra các tiêu chí đánh giá sau:

- Tính tự nhiên của ngôn ngữ: Tiêu chí này đánh giá mức độ trôi chảy, gần gũi với cách nói hàng ngày của người Việt. Câu hỏi phải tránh lỗi diễn đạt cứng nhắc, thiếu logic hoặc dùng từ ngữ không phù hợp với văn phong giao tiếp.
- Tính phù hợp với ngữ cảnh: Đảm bảo câu hỏi luôn bám sát chủ đề chính "Ảnh hưởng của mạng xã hội đến đời sống sinh viên", không lan man sang vấn đề khác.
- Tính cân bằng góc nhìn: Câu hỏi cần bao quát cả mặt tích cực (kết nối, học hỏi) và tiêu cực (nghiện, stress) của mạng xã hội, tránh thiên vị một chiều.
- Tính sáng tạo: Câu hỏi mới có đưa ra cách tiếp cận khác biệt so với câu gốc nhờ áp dụng linh hoạt 5 luật biến đổi, nhưng vẫn giữ nguyên ý nghĩa cốt lõi.
- Độ phức tạp của câu hỏi: Đánh giá sự đa dạng về mức độ dễ/khó của câu hỏi, từ đơn giản (đóng/mở) đến phức tạp (đòi hỏi suy luận).
- Tính khách quan: Câu hỏi phải trung lập, không áp đặt câu trả lời hoặc chứa ngôn từ mang tính định hướng.
- Khả năng ứng dụng thực tế: Câu hỏi phải có thể trả lời được trong thực tế, không quá mơ hồ hoặc yêu cầu thông tin không tồn tại.

Từ những tiêu chí đã được xác định, nhóm chọn ra 15% số câu hỏi được ChatGPT sinh ra (90 câu) để tiến hành đánh giá theo thang đo từ 1 đến 5 theo quy tắc từ Rất kém đến Rất tốt. Kết quả đánh giá như trong Bảng 1.4 bên dưới.

Bảng 1.4. Bảng đánh giá bộ dữ liệu sinh thêm.

Tiêu chí	Điểm trung bình
----------	-----------------

Tính tự nhiên của ngôn ngữ	3
Tính phù hợp với ngữ cảnh	4
Tính cân bằng góc nhìn	4
Tính sáng tạo	5
Độ phức tạp của câu hỏi	5
Tính khách quan	4
Khả năng ứng dụng thực tế	4
Điểm trung bình chung	4.14

1.5. Lưu bộ dữ liệu:

Lưu bộ dữ liệu đã sinh ra ở trên thành 2 bộ dữ liệu tương ứng: Data, DChatGPT. Mỗi bộ dữ liệu đều có câu hỏi/trả lời tương ứng. Nộp bộ dữ liệu lên LMS.

CHƯƠNG 2

ĐÁNH GIÁ MỨC ĐỘ HÀI LÒNG CỦA SINH VIÊN ĐỐI VỚI CHẤT LƯỢNG DỊCH VỤ CỦA THƯ VIỆN TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THÀNH PHỐ HỒ CHÍ MINH

2.1. Câu hỏi đặt ra:

Hiện nay, trường Đại học Công nghiệp Thành phố Hồ Chí Minh (IUH) đang có rất nhiều sinh viên đang theo học, nhu cầu về tham khảo tài liệu, sử dụng thiết bị công nghệ, chỗ nghỉ ngơi, không gian làm việc nhóm rất cao. Đặc biệt, IUH còn thường xuyên tổ chức các tiết học online đối với một số môn lý thuyết, trong khi lịch học của sinh viên có buổi thực hành ở trường, nên việc sử dụng thư viện để học online cũng là thực trạng đáng chú ý.

Nhận thấy việc nâng cao chất lượng dịch vụ của thư viện IUH là cần thiết và quan trọng, nhóm đã thống nhất chọn đề tài: *“Đánh giá mức độ hài lòng của sinh viên đối với chất lượng dịch vụ của thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh”* nhằm mục đích khai thác thông tin, ý kiến, đào sâu, phân tích rõ và đánh giá tổng quan về trải nghiệm sử dụng dịch vụ thư viện của sinh viên.

Để nghiên cứu và tìm hiểu sâu về mức độ hài lòng của sinh viên, nhóm xác định 2 câu hỏi nghiên cứu như sau:

- Câu hỏi 1: Mức độ hài lòng của sinh viên trường Đại học Công nghiệp Thành phố Hồ Chí Minh (IUH) đối với dịch vụ thư viện hiện tại như thế nào?
- Lý do đặt câu hỏi:
 - + Với số lượng sinh viên đông đảo tại IUH, thư viện trở thành nơi không chỉ để tra cứu tài liệu mà còn là không gian học tập, làm việc nhóm. Đặc biệt, sự kết hợp giữa học online và offline khiến sinh viên cần tiếp cận các dịch vụ thư viện linh hoạt hơn.
 - + Đa dạng dịch vụ cần đánh giá vì thư viện không chỉ cung cấp sách mà còn bao gồm các dịch vụ như wifi, máy tính, không gian học tập, không gian làm việc

nhóm... Việc đánh giá mức độ hài lòng sẽ giúp xác định dịch vụ nào đang đáp ứng tốt và dịch vụ nào cần cải thiện.

- + Phản hồi từ sinh viên là vô cùng cần thiết bởi sinh viên là đối tượng trực tiếp sử dụng dịch vụ, do đó, ý kiến của họ sẽ phản ánh chính xác nhất chất lượng thực tế của thư viện.
- Ý nghĩa của câu hỏi:
 - + Cung cấp dữ liệu định lượng và định tính: Kết quả từ câu hỏi này sẽ cho biết tỷ lệ sinh viên hài lòng/không hài lòng, đồng thời qua các câu trả lời mở, có thể thu thập được ý kiến chi tiết về từng khía cạnh dịch vụ.
 - + Làm cơ sở để cải thiện dịch vụ: Ví dụ, nếu phần lớn sinh viên không hài lòng với chất lượng wifi, nhà trường có thể đầu tư nâng cấp hệ thống mạng. Ngược lại, nếu dịch vụ cho mượn sách được đánh giá cao, có thể nhân rộng mô hình này.
- Câu hỏi 2: Những yếu tố nào ảnh hưởng đến sự hài lòng của sinh viên trường IUH đối với thư viện?
 - Lý do đặt câu hỏi:
 - + Sự hài lòng của sinh viên không chỉ phụ thuộc vào một yếu tố đơn lẻ mà là tổng hòa của nhiều khía cạnh như chất lượng tài liệu, cơ sở vật chất, thái độ nhân viên, tốc độ internet... Việc xác định rõ yếu tố nào có ảnh hưởng lớn nhất sẽ giúp nhà trường tập trung nguồn lực vào đúng trọng tâm.
 - + Với sự phát triển của công nghệ, nhu cầu của sinh viên không còn dừng lại ở việc mượn sách mà còn bao gồm truy cập tài liệu số, sử dụng không gian làm việc nhóm... Do đó, cần hiểu rõ yếu tố nào đang thực sự quan trọng trong bối cảnh hiện tại.
 - Ý nghĩa của câu hỏi:

- + Xác định yếu tố trọng tâm cần cải thiện: Ví dụ, nếu kết quả cho thấy "chất lượng wifi" là yếu tố ảnh hưởng lớn nhất đến sự hài lòng, nhà trường có thể tăng thêm số lượng modem wifi được bố trí tại thư viện hoặc nâng cấp gói dịch vụ wifi.
- + Phân nhóm đối tượng sinh viên: Câu hỏi này có thể giúp phân tích sự khác biệt về mức độ hài lòng giữa các nhóm sinh viên (ví dụ: sinh viên năm nhất với sinh viên năm cuối).
- + Thay vì chỉ tập trung vào một yếu tố, kết quả nghiên cứu có thể đề xuất nhiều giải pháp đồng bộ, ví dụ: vừa nâng cấp cơ sở vật chất, vừa cải thiện chất lượng tài liệu.

Trả lời 2 câu hỏi này, nhóm sẽ giải quyết được mục tiêu nghiên cứu mà nhóm hướng tới, từ đó có những kết luận chính xác và giải pháp phù hợp cho thực trạng đã nêu. Thông qua khảo sát này, nhóm muốn nhận được các ý kiến, đánh giá của sinh viên, thu thập những trải nghiệm sử dụng thư viện, những mong muốn của sinh viên về vấn đề cải thiện chất lượng dịch vụ. Những ý kiến phản hồi của sinh viên sẽ được chúng tôi phục vụ cho quá trình phân tích, trực quan từ đó có những nhận xét, đánh giá tổng quan, chính xác để đề xuất ra những giải pháp phù hợp.

2.2. Xây dựng câu hỏi khảo sát:

Để có thể trả lời được 2 câu hỏi nghiên cứu mà nhóm đã đặt ra, nhóm đã thực hiện khảo sát với bộ 13 câu hỏi. Các câu hỏi trong phiếu khảo sát được nhóm xác định định theo quy trình sau:

- Bước 1: Xác định mục tiêu, mục đích khảo sát:
 - + Mục tiêu: thu thập được thông tin về trải nghiệm sử dụng, nhận xét, đánh giá, ý kiến của sinh viên để phục vụ cho quá trình phân tích dữ liệu. Bên cạnh đó, nhóm cũng thu thập những mong muốn của sinh viên để đề xuất ra những giải pháp.

- + Mục đích: dữ liệu thu được của khảo sát này giúp cho việc phân tích, trực quan và đánh giá dữ liệu của nhóm trở nên có ý nghĩa, chính xác. Từ đó nhóm có những đề xuất tối ưu cho nhà trường cải thiện chất lượng dịch vụ thư viện.
- Bước 2: Xác định khía cạnh cần thiết để khảo sát:
 Các khía cạnh của thư viện mà nhóm quan tâm là:
 - Chất lượng của các thiết bị điện tử và kết nối mạng.
 - Không gian thư viện.
 - Chất lượng phục vụ của cán bộ, nhân viên.
 Một số khía cạnh khác:
 - Mong muốn của sinh viên.
- Bước 3: xây dựng bộ câu hỏi khảo sát:

Nhóm thực hiện xây dựng bộ câu hỏi khảo sát dựa theo các khía cạnh đã xác định ở trên:

- Chất lượng các thiết bị điện tử và kết nối (Nhóm I):

Câu hỏi	Lý giải
Thư viện có đủ số lượng máy tính để đáp ứng nhu cầu sử dụng của sinh viên?	Câu hỏi này giúp đánh giá năng lực đáp ứng nhu cầu cơ bản, từ đó cân nhắc đầu tư thêm trang thiết bị nếu cần thiết.
Các máy tính trong thư viện hoạt động tốt, không bị giật lag khi sử dụng?	Máy tính giật lag làm gián đoạn việc học tập, gây khó chịu. Đánh giá này giúp xác định tần suất bảo trì thiết bị và nâng cấp phần cứng/phần mềm để đảm bảo trải nghiệm mượt mà.
Bàn ghế và ổ cắm điện tại khu vực máy tính được bố trí hợp lý, thuận tiện cho	Không gian chật chội hay thiếu ổ cắm sẽ cản trở việc sử dụng laptop cá nhân. Câu

sinh viên học tập?	hỏi nhằm tối ưu hóa bố cục khu vực học tập, đảm bảo tiện nghi tối đa.
Tình trạng mạng wifi trong thư viện hoạt động ổn định và dễ dàng truy cập?	Mạng yếu khiến sinh viên không thể truy cập tài liệu online hoặc tham gia lớp học trực tuyến. Kết quả đánh giá là cơ sở để nâng cấp băng thông hoặc điều chỉnh vị trí, số lượng modem wifi.
Hệ thống máy lạnh trong thư viện hoạt động tốt?	Hệ thống điều hòa quá lạnh hoặc không đủ lạnh có thể gây khó chịu, ảnh hưởng đến hiệu quả học tập và thời gian sử dụng thư viện. Câu hỏi nhằm xác định nhiệt độ hiện tại của máy lạnh tại thư viện có phù hợp chưa để có điều chỉnh hợp lý.

Bảng 2.2.1. Câu hỏi thuộc khía cạnh Chất lượng các thiết bị điện tử và kết nối và lý giải kèm theo.

- Không gian thư viện (Nhóm II):

Câu hỏi	Lý giải
Không gian thư viện đủ rộng có thể phục vụ cho số lượng lớn sinh viên cùng lúc?	Không gian chật chội sẽ hạn chế khả năng tiếp nhận sinh viên, đặc biệt trong giờ cao điểm. Câu hỏi này giúp đánh giá nhu cầu mở rộng hoặc tối ưu hóa bố trí không gian hiện có.
Môi trường học tập trong thư viện yên tĩnh, giúp sinh viên tập trung hơn khi	Sự yên tĩnh là yếu tố then chốt để sinh viên tập trung học tập. Đánh giá này giúp

học?	xác định hiệu quả của các biện pháp giữ trật tự và phân chia khu vực học tập phù hợp.
Số lượng phòng họp nhóm đáp ứng được nhu cầu làm việc nhóm của sinh viên?	Nhu cầu làm việc nhóm ngày càng tăng, đặc biệt với phương pháp học tập hiện đại. Thiếu phòng họp sẽ buộc sinh viên phải tìm không gian khác, giảm hiệu quả học tập.
Cơ sở vật chất trong các phòng họp nhóm ở thư viện luôn được trang bị đầy đủ?	Bảng, viết, máy quạt, hệ thống chiếu sáng đầy đủ là yếu tố thiết yếu hỗ trợ làm việc nhóm hiệu quả. Đánh giá này giúp cải thiện chất lượng dịch vụ phòng họp nhóm.
Khu vực ngồi nghỉ của sinh viên tại thư viện được bố trí hợp lý?	Khu vực nghỉ ngơi thoải mái giúp sinh viên thư giãn giữa các giờ học, nâng cao hiệu suất học tập. Bố trí hợp lý cần đảm bảo tiện nghỉ nhưng không ảnh hưởng khu vực học tập.

Bảng 2.2.2. Câu hỏi thuộc khía cạnh Không gian thư viện và lý giải kèm theo.

- Chất lượng phục vụ của các bộ, nhân viên (Nhóm III):

Câu hỏi	Lý giải
Bạn có nhận xét thế nào về chất lượng phục vụ của cán bộ thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh?	Câu hỏi này nhằm đánh giá thái độ phục vụ, chuyên môn của cán bộ thư viện, Kết quả thu được sẽ phục vụ cho việc điều

Minh?	chính, cải thiện chất lượng phục vụ của đội ngũ này.
-------	--

Bảng 2.2.3. Câu hỏi thuộc khía cạnh Chất lượng phục vụ của cán bộ, nhân viên và lý giải kèm theo.

- Mong muốn của sinh viên (Nhóm III):

Câu hỏi	Lý giải
Theo bạn, thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh cần cải thiện những gì để nâng cao chất lượng dịch vụ?	Câu hỏi mở này thu thập ý kiến trực tiếp từ người sử dụng để xác định những vấn đề cần ưu tiên cải thiện. Đây là nguồn thông tin quý giá giúp thư viện đầu tư đúng trọng tâm vào các hạng mục mang lại giá trị cao nhất cho sinh viên.
Bạn có mong muốn thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh thay thế/nâng cấp thiết bị nào không?	Trong bối cảnh công nghệ phát triển, câu hỏi này giúp xác định nhu cầu cụ thể về trang thiết bị (máy tính, hệ thống wifi,...). Thông tin thu được sẽ là cơ sở để đề xuất các giải pháp phù hợp với nhu cầu thực tế của sinh viên.

Bảng 2.2.4. Câu hỏi thuộc khía cạnh Mong muốn của sinh viên và lý giải kèm theo.

2.3. Phân tích:

2.3.1. Kết quả khảo sát:

2.3.1.1. Cách thức tiến hành khảo sát:

Các câu hỏi khảo sát được nhóm thiết kế thành bảng câu hỏi khảo sát và in ra giấy. Mỗi phiếu khảo sát chứa 13 câu hỏi khảo sát mà nhóm đã xác định cùng với các câu hỏi về thông tin cá nhân của người thực hiện khảo sát.

Trong đó, các câu hỏi thuộc Nhóm I và Nhóm II được thiết kế theo thang đo Likert với mức điểm từ 1 đến 5 theo quy tắc:

- 1- Hoàn toàn không đồng ý.
- 2- Không đồng ý.
- 3- Bình thường.
- 4- Đồng ý.
- 5- Hoàn toàn đồng ý.

Với định dạng câu trả lời như trên, nhóm muốn lượng hóa mức độ hài lòng của sinh viên đối với chất lượng dịch vụ thư viện IUH. Các câu hỏi thuộc Nhóm III là các câu hỏi mở với định dạng câu trả lời là text, nhằm thu thập ý kiến, đánh giá, mong muốn của sinh viên đối với chất lượng dịch vụ thư viện IUH

Phiếu khảo sát sẽ do các thành viên trong nhóm trực tiếp gửi đến người thực hiện khảo sát tại cơ sở chính và cơ sở Nguyễn Văn Dung trường Đại học Công nghiệp Thành phố Hồ Chí Minh, người thực hiện khảo sát mất khoảng 3-5 phút để điền câu trả lời vào phiếu khảo sát. Sau đó thành viên nhóm sẽ thu lại phiếu khảo sát và kiểm tra ngay sau khi người thực hiện khảo sát điền xong.

2.3.1.2. Đối tượng khảo sát:

Đối tượng mà nhóm hướng tới là các sinh viên theo học tại trường Đại học Công nghiệp Thành phố Hồ Chí Minh bởi đây là những người trực tiếp sử dụng dịch vụ thư viện tại trường nên học có thể đưa ra những cảm nhận, đánh giá, ý kiến chính xác với thực tế và đúng với trải nghiệm của họ.

2.3.1.3. Số lượng đối tượng tham gia khảo sát:

Nhóm thực hiện khảo sát với 60 đối tượng ngẫu nhiên nhằm thu được thông tin một cách phong phú và đa dạng, tổng quan. Lượng thông tin đủ lớn có thể giúp nhóm đưa ra những nhận định khách quan, tổng thể, đề xuất những giải pháp hợp lí.

2.3.1.4. Thời gian khảo sát:

Nhóm phân chia nhiệm vụ cho các thành viên thực hiện đồng thời việc khảo sát và nhập dữ liệu vào file Excel. Thời gian kéo dài 2 ngày cho việc khảo sát và 3 tiếng cho việc nhập dữ liệu, xử lý dữ liệu bằng Python.

2.3.1.5. Kết quả thu được:

Sau quá trình khảo sát nhóm thu được đầy đủ 60 phiếu khảo sát với mỗi phiếu khảo sát được người thực hiện khảo sát điền đầy đủ thông tin và không bỏ trống. Trong các phiếu khảo sát đảm bảo có đa dạng các khoa, khóa và ngành học khác nhau nhằm đảm bảo độ khách quan và tổng quan của dữ liệu.

2.3.2. Phân tích dữ liệu:

Sau khi thực hiện khảo sát, các thành viên trong nhóm nhập dữ liệu đồng thời vào file excel. Do dữ liệu đã được kiểm tra ngay từ bước khảo sát, đảm bảo người khảo sát điền đúng và đủ các câu trả lời nên bộ dữ liệu mà nhóm thu được là không có dữ liệu trống hay thiếu. Đồng thời, định dạng trả lời của các câu hỏi được thiết dạng Likert trong phiếu khảo sát là các số từ 1 đến 5 trừ những câu hỏi mở thuộc Nhóm III là dạng text (nhóm chúng tôi sẽ phân tích trực tiếp trên dữ liệu dạng text của các câu hỏi này) nên bộ dữ liệu của nhóm có thể được sử dụng trực tiếp cho quá trình phân tích mà không cần phải tiền xử lý hay mã hóa dữ liệu.

Nhóm chúng tôi thực hiện phân tích dữ liệu để có những hiểu biết sâu sắc hơn về đặc trưng, xu hướng đánh giá, cảm nhận chung của sinh viên đối với chất lượng dịch vụ thư viện IUH. Kết quả phân tích như sau:

Câu 1: Thư viện có đủ máy tính để đáp ứng nhu cầu sử dụng của sinh viên?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về mức độ đáp ứng của số lượng máy tính đối với nhu cầu sử dụng của sinh viên. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:

1- Hoàn toàn không đồng ý.

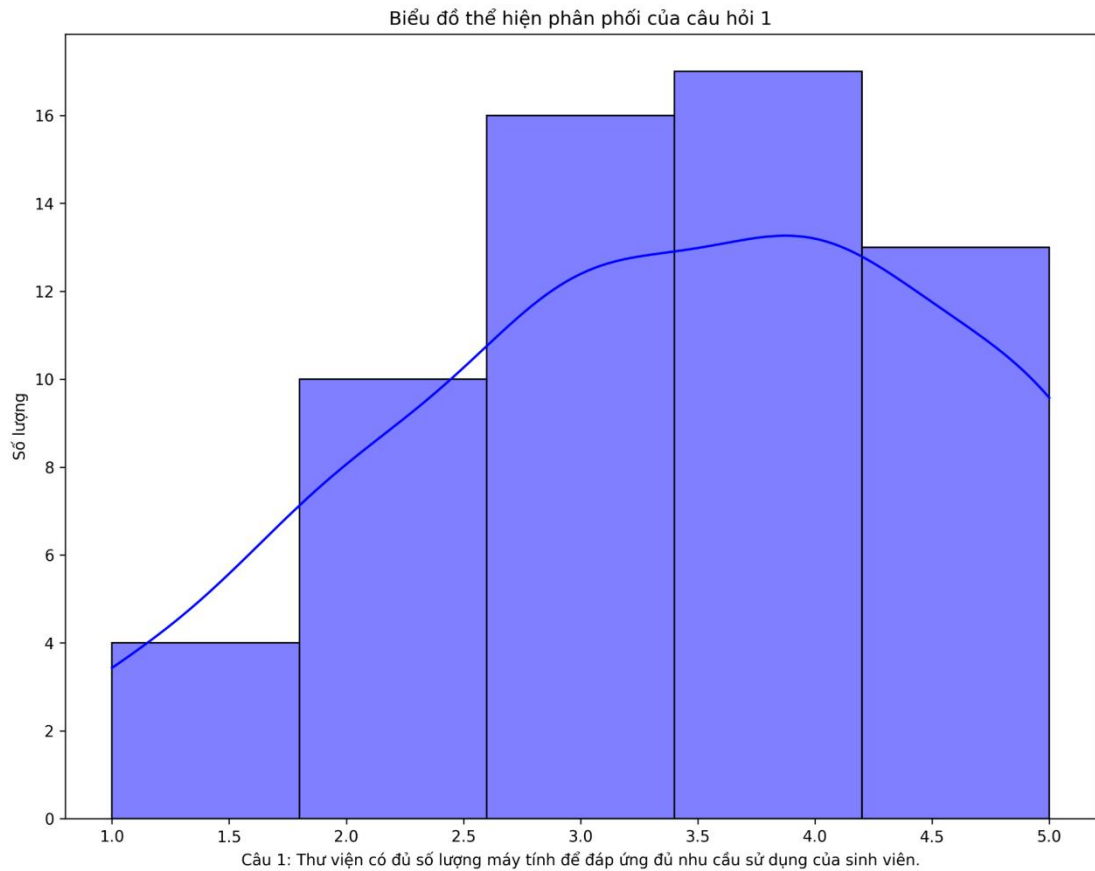
2- Không đồng ý.

3- Bình thường.

4- Đồng ý.

5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.416667
 - + std: 1.197337
 - + min : 1.000000
 - + 25%: 3.000000
 - + 50%: 3.500000
 - + 75%: 4.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: ~3.42
 - + Median: 5.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:

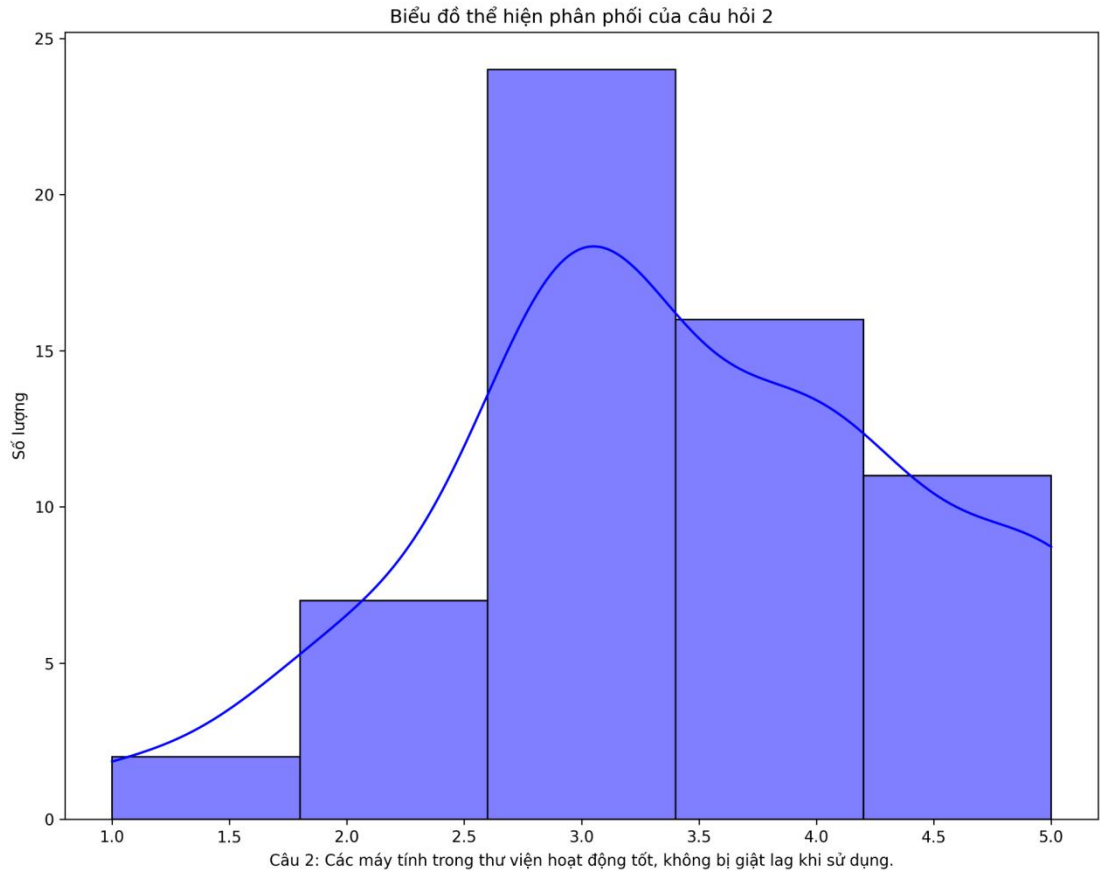


Hình 2.3.2.1. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 1.

Câu 2: Các máy tính trong thư viện hoạt động tốt, không bị giật lag khi sử dụng?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về mức độ đáp ứng của chất lượng máy tính đối với nhu cầu sử dụng của sinh viên. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.
 - 5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.450000
 - + std: 1.032112
 - + min: 1.000000
 - + 25%: 3.000000
 - + 50%: 3.000000
 - + 75%: 4.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: 3.45
 - + Median: 3.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:



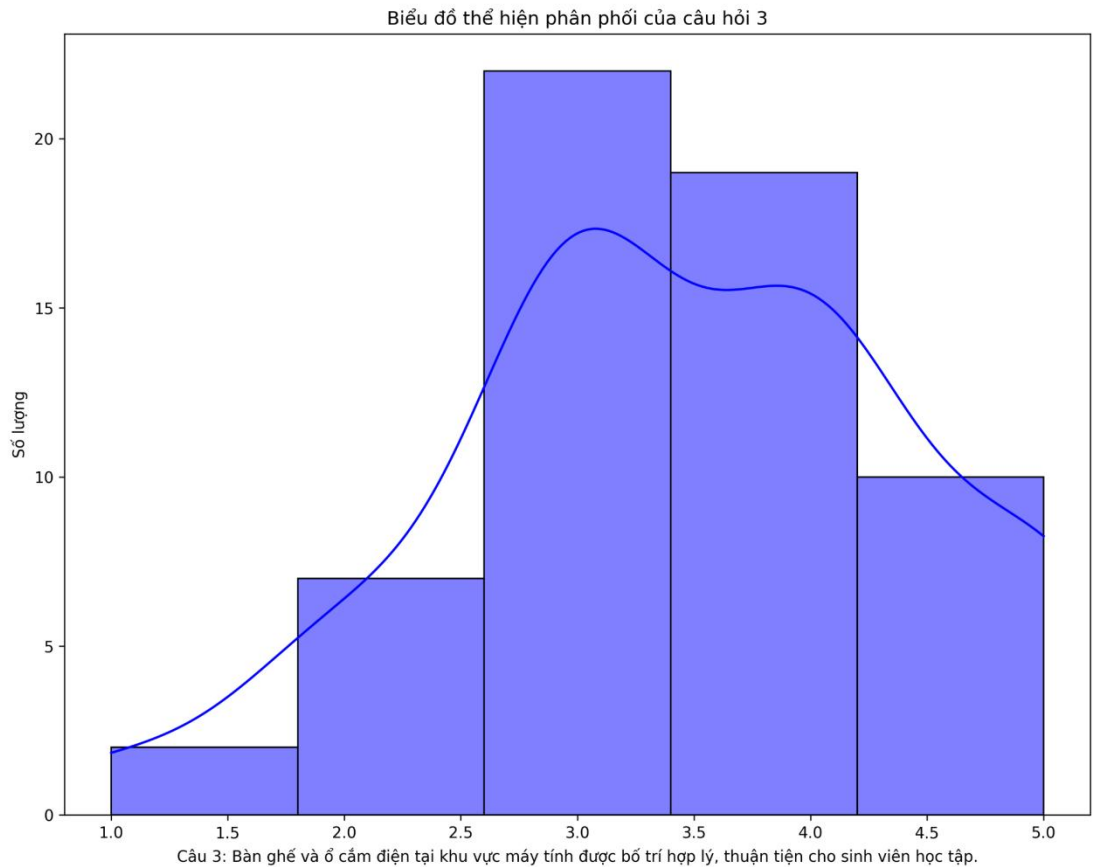
Hình 2.3.2.2. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 2.

Câu 3: Bàn ghế và ổ cắm điện tại khu vực máy tính được bố trí hợp lý, thuận tiện cho sinh viên học tập?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về mức độ hợp lý khi bố trí bàn ghế và ổ cắm điện tại khu vực máy tính. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.

5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.466667
 - + std: 1.016252
 - + min: 1.000000
 - + 25%: 3.000000
 - + 50%: 3.000000
 - + 75%: 4.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: ~3.47
 - + Median: 3.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:



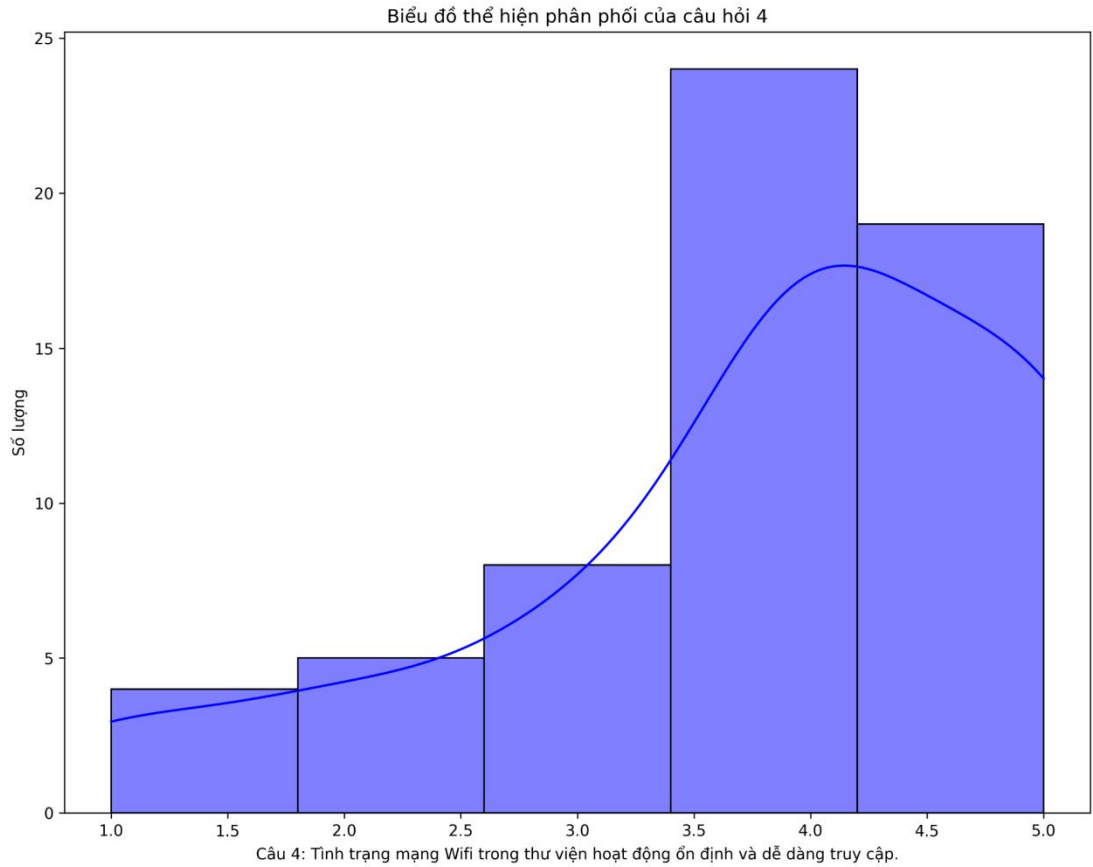
Hình 2.3.2.3. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 3.

Câu 4: Tình trạng mạng Wifi trong thư viện hoạt động ổn định và dễ dàng truy cập?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về chất lượng mạng wifi trong thư viện. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.

5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.816667
 - + std: 1.171580
 - + min: 1.000000
 - + 25%: 3.000000
 - + 50%: 4.000000
 - + 75%: 5.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: ~3.82
 - + Median: 4.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:

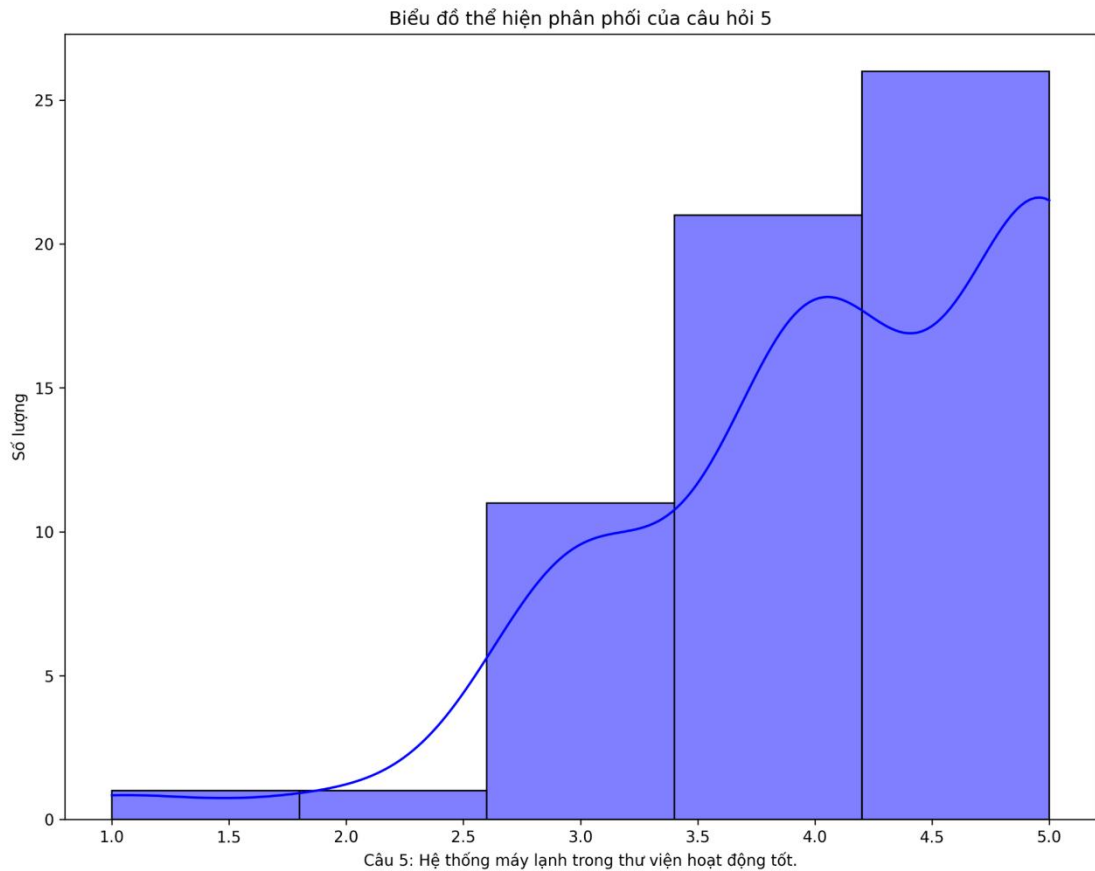


Hình 2.3.2.4. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 4.

Câu 5: Hệ thống máy lạnh trong thư viện hoạt động tốt?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về chất lượng hệ thống máy lạnh trong thư viện. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.
 - 5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 4.166667
 - + std: 0.905102
 - + min: 1.000000
 - + 25%: 4.000000
 - + 50%: 4.000000
 - + 75%: 5.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: ~4.17
 - + Median: 4.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:



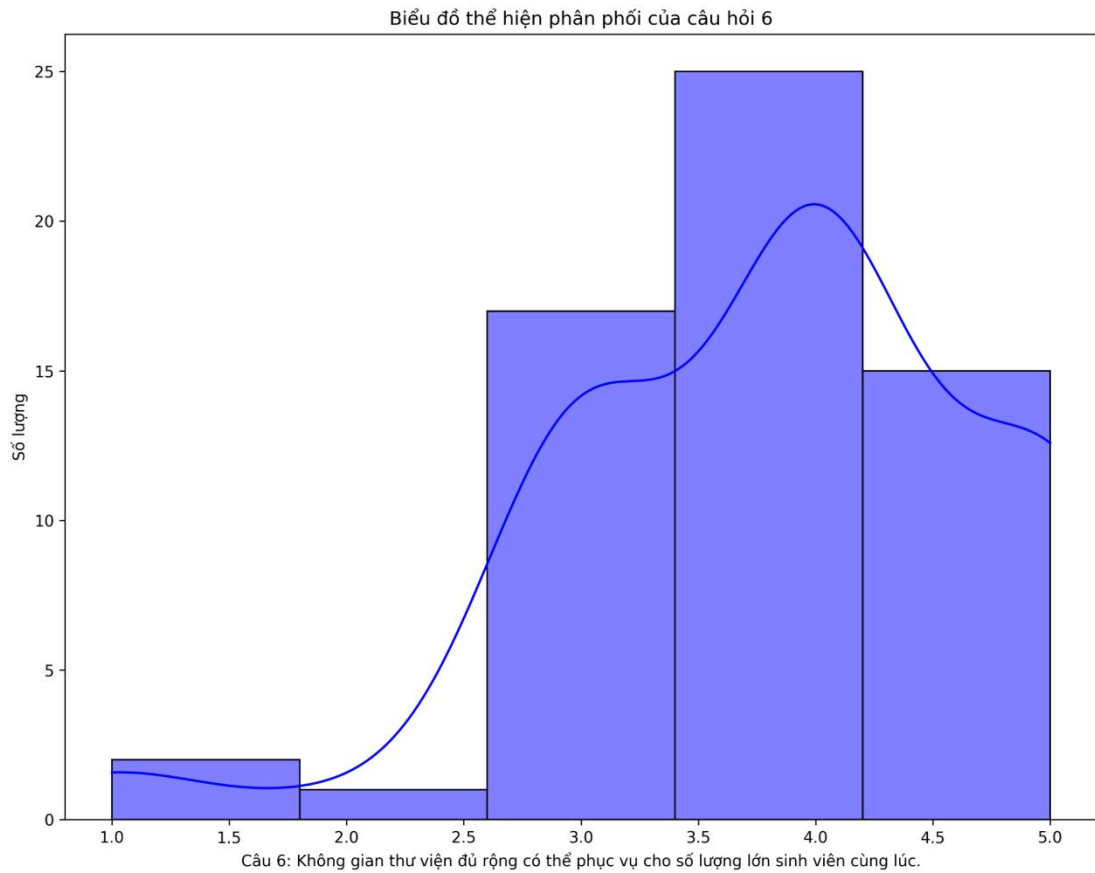
Hình 2.3.2.5. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 5.

Câu 6: Không gian thư viện đủ rộng có thể phục vụ cho số lượng lớn sinh viên cùng lúc?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về độ rộng rãi, thoải mái của thư viện đảm bảo phục vụ đủ nhu cầu của sinh viên. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.

5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.833333
 - + std: 0.941810
 - + min: 1.000000
 - + 25%: 3.000000
 - + 50%: 4.000000
 - + 75%: 4.250000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: ~3.83
 - + Median: 4.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:



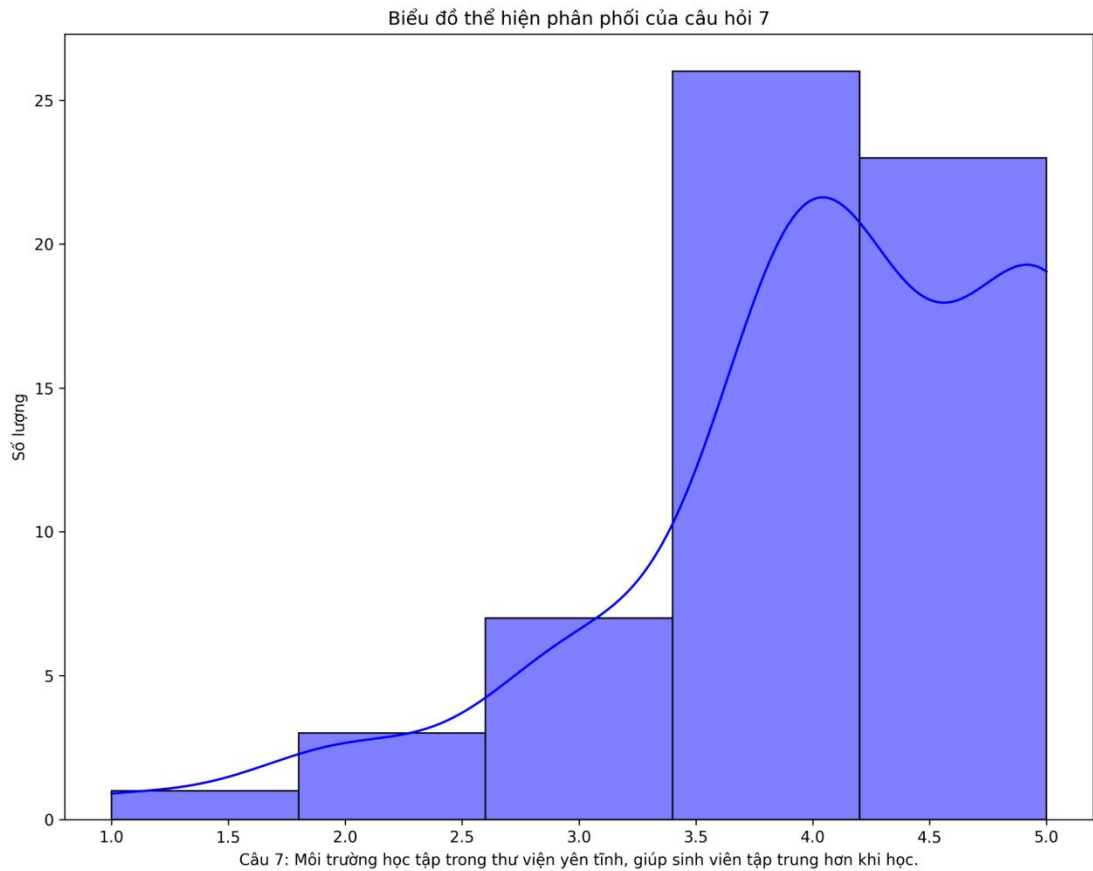
Hình 2.3.2.6. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 6.

Câu 7: Môi trường học tập trong thư viện yên tĩnh, giúp sinh viên tập trung hơn khi học?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về mức độ yên tĩnh của không gian bên trong thư viện. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.

5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 4.116667
 - + std: 0.922261
 - + min: 1.000000
 - + 25%: 4.000000
 - + 50%: 4.000000
 - + 75%: 5.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: ~4.11
 - + Median: 4.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:



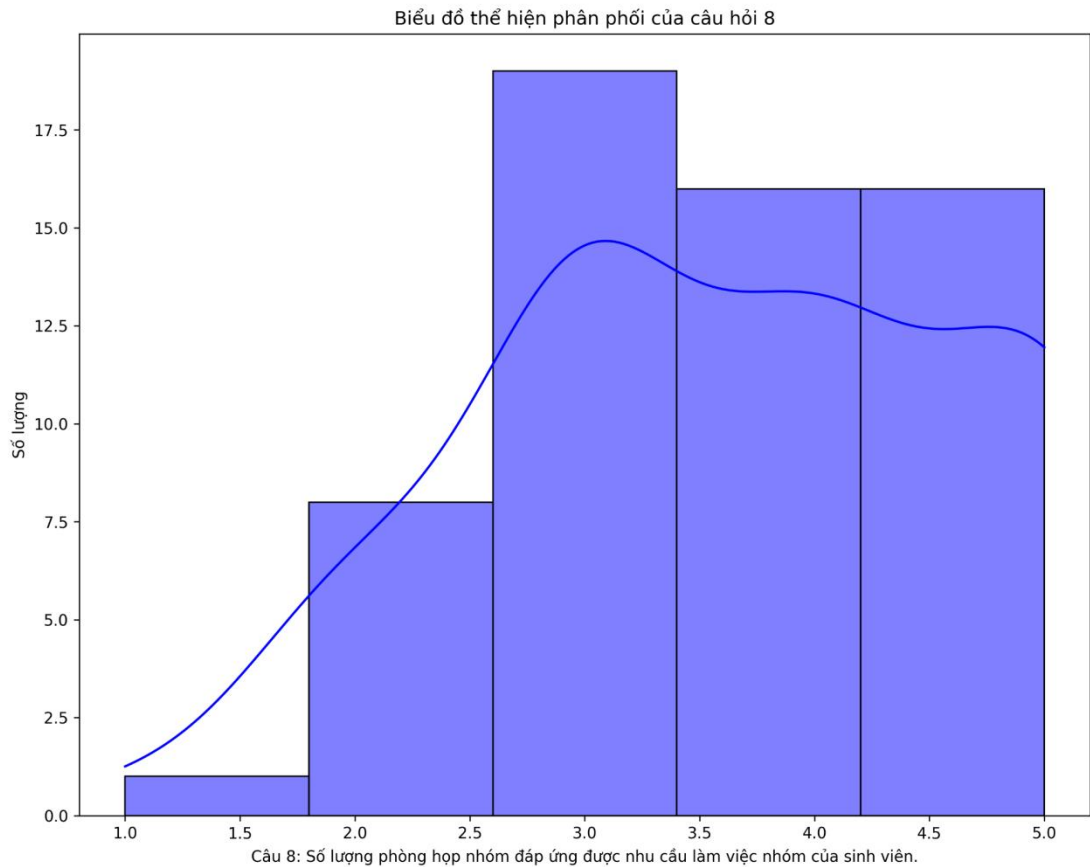
Hình 2.3.2.7. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 7.

Câu 8: Số lượng phòng họp nhóm đáp ứng được nhu cầu làm việc nhóm của sinh viên?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về mức độ đáp ứng của số lượng phòng họp nhóm đối với nhu cầu của sinh viên. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.

5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.633333
 - + std: 1.073039
 - + min: 1.000000
 - + 25%: 3.000000
 - + 50%: 4.000000
 - + 75%: 5.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: ~3.63
 - + Median: 4.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:



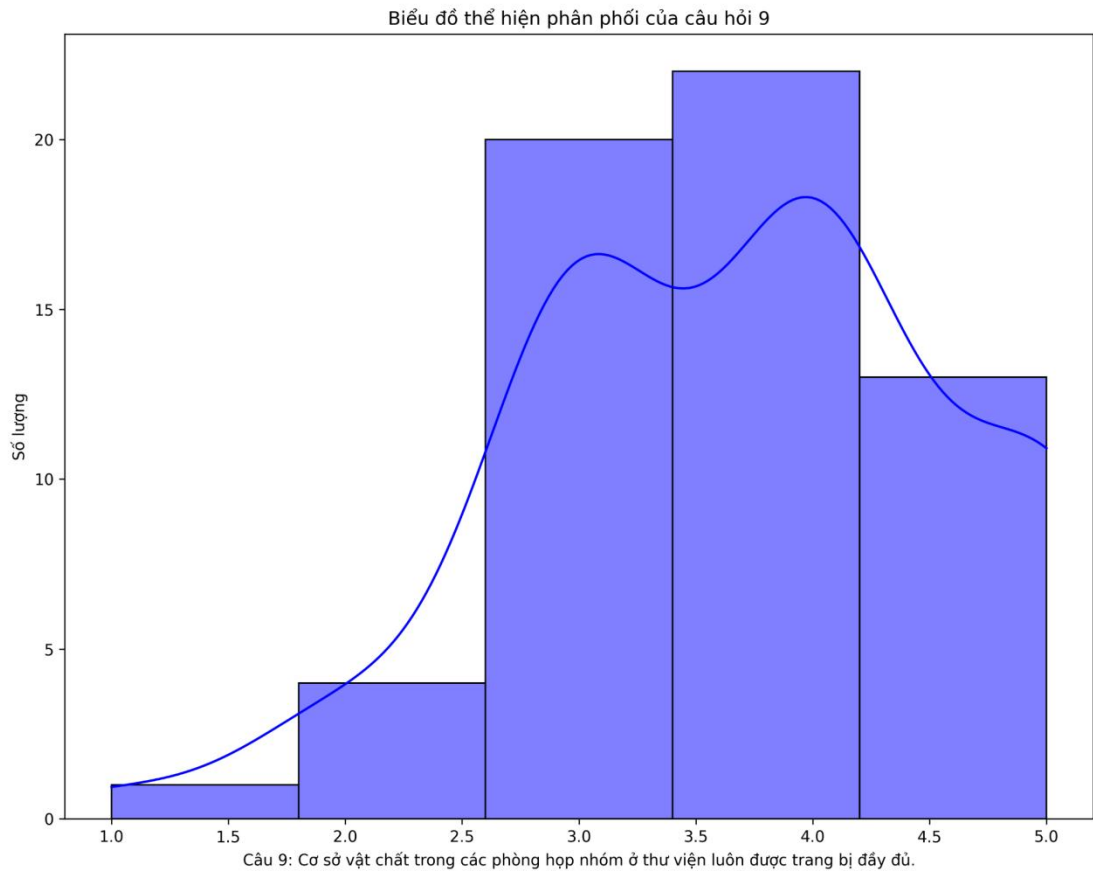
Hình 2.3.2.8. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 8.

Câu 9: Cơ sở vật chất trong các phòng họp nhóm ở thư viện luôn được trang bị đầy đủ?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về mức độ đáp ứng của cơ sở vật chất trong phòng họp nhóm. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.

5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.700000
 - + std: 0.944206
 - + min: 1.000000
 - + 25%: 3.000000
 - + 50%: 4.000000
 - + 75%: 4.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: 3.7
 - + Median: 4.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:

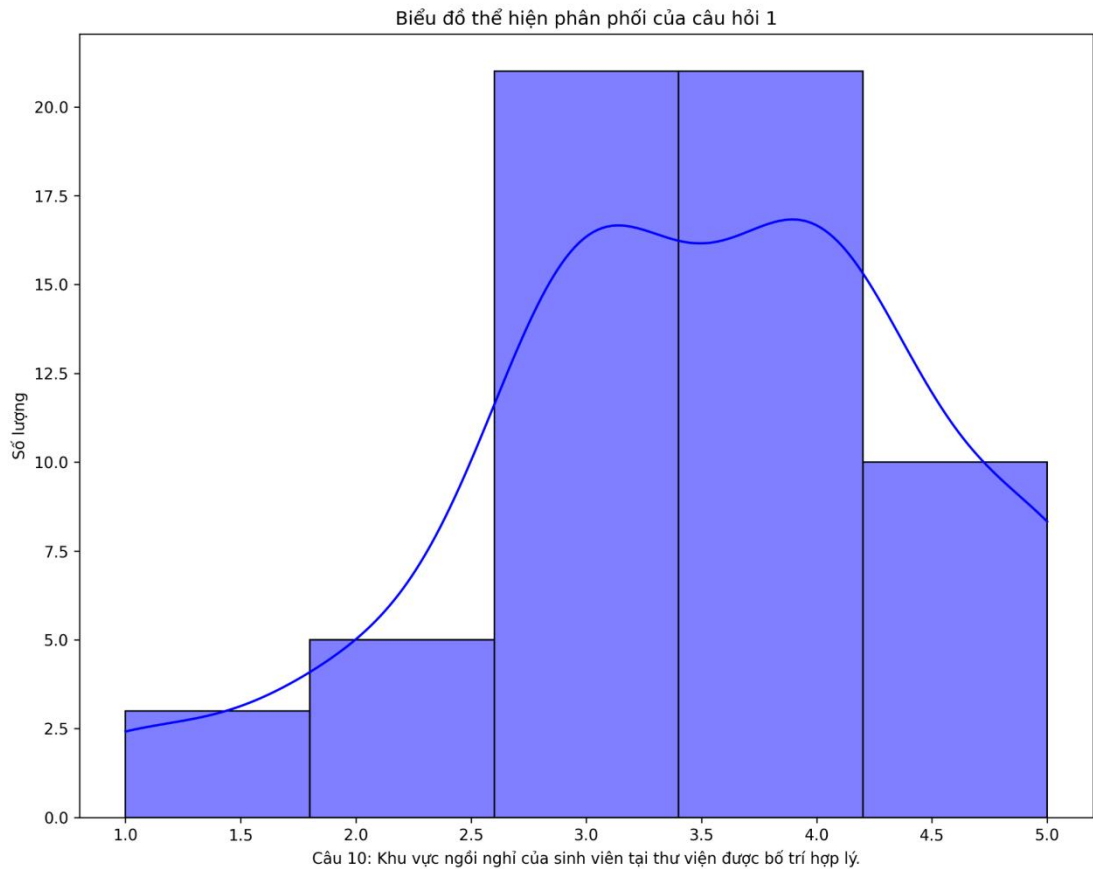


Hình 2.3.2.9. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 9.

Câu 10: Khu vực nghỉ ngơi của sinh viên tại thư viện được bố trí hợp lý?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về bố trí khu vực nghỉ ngơi dành cho sinh viên. Bộ dữ liệu là các con số từ 1 đến 5 ứng với các câu trả lời ở định dạng Likert với thang đo từ 1 đến 5 ứng với quy tắc:
 - 1- Hoàn toàn không đồng ý.
 - 2- Không đồng ý.
 - 3- Bình thường.
 - 4- Đồng ý.
 - 5- Hoàn toàn đồng ý.

- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60.000000
 - + mean: 3.500000
 - + std: 1.033342
 - + min: 1.000000
 - + 25%: 3.000000
 - + 50%: 4.000000
 - + 75%: 4.000000
 - + max: 5.000000
- Để tìm mối tương quan giữa câu hỏi khảo sát và kết quả, nhóm chúng tôi sử dụng các giá trị Mean và Median để làm ngưỡng đánh giá. Sau quá trình thực hiện, kết quả mà nhóm nhận được là:
 - + Mean: 3.5
 - + Median: 4.00
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:

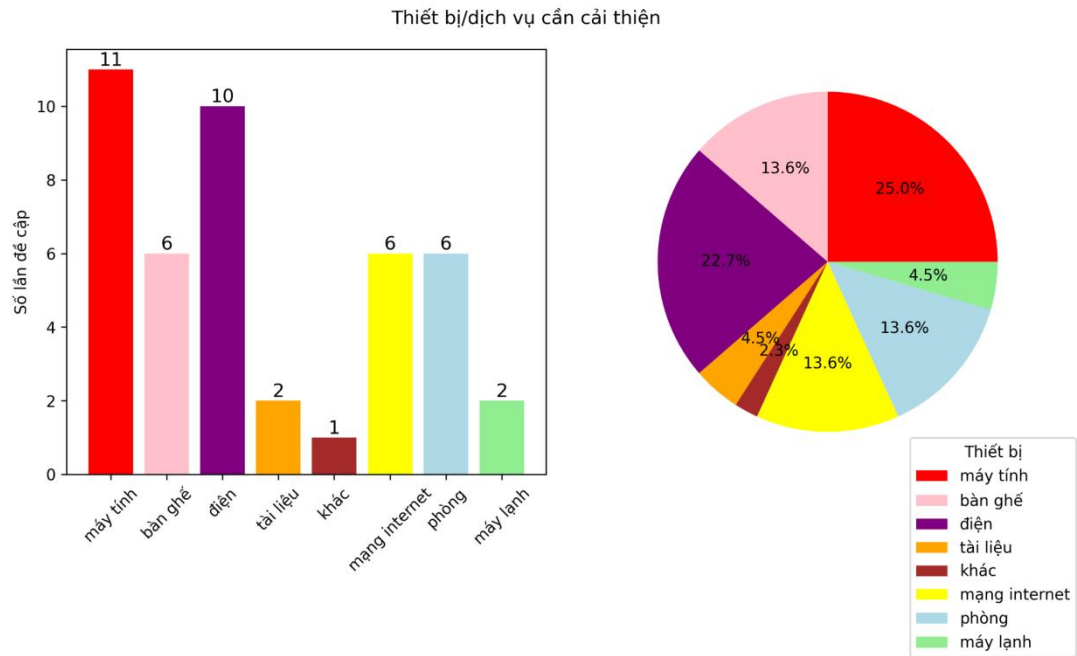


Hình 2.3.2.10. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 10.

Câu 11: Theo bạn, thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh cần cải thiện những gì để nâng cao chất lượng dịch vụ?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi ý kiến của họ về những thiết bị/dịch vụ/tài liệu hay các thứ liên quan thư viện IUH cần cải thiện để nâng cao chất lượng dịch vụ. Bộ dữ liệu là các câu trả lời ngắn dạng text.
- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60
 - + unique: 50
 - + top: 'Không có'

- + freq: 7
- Tìm mối tương quan giữa câu hỏi và kết quả:
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:

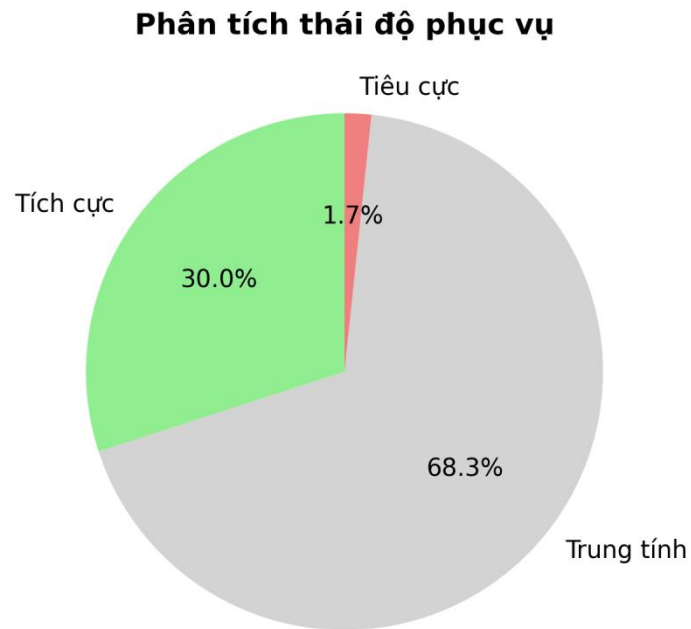


Hình 2.3.2.11. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 11.

Câu 12: Bạn có nhận xét thế nào về chất lượng phục vụ của cán bộ thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi ý kiến của họ về chất lượng phục vụ của nhân viên thư viện IUH. Bộ dữ liệu là các câu trả lời ngắn dạng text.
- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60
 - + unique: 34
 - + top: Tốt
 - + freq: 14

- Tìm mối tương quan giữa câu hỏi và kết quả:
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:

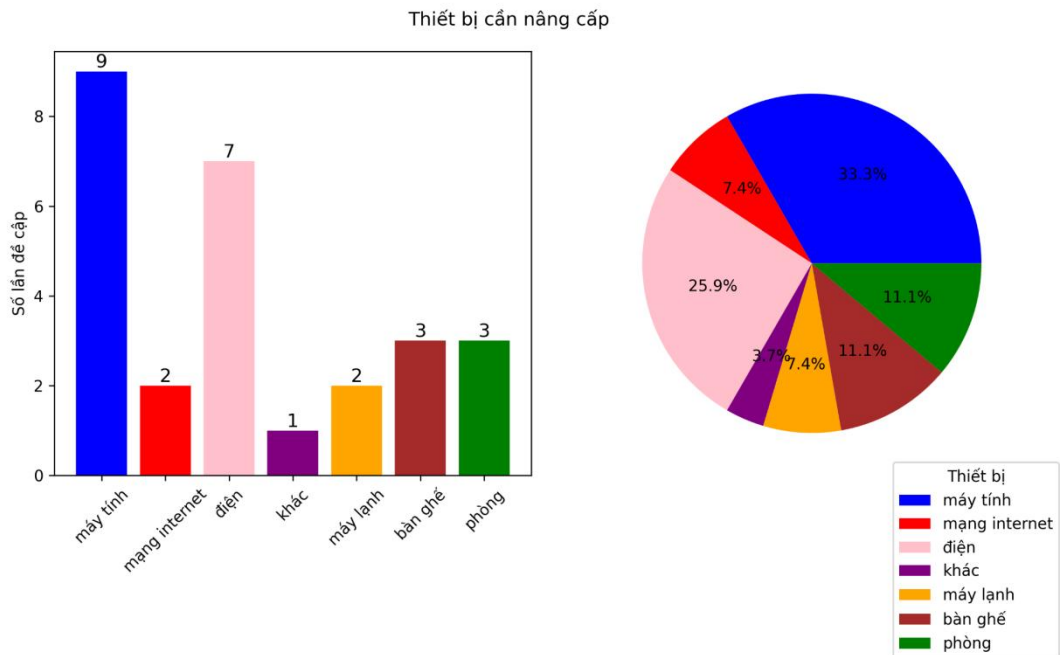


Hình 2.3.2.12. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 12.

Câu 13: Bạn có mong muốn thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh thay thế/nâng cấp thiết bị nào không?

- Mô tả về dữ liệu: đây là câu trả lời của 60 sinh viên thực hiện khảo sát cho câu hỏi về đề xuất của họ về các thiết bị mà thư viện nên thay thế hoặc nâng cấp. Bộ dữ liệu là các câu trả lời ngắn dạng text.
- Nhóm đã thực hiện các thống kê căn bản, kết quả như sau:
 - + count: 60
 - + unique: 34

- + top: Không có
- + freq: 13
- Tìm mối tương quan giữa câu hỏi và kết quả:
- Xác định các yếu tố quan trọng ảnh hưởng đến kết quả:
- Từ bộ dữ liệu và quá trình phân tích, nhóm có biểu đồ trực quan như sau:



Hình 2.3.2.13. Biểu đồ trực quan hóa phân phối dữ liệu của câu hỏi 13.

2.4. Kết luận:

2.4.1. Kết luận và đề xuất giải pháp:

Từ bài tập này, chúng tôi có thêm được kiến thức và kỹ năng để thực hiện một đề án chuyên ngành, từ việc chọn chủ đề, đặt ra câu hỏi nghiên cứu đến giải quyết vấn đề bằng các kỹ thuật xử lý, phân tích dữ liệu bằng Python. Thông qua nhiệm vụ phân tích dữ liệu bằng Python, chúng tôi có thêm nhiều hiểu biết về các kỹ thuật phân tích dữ liệu, nâng cao kỹ năng lập trình đặc biệt về là các kỹ năng thông kê và trực quan hóa dữ liệu.

Bên cạnh đó, bài tập này cũng là cơ hội để các thành viên trong nhóm làm việc với nhau, nâng cao kỹ năng làm việc nhóm từ việc phân chia nhiệm vụ, phối hợp làm việc

đến quản lý tiến trình, kiểm tra kết quả làm việc của nhau. Việc lập ra bảng câu hỏi khảo sát giúp chúng tôi hiểu rõ hơn về quy trình khảo sát từ việc xác định mục tiêu, xác định câu hỏi khảo sát, xây dựng bảng câu hỏi, xác định đối tượng đến việc đi khảo sát.

Cuối cùng, từ việc dùng Python để phân tích, đánh giá và trực quan hóa bộ dữ liệu thu được sau quá trình khảo sát, nhóm có được một số kết luận và giải pháp như sau:

- Kết luận:
 - + Đối với các khía cạnh *Chất lượng các thiết bị và kết nối mạng*: sinh viên hài lòng nhất đối với chất lượng hệ thống máy lạnh (điểm 5 được chọn nhiều nhất, chiếm 43.3%), kết nối wifi (điểm 4 được chọn nhiều nhất, chiếm 40%) và số lượng máy tính của thư viện (điểm 4 được chọn nhiều nhất, chiếm 28.3%). Về vấn đề chất lượng máy tính, bố trí ổ cắm điện thì sinh viên đánh giá ở mức độ bình thường.
 - + Đối với khía cạnh *Không gian thư viện*: hầu hết các câu hỏi đều được sinh viên chọn các mức điểm cao, ngoại trừ vấn đề về số lượng phòng họp nhóm sinh viên đánh giá ở mức độ trung bình (điểm 3 được chọn nhiều nhất, chiếm 31.7%).
 - + Đối với khía cạnh *Chất lượng phục vụ của cán bộ, nhân viên*: đa số sinh viên tỏ thái độ trung tính đối với thái độ/chất lượng phục vụ của cán bộ, nhân viên thư viện (ý kiến thể hiện sự trung tính chiếm 68.3%).
 - + Đối với khía cạnh *Mong muốn của sinh viên*: đối với câu hỏi 11 và 13 phần lớn sinh viên có ý kiến đề góp ý cải thiện/nâng cấp thiết bị, cụ thể, tỉ lệ các câu trả lời có ý kiến của 2 câu hỏi này đều hơn 95% trong 60 câu trả lời nhận được. Sinh viên có mong muốn thư viện cải thiện hệ thống máy tính với số lượng các câu trả lời ở câu hỏi này có chứa cụm từ “máy tính” hay các cụm từ đồng nghĩa với nó chiếm tỉ lệ 25% trong tổng số các câu trả lời có ý kiến. Đồng thời sinh viên cũng có ý kiến mong muốn thư viện nâng cấp/thay thế hệ thống máy tính với số lượng các câu trả lời của câu hỏi này có chứa cụm từ “máy tính” hay các cụm từ đồng nghĩa chiếm tỉ lệ 33.3% trong tổng số các câu trả lời có ý kiến.

- + Nhìn chung, sinh viên có mức độ hài lòng chung tương đối cao với dịch vụ thư viện IUH, cụ thể top 2 điểm số được chọn nhiều nhất trong số 10 câu hỏi dạng Likert là điểm 3 và điểm 4 với số lượng lần lượt là 207 và 165 (ứng với tỉ lệ 34.5% và 27.5%). Nhận xét đối với chất lượng phục vụ của cán bộ, nhân viên thư viện đa số là trung tính. Các yếu tố ảnh hưởng lớn đến độ hài lòng của sinh viên đối với dịch vụ thư viện là các vấn đề về máy tính, hệ thống máy lạnh, bàn ghế và ổ cắm điện.
- Giải pháp: từ kết quả thu về của câu hỏi 11 và 13, đồng thời xem xét kết quả các câu hỏi thuộc *Nhóm I* và *Nhóm II*, nhóm có một số đề xuất sau để thư viện tập trung cải thiện chất lượng dịch vụ:
 - + Nhà trường nên tập trung nâng cấp hệ thống máy tính tại khu vực tự học, có thể xem xét nâng cấp bộ xử lý trung tâm để cải thiện tốc độ phản hồi hoặc xem xét thay thế máy tính mới.
 - + Hệ thống điện, ổ cắm điện cũng đang là vấn đề mà nhiều sinh viên mong muốn được cải thiện. Thư viện có thể xem xét bố trí lại các ổ điện tại các bàn tự học hoặc tăng thêm số lượng ổ cắm tại khu vực này.
 - + Sinh viên có mức độ hài lòng cao đối với hệ thống máy lạnh hiện tại, vì thế thư viện nên chú trọng việc bảo trì và bảo dưỡng cho hệ thống máy lạnh nhằm duy trì chất lượng không khí và nhiệt độ, thông qua đó đảm bảo cho sinh viên có trải nghiệm tốt khi sử dụng thư viện.

2.4.2. Hạn chế:

Sau quá trình thực hiện bài tập này, nhóm nhận ra một số điểm hạn chế cần được khắc phục như sau:

- Cách thiết kế câu hỏi khảo sát chưa thực sự tối ưu, dễ gây hiểu nhầm cho đối tượng khảo sát, nếu điều chỉnh cho tối ưu hơn, dữ liệu thu về sẽ có nhiều giá trị cho quá trình thống kê, phân tích.

- Mẫu đối tượng khảo sát chưa thực sự phổ biến và đa dạng, cần mở rộng thêm mẫu khảo sát và gia tăng thêm khoa và khóa của sinh viên để kết quả mang tính khách quan và phong phú nhất có thể.
- Quá trình phân tích dữ liệu bằng Python còn nhiều khó khăn, thiếu kỹ thuật, cần tối ưu hóa quy trình để có được những phân tích chuyên sâu hơn.

2.4.3. Hướng phát triển trong tương lai:

Trong tương lai, nhóm sẽ hướng tới hoàn thiện kỹ năng của các thành viên trong nhóm cả về kỹ năng chuyên ngành và các kỹ năng liên quan đến việc thực hiện đồ án, cụ thể:

- Hướng tới hoàn chỉnh kỹ năng xây dựng bộ câu hỏi khảo sát; tối ưu các câu hỏi sao cho người thực hiện khảo sát dễ hiểu để học cung cấp những thông tin có giá trị, phục vụ cho quá trình phân tích dữ liệu.
- Mở rộng quy mô khảo sát và đối tượng khảo sát, gia tăng số lượng phiếu khảo sát và đa dạng hóa đối tượng khảo sát để kết quả thu về mang tính khách quan và tổng quát. Tìm hiểu thêm các công cụ khảo sát trực tuyến để gia tăng số lượng đối tượng khảo sát đồng thời giảm thời gian, chi phí.
- Luyện tập, nâng cao kỹ năng lập trình phân tích dữ liệu và các tác vụ liên quan. Học hỏi, trau dồi thêm các kỹ năng phân tích dữ liệu chuyên sâu, tối ưu hóa code và thời gian chạy; học thêm các dạng biểu đồ để trực quan hóa dữ liệu một cách chi tiết và dễ hiểu.

CHƯƠNG 3

KHAI PHÁ DỮ LIỆU TỪ CÂU NÓI CỦA NGƯỜI NỔI TIẾNG

3.1. Thu thập dữ liệu:

Để thu thập dữ liệu về các câu nói của Những người nổi tiếng và thông tin của họ. Chúng tôi đã sử dụng thư viện selenium kết hợp với trình duyệt tự động ChromeDriver để truy cập và thao tác trên trang web: <http://quotes.toscrape.com/>

Việc thu thập dữ liệu được chúng tôi thực hiện qua các bước sau:

- Bước 1: Thu thập dữ liệu và mô tả cấu trúc trang web:
 - + Đầu tiên, chúng tôi cài đặt các thư viện cần thiết như sử dụng selenium để truy cập và thu thập nội dung HTML của trang web và cấu hình ChromeDriver để tương thích với trình duyệt Chrome mà tôi đang sử dụng.

Code:

```
# Khai báo thư viện
from selenium import webdriver
from selenium.webdriver.common.by import By
import time
```

- + Sau khi đã thiết lập xong công cụ, chúng tôi viết code để cào dữ liệu từ trang web để lấy toàn bộ mã HTML và lưu vào một file tên là kq.txt để thuận tiện cho việc phân tích cấu trúc trang.

Code:

```
# Cào dữ liệu từ trang web. lưu kết quả vào file kq.txt
Driver = webdriver.Chrome()
Driver.get("https://quotes.toscrape.com/")
time.sleep(1)
ketqua_html = Driver.page_source
```

```

ketqua_html.split('¥n')
# Lưu vào file
with open('kq.txt', 'w', encoding='utf-8') as file:
    for line in ketqua_html.split('¥n'):
        file.write(line + '¥n')
print('Lưu thành công file kq.txt')

```

Kết quả:

Lưu thành công file kq.txt

Hình 3.1.1. Kết quả của việc lưu file thành công

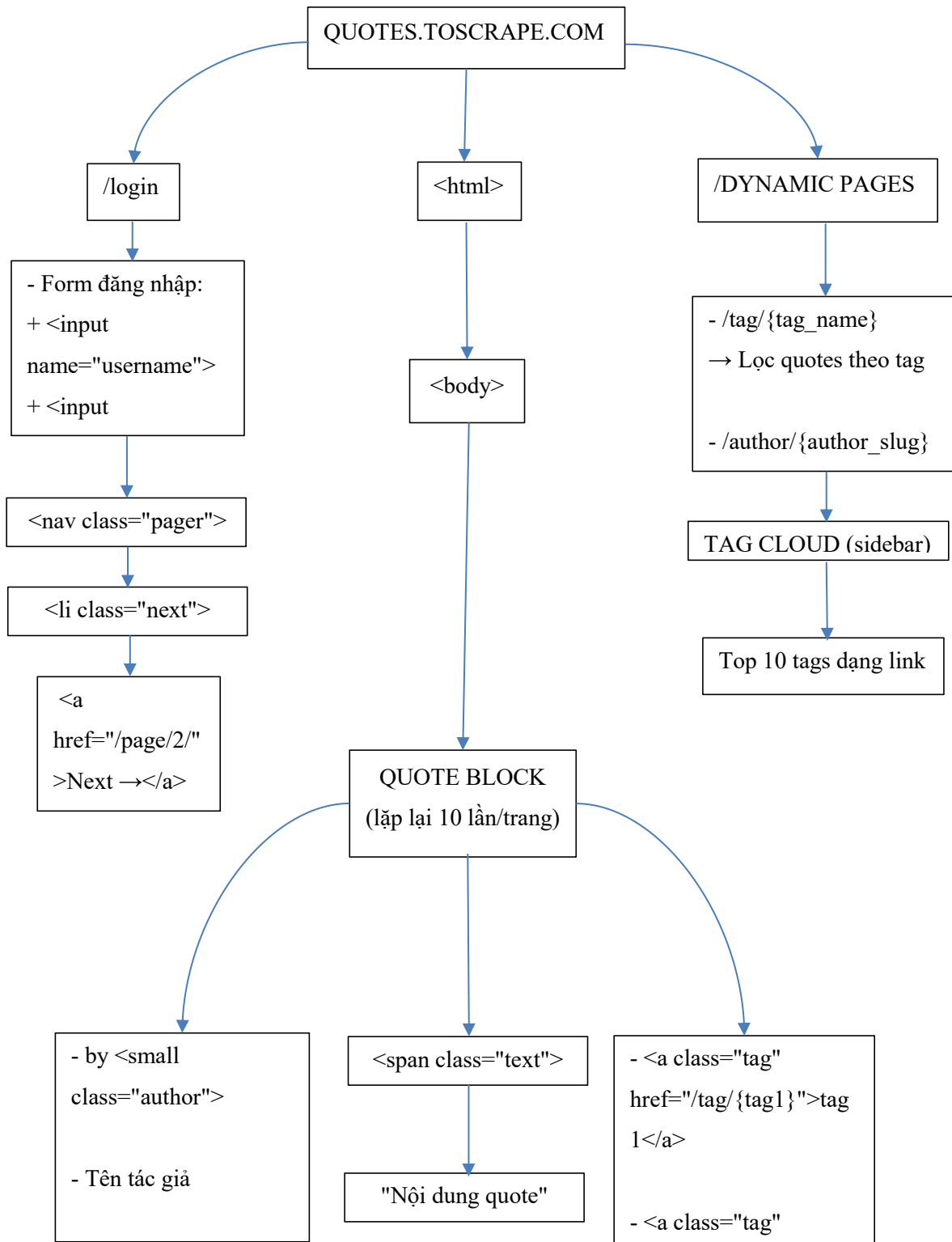
```

1 <html lang="en"><head>
2   <meta charset="UTF-8">
3   <title>Quotes to Scrape</title>
4   <link rel="stylesheet" href="/static/bootstrap.min.css">
5   <link rel="stylesheet" href="/static/main.css">
6
7
8 </head>
9 <body>
10   <div class="container">
11     <div class="row header-box">
12       <div class="col-md-8">
13         <h1>
14           <a href="/" style="text-decoration: none">Quotes to Scrape</a>
15         </h1>
16       </div>
17       <div class="col-md-4">
18         <p>
19
20           <a href="/login">Login</a>
21
22         </p>
23       </div>
24     </div>

```

Hình 3.1.2. Dữ liệu trong file kq.txt

- + Sau khi xem qua mã nguồn HTML chúng tôi nhận thấy mỗi câu nói trên trang web được đặt trong một thẻ <div> với class “quote”. Bên trong mỗi “quote” là các thông tin gồm nội dung câu nói, tên tác giả, và một đường dẫn dẫn đến trang riêng của tác giả.
- + Mô tả cấu trúc trang Web bằng sơ đồ sau:



Hình 3.1.3. Sơ đồ mô tả về Cấu trúc trang Web: <http://quotes.toscrape.com/>.

- Bước 2: Xử lý dữ liệu đã thu thập

Sau khi có dữ liệu từ trang web, chúng tôi tiến hành viết code để xử lý và trích xuất thông tin cần thiết để đáp ứng các yêu cầu đã đề ra:

- + Trích xuất các câu nói nổi tiếng: Tôi sử dụng Selenium để tìm tất cả các phần tử div mang class “quote” và lưu lại toàn bộ HTML của chúng vào biến result. Sau đó in kết quả ra màn hình để kiểm tra trực tiếp.

Code:

#a) Đọc tất cả các thẻ html(div) với lớp là 'quote' và lưu nó trong biến result, hiển thị giá trị result ra màn hình/

```
quotes = Driver.find_elements(By.CLASS_NAME, 'quote')
```

#lưu mã HTML của mỗi thẻ vào biến 'result'

```
result = [quote.get_attribute('outerHTML') for quote in quotes]
```

#Hiển thị giá trị của biến result ra màn hình

for i in result:

```
    print(i)
```

Kết quả:

```
<div class="quote" itemscope="" itemtype="http://schema.org/CreativeWork">
  <span class="text" itemprop="text">"The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking."</span>
  <span>by <small class="author" itemprop="author">Albert Einstein</small>
  <a href="/author/Albert-Einstein">(about)</a>
</span>
<div class="tags">
  Tags:
  <meta class="keywords" itemprop="keywords" content="change,deep-thoughts,thinking,world">

  <a class="tag" href="/tag/change/page/1/">change</a>

  <a class="tag" href="/tag/deep-thoughts/page/1/">deep-thoughts</a>

  <a class="tag" href="/tag/thinking/page/1/">thinking</a>

  <a class="tag" href="/tag/world/page/1/">world</a>

</div>
```

Hình 3.1.4. Kết quả đọc tất cả các thẻ html(div) với lớp là ‘quote’.

- + Lọc ra tên tác giả: Từ dữ liệu trong result, tôi tiếp tục tìm các phần tử chứa tag <small> với class “author”, đây chính là nơi lưu tên tác giả cho từng câu nói và in toàn bộ danh sách tác giả ra màn hình.

Code:

b) Tìm biến result vừa rồi các dữ liệu có chứa nhãn small với class là author và in kết quả ra màn hình

```
from bs4 import BeautifulSoup

for quote_html in result:
    soup = BeautifulSoup(quote_html, 'html.parser') # Thay 'lxml' bằng
    'html.parser' để không cần cài thêm thư viện
    author_tag = soup.find('small', class_='author')

    # Sử dụng f-string để format output đẹp hơn
    print(f'Tác giả: {author_tag.text if author_tag else 'Không tìm thấy tác giả'}')

# Thêm dấu phân cách giữa các quote
    print("-----")
```

Kết quả:

```
Tác giả: Albert Einstein
-----
Tác giả: J.K. Rowling
-----
Tác giả: Albert Einstein
-----
Tác giả: Jane Austen
-----
Tác giả: Marilyn Monroe
-----
Tác giả: Albert Einstein
-----
Tác giả: André Gide
-----
Tác giả: Thomas A. Edison
-----
Tác giả: Eleanor Roosevelt
-----
Tác giả: Steve Martin
```

Hình 3.1.5. Kết quả danh sách các tác giả có chứa nhãn small.

- + Viết hàm in ra toàn bộ thông tin về tác giả: Tôi viết một hàm riêng tên là `tacgiaLink()` trích xuất các thông tin gồm:
 - Tên tác giả
 - Đường link cá nhân
 - Ngày tháng năm sinh
 - Một câu nói tiêu biểu của họ
- + Lưu trữ dữ liệu: Sau khi thu thập được thông tin chi tiết của từng tác giả và câu nói, sau đó lưu toàn bộ kết quả vào một file CSV tên là `Quote.csv`, với định dạng mỗi dòng trong file là một tác giả cùng với câu nói tương ứng đảm bảo chương trình sẽ tự động lặp lại ít nhất 40 lần, như yêu cầu của đề bài.

Code:

#c) Viết Hàm `tacgiaLink()` để lấy nội dung của mỗi tác giả. Mỗi Tác giả in các nội dung

Tên tác giả, Đường link của tác giả, Ngày tháng năm sinh, Và câu nói nổi tiếng

```
from selenium import webdriver
from selenium.webdriver.common.by import By
import time
import csv
from selenium.common.exceptions import NoSuchElementException
```

```
def tacgiaLink():
```

```
    # Khởi tạo trình duyệt Chrome
```

```
    driver = webdriver.Chrome()
```

```
    driver.get('https://quotes.toscrape.com/')

```

```
    time.sleep(2) # Chờ 2 giây để trang web load hoàn tất
```

```

collected_data = [] # Danh sách lưu dữ liệu thu thập

try:
    # Lặp cho đến khi thu thập đủ 40 mục dữ liệu
    while len(collected_data) < 40:
        # Tìm tất cả các quote trên trang hiện tại
        quotes = driver.find_elements(By.CSS_SELECTOR, 'div.quote')

        for quote in quotes:
            try:
                # Trích xuất thông tin từ quote
                author_name = quote.find_element(By.CSS_SELECTOR,
'small.author').text
                author_link = quote.find_element(By.CSS_SELECTOR,
'a[href*="/author/"]').get_attribute('href')
                text_quote = quote.find_element(By.CSS_SELECTOR,
'span.text').text

                # Truy cập trang chi tiết tác giả
                driver.get(author_link)
                time.sleep(2)

                # Trích xuất thông tin ngày sinh và nơi sinh
                birth_date = driver.find_element(By.CSS_SELECTOR, '.author-
born-date').text

```

```

# Hiển thị thông tin ra màn hình
print("¥n" + "-----")
print(f'Tên tác giả: {author_name}')
print(f'Đường link của tác giả: {author_link}')
print(f'Ngày sinh: {birth_date}')
print(f'Câu nói nổi tiếng: {text_quote}')
print("-----" + "¥n")

# Thêm dữ liệu vào danh sách
collected_data.append([author_name, author_link, birth_date,
text_quote])

# Quay lại trang danh sách quotes
driver.back()
time.sleep(2)

# Dừng nếu đã thu thập đủ 40 mục
if len(collected_data) >= 40:
    break

except NoSuchElementException as e:
    print(f'Lỗi không tìm thấy phần tử: {e}')
    continue

# Chuyển trang nếu chưa thu thập đủ dữ liệu
if len(collected_data) < 40:
    try:

```

```

        next_button = driver.find_element(By.CSS_SELECTOR, 'li.next >
a')

        next_button.click()
        time.sleep(2)
    except NoSuchElementException:
        print("Không tìm thấy nút chuyển trang")
        break

#d) Lưu kết quả vào file Quote.csv
# Lưu dữ liệu vào file CSV
with open('Quote.csv', 'w', newline="", encoding='utf-8-sig') as csvfile:
    writer = csv.writer(csvfile)
    writer.writerow(['Tác giả', 'Link', 'Ngày sinh', 'Câu nói'])
    writer.writerows(collected_data)

except Exception as e:
    print(f"Có lỗi xảy ra trong quá trình thu thập: {e}")
finally:
    driver.quit()
    print("Đã hoàn thành việc thu thập dữ liệu!")

# Gọi hàm thực thi
tacgiaLink()

```

Kết quả:

 Tên tác giả: Albert Einstein
 Đường link của tác giả: <https://quotes.toscrape.com/author/Albert-Einstein>
 Ngày sinh: March 14, 1879
 Câu nói nổi tiếng: "The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking."

 Tên tác giả: J.K. Rowling
 Đường link của tác giả: <https://quotes.toscrape.com/author/J-K-Rowling>
 Ngày sinh: July 31, 1965
 Câu nói nổi tiếng: "It is our choices, Harry, that show what we truly are, far more than our abilities."

Hình 3.1.6. Kết quả danh sách thông tin các tác giả.

Kết quả dữ liệu thu thập được:

- Sau khi thu thập dữ liệu thành công, chúng tôi có được một tập dữ liệu chứa đựng những câu nói truyền cảm hứng cùng với thông tin của tác giả và đường liên kết tương ứng.
- Dữ liệu này bao gồm thông tin về các câu nói ý nghĩa và sâu sắc từ những tác giả nổi tiếng và đường liên kết đến trang cá nhân của họ trên trang web. Điều này giúp người đọc hiểu rõ hơn về nguồn gốc cũng như triết lý quan điểm của tác giả thông qua các câu nói nổi tiếng của họ.

3.2. Khai phá dữ liệu:

3.2.1. Xử lý dữ liệu - Data Imputation:

Yêu cầu 1: Bạn hãy thêm vào Trường STT và điền tự động dữ liệu của trường này?

- Sau khi hoàn thành bước thu thập dữ liệu và lưu vào file Quote.csv, để thêm vào Trường STT (số thứ tự) vào đầu bảng bằng cách sử dụng insert(). Cột STT thêm vào được tự động đánh số thông qua vòng for() bắt đầu từ 0. Nhờ đó giúp việc theo dõi, đối chiếu và xử lý dễ dàng và thuận tiện hơn.
- Sau đó tôi lưu dữ liệu đã chỉnh sửa vào file mới QuoteInsert_1.csv

Code:

- Bạn hãy thêm vào Trường STT và điền tự động dữ liệu của trường này?

```
df.insert(0, 'STT', range(1, 1 + len(df)))
```

```
df.head(30)
```

```
# Lưu vào một file mới
df.to_csv('QuoteInsert_1.csv', index=False)
print('Lưu file QuoteInsert_1.csv chứa cột STT')
```

Kết quả:

	STT	Tác giả	Link	Ngày sinh	Câu nói
0	1	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	March 14, 1879	"The world as we have created it is a process ...
1	2	J.K. Rowling	https://quotes.toscrape.com/author/J-K-Rowling	July 31, 1965	"It is our choices, Harry, that show what we t...
2	3	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	March 14, 1879	"There are only two ways to live your life. On...
3	4	Jane Austen	https://quotes.toscrape.com/author/Jane-Austen	December 16, 1775	"The person, be it gentleman or lady, who has ...
4	5	Marilyn Monroe	https://quotes.toscrape.com/author/Marilyn-Monroe	June 01, 1926	"Imperfection is beauty, madness is genius and...

Lưu file QuoteInsert_1.csv chứa cột STT

Hình 3.2.1.1. Kết quả sau khi thêm cột STT và lưu vào file mới.

Yêu cầu 2: Một số giá trị của dữ liệu Trường ngày sinh chưa có, bạn hãy đề xuất cách điền?

- Với file QuoteInsert_1.csv (file sau khi đã thêm cột STT). Một số giá trị của Trường ngày sinh chưa có, nên chúng tôi có thực hiện một số phương án để điền vào ô giá trị đó như sau:
 - + Phương án 1: Tìm kiếm nếu như có thêm thông tin như năm mất hoặc tuổi lúc qua đời thì có thể suy ngược lại ngày sinh. Từ các nguồn công khai như Wikipedia, FamousBirthdays hoặc các trang lưu trữ tiểu sử tác giả.
 - + Phương án 2: Sử dụng ngày sinh trung bình của các tác giả khác, chúng tôi sẽ tính ngày sinh trung bình của những tác giả đã có đầy đủ thông tin về ngày mất và tính giá trị trung bình để điền vào các dữ liệu còn thiếu.
 - + Phương án 3: Nếu như không tìm được hoặc thông tin không chắc chắn, chúng tôi quyết định đánh dấu lại là “Không rõ” để tránh điền dữ liệu suy đoán có thể

gây sai lệch kết quả phân tích và thay các giá trị “Không rõ” bằng None trong code.

- Trong ba phương án trên thì chúng tôi đề xuất hai phương án là kết hợp phương án 1 và phương án 3 vì có thể thu thập được các thông tin từ các người nổi tiếng đã public và cập nhật liên tục trên các nền tảng công khai nên có thể dễ dàng lấy thông tin và đối với các thông tin chưa rõ hay chưa được cập nhật thì sẽ thực hiện theo phương án 3.

Code:

- Một số giá trị của dữ liệu Trường ngày sinh chưa có, bạn hãy đề xuất cách điền?

Đọc File QuoteInsert_1.csv

df = pd.read_csv('QuoteInsert_1.csv')

List danh sách năm mất

Nammat_tacgia = {

'Albert Einstein' : 1955,

'J.K. Rowling' : None,

'Jane Austen' : 1817,

'Marilyn Monroe' : 1962,

'André Gide' : 1951,

'Thomas A. Edison' : 1931,

'Eleanor Roosevelt' : 1962,

'Steve Martin' : None,


```

'Bob Marley' : 1981,
'Dr. Seuss' : 1991,
'Douglas Adams' : 2001,
'Elie Wiesel' : 2016,
'Friedrich Nietzsche' : 1900,
'Mark Twain' : 1910,
'Allen Saunders' : 1986,
'Pablo Neruda' : 1973,
'Ralph Waldo Emerson' : 1882,
'Mother Teresa' : 1997,
'Garrison Keillor' : None,
'Jim Henson' : 1990,
'Charles M. Schulz' : 2000,
'William Nicholson' : None,
'Jorge Luis Borges' : 1986,
'George Eliot' : 1880

}

# thêm cột năm mất vào dataframe
df['Năm mất'] = df['Tác giả'].apply(lambda x: Nammat_tacgia[x])

```

df

Kết quả:

STT	Tác giả		Link	Ngày sinh	Câu nói	Năm mất
0	1	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	March 14, 1879	"The world as we have created it is a process ...	1955.0
1	2	J.K. Rowling	https://quotes.toscrape.com/author/J-K-Rowling	July 31, 1965	"It is our choices, Harry, that show what we t...	NaN
2	3	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	March 14, 1879	"There are only two ways to live your life. On...	1955.0
3	4	Jane Austen	https://quotes.toscrape.com/author/Jane-Austen	December 16, 1775	"The person, be it gentleman or lady, who has ...	1817.0
4	5	Marilyn Monroe	https://quotes.toscrape.com/author/Marilyn-Monroe	June 01, 1926	"Imperfection is beauty, madness is genius and...	1962.0
5	6	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	March 14, 1879	"Try not to become a man of success. Rather be...	1955.0
6	7	André Gide	https://quotes.toscrape.com/author/Andre-Gide	November 22, 1869	"It is better to be hated for what you are tha...	1951.0
7	8	Thomas A. Edison	https://quotes.toscrape.com/author/Thomas-A-Ed...	February 11, 1847	"I have not failed. I've just found 10,000 way...	1931.0
8	9	Eleanor Roosevelt	https://quotes.toscrape.com/author/Eleanor-Roo...	October 11, 1884	"A woman is like a tea bag; you never know how...	1962.0
9	10	Steve Martin	https://quotes.toscrape.com/author/Steve-Martin	August 14, 1945	"A day without sunshine is like, you know, nig...	NaN

Hình 3.2.1.2. Kết quả bộ dữ liệu sau khi thêm cột Năm mất.

Yêu cầu 3: Bạn hãy thêm vào Trường Tuổi (Tuổi) và đề xuất cách điền tuổi của các tác giả?

- Sau khi thêm cột “Năm mất” chúng tôi tiến hành tính tuổi (Tuổi) cho mỗi tác giả. Đầu tiên, chúng tôi bổ sung thêm cột “Tuổi”. Để tính tuổi bằng cách viết hàm `TinhTuoi()` dựa trên nguyên tắc:
 - + Nếu tác giả đã qua đời, thì tuổi = năm mất – năm sinh.
 - + Nếu tác giả còn sống hoặc chưa rõ dữ liệu năm mất, thì tuổi sẽ cập nhật để giá trị (NaN) để tránh suy đoán sai.
- Kết quả, chúng tôi sẽ thu được một bộ dữ liệu hoàn chỉnh (lưu vào file `QuoteInsert_Tuoi.csv`) để thực hiện tiếp tục cho việc thống kê và phân tích. Thông qua cách thức điền giá trị như trên sẽ giúp chúng ta giảm thiểu tình trạng dữ liệu sai và không chính xác.

Code:

- Bạn hãy thêm vào Trường Tuổi (Tuổi) và đề xuất cách điền tuổi của các tác giả?

```
def TinhTuoi(SinhNhat, Nammat):
```

```
    try:
```

```
        if Nammat is None:
```

```
            return 2025 - SinhNhat.year
```

```
        else:
```

```
            return Nammat - SinhNhat.year
```

```
    except AttributeError:
```

```
        return None
```

```
df['Ngày sinh'] = pd.to_datetime(df['Ngày sinh'], errors='coerce')
```

```
df['Tuổi'] = df.apply(lambda x: TinhTuoi(x['Ngày sinh'], x['Năm mất']), axis=1)
```

```
# save file
```

```
df.to_csv('QuoteInsert_Tuoi.csv', index=False)
```

```
df
```

Kết quả:

	STT	Tác giả	Link	Ngày sinh	Câu nói	Năm mất	Tuổi
0	1	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	1879-03-14	"The world as we have created it is a process ...	1955.0	76.0
1	2	J.K. Rowling	https://quotes.toscrape.com/author/J-K-Rowling	1965-07-31	"It is our choices, Harry, that show what we t...	NaN	NaN
2	3	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	1879-03-14	"There are only two ways to live your life. On...	1955.0	76.0
3	4	Jane Austen	https://quotes.toscrape.com/author/Jane-Austen	1775-12-16	"The person, be it gentleman or lady, who has ...	1817.0	42.0
4	5	Marilyn Monroe	https://quotes.toscrape.com/author/Marilyn-Monroe	1926-06-01	"Imperfection is beauty, madness is genius and...	1962.0	36.0
5	6	Albert Einstein	https://quotes.toscrape.com/author/Albert-Eins...	1879-03-14	"Try not to become a man of success. Rather be...	1955.0	76.0
6	7	André Gide	https://quotes.toscrape.com/author/Andre-Gide	1869-11-22	"It is better to be hated for what you are tha...	1951.0	82.0
7	8	Thomas A. Edison	https://quotes.toscrape.com/author/Thomas-A-Ed...	1847-02-11	"I have not failed. I've just found 10,000 way...	1931.0	84.0
8	9	Eleanor Roosevelt	https://quotes.toscrape.com/author/Eleanor-Roo...	1884-10-11	"A woman is like a tea bag; you never know how...	1962.0	78.0
9	10	Steve Martin	https://quotes.toscrape.com/author/Steve-Martin	1945-08-14	"A day without sunshine is like, you know, nig...	NaN	NaN

Hình 3.2.1.3. Kết quả bộ dữ liệu sau khi thêm Tuổi.

3.2.2. Khám phá dữ liệu - Data Exploration:

Sau khi đã hoàn thành việc xử lý dữ liệu thu thập từ trang web chúng tôi có một bộ dữ liệu hoàn chỉnh. Tiếp theo chúng tôi tiến hành khám phá dữ liệu thông qua các yêu cầu sau.

Yêu cầu 1: Thống kê về tác giả và câu nói nổi tiếng có trong bộ dữ liệu

- Chúng tôi sẽ tiến hành đọc file (QuoteInsert_Tuoi.csv) và thực hiện thống kê số lượng tác giả cũng như các câu nói nổi tiếng trong tập dữ liệu.
- Chúng tôi sử dụng thư viện Pandas để xử lý dữ liệu và hàm `value_counts()` để thống kê các tác giả và câu nói nổi tiếng. Sau đó in ra kết quả các thống kê.
- Kết quả sau khi thống kê là số lượng các câu nói nổi tiếng và số lượng tác giả và những câu nói phổ biến nhất.

Code:

#- Thống kê về tác giả và câu nói nổi tiếng có trong bộ dữ liệu

try:

```
du_lieu = pd.read_csv('QuoteInsert_Tuoi.csv')
```

```
cot_du_lieu = du_lieu.columns.tolist()
```

except Exception as loi:

```
cot_du_lieu = str(lois)
```

```

if isinstance(cot_du_lieu, list):
    # Tính số lượng tác giả và số câu nói nổi tiếng của từng tác giả
    thong_ke_tac_gia = du_lieu['Tác giả'].value_counts()
    thong_ke_cau_noi = du_lieu['Câu nói'].value_counts()

    # Tạo dictionary lưu trữ thông tin thống kê
    tong_quan_thong_ke = {
        'Số tác giả khác nhau': len(thong_ke_tac_gia),
        'Số câu nói khác nhau': len(thong_ke_cau_noi),
        'Tác giả nổi bật': thong_ke_tac_gia.head().to_dict(),
        'Câu nói nổi bật': thong_ke_cau_noi.head().to_dict()
    }
else:
    # Nếu có lỗi, biến cot_du_lieu sẽ chứa thông báo lỗi
    tong_quan_thong_ke = cot_du_lieu

tong_quan_thong_ke

```

Kết quả:

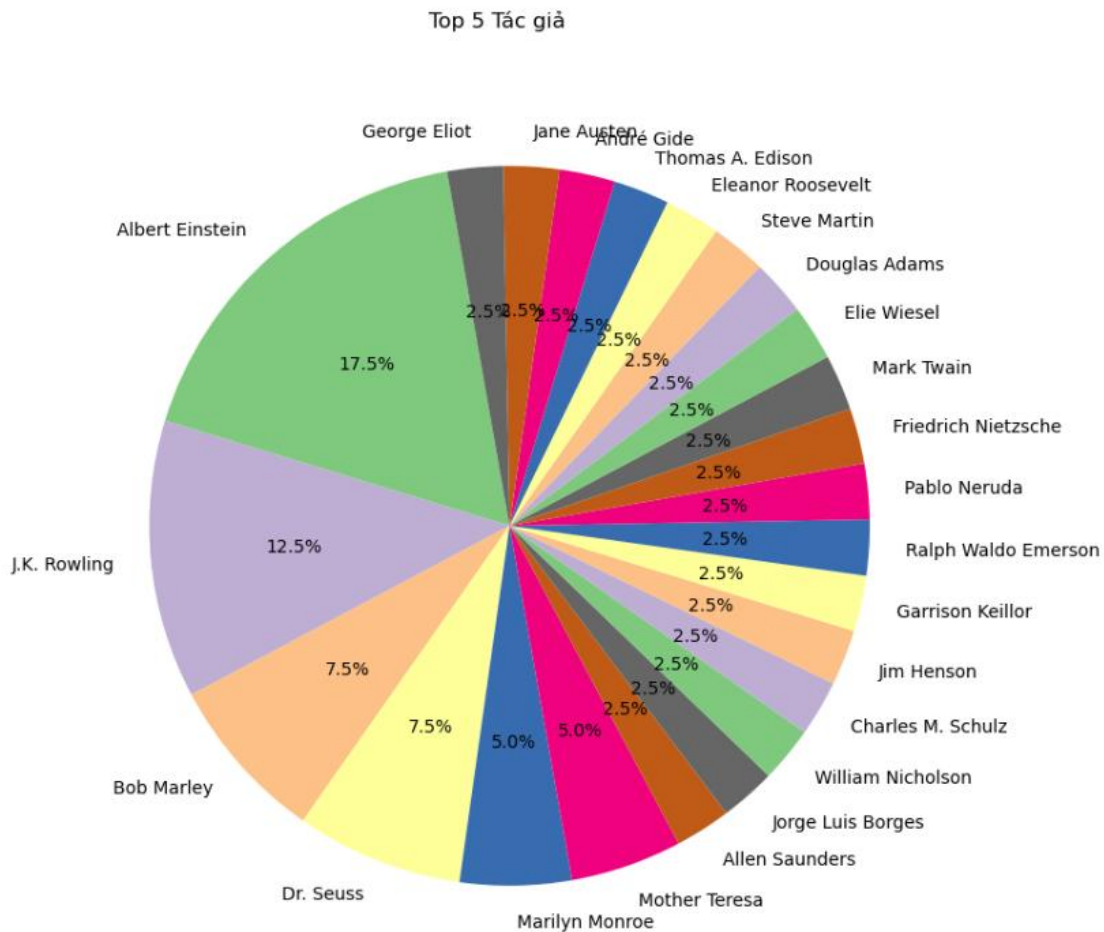
```

{'Số tác giả khác nhau': 24,
 'Số câu nói khác nhau': 40,
 'Tác giả nổi bật': {'Albert Einstein': 7,
 'J.K. Rowling': 5,
 'Bob Marley': 3,
 'Dr. Seuss': 3,
 'Marilyn Monroe': 2},
 'Câu nói nổi bật': {'"The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking."': 1,
 'It is our choices, Harry, that show what we truly are, far more than our abilities.'': 1,
 'If you judge people, you have no time to love them.'': 1,
 'Anyone who thinks sitting in church can make you a Christian must also think that sitting in a garage can make you a car.'': 1,
 'Beauty is in the eye of the beholder and it may be necessary from time to time to give a stupid or misinformed beholder a black eye.'': 1}}

```

Hình 3.2.2.1. Kết quả thống kê về tác giả và câu nói nổi tiếng có trong bộ dữ liệu.

- Chúng tôi vẽ biểu đồ pie chart để thống kê Top 5 Tác giả theo số lượng câu nói nổi tiếng. Tính theo phần trăm.



Hình 3.2.2.2. Biểu đồ thể hiện thống kê các tác giả và câu nói nổi tiếng theo phần trăm.

- Dựa trên biểu đồ, có thể nhận thấy rằng các tác giả như Albert Einstein, J.K. Rowling, Bob Marley, Dr. Seuss và Marilyn Monroe có số lượng câu nói nổi bật cao hơn đáng kể so với các tác giả khác. Điều này cho thấy họ là những nhân vật có nhiều phát ngôn được trích dẫn và ghi nhớ rộng rãi.

Yêu cầu 2: Thống kê về năm sinh và độ tuổi của các tác giả

- Nhằm hạn chế các lỗi tiềm ẩn trong quá trình xử lý dữ liệu do thiếu thông tin, nhóm chúng tôi đã kiểm tra sự tồn tại của hai cột “*Năm sinh*”(năm sinh) và “*Tuổi*” (tuổi) trong tập dữ liệu.

- Nếu cả hai trường dữ liệu này đều đầy đủ thì chúng tôi sẽ thực hiện bước tiếp theo là thống kê năm sinh của các tác giả. Để thực hiện yêu cầu này chúng tôi sử dụng phương thức `value_counts().sort_index()` nhằm đếm số lượng tác giả ứng với từng năm sinh, đồng thời sắp xếp kết quả theo thứ tự thời gian để dễ dàng quan sát phân bố.
- Tương tự, quá trình thống kê cũng được thực hiện cho cột “*Tuổi*”, với kết quả được sắp xếp theo giá trị độ tuổi để phản ánh rõ hơn về cấu trúc tuổi của các tác giả trong tập dữ liệu.

Code:

#- Thống kê về năm sinh và độ tuổi của các tác giả

*# Kiểm tra xem hai cột 'Năm sinh' và 'Tuổi' có tồn tại trong DataFrame không
if 'Ngày sinh' in du_lieu.columns and 'Tuổi' in du_lieu.columns:*

Thống kê số lượng tác giả theo năm sinh và tuổi

thong_ke_nam_sinh = du_lieu['Ngày sinh'].value_counts().sort_index()

thong_ke_tuoi = du_lieu['Tuổi'].value_counts().sort_index()

In ra thống kê

print(" Thống kê Năm sinh:")

print(thong_ke_nam_sinh)

print("\n Thống kê Tuổi:")

print(thong_ke_tuoi)

Biểu đồ ngang cho thống kê năm sinh

plt.figure(figsize=(10, 7))

thong_ke_nam_sinh.plot(kind='barh', color='skyblue')

plt.title('Thống kê Năm sinh')

```

plt.xlabel('Số lượng tác giả')
plt.ylabel('Năm sinh')
plt.gca().invert_yaxis() # Hiển thị năm sinh mới nhất ở trên
plt.tight_layout()
plt.show()

# Biểu đồ cột cho thống kê tuổi
plt.figure(figsize=(10, 5))
thong_ke_tuoi.plot(kind='bar', color='lightgreen')
plt.title('Thống kê Tuổi')
plt.xlabel('Tuổi')
plt.ylabel('Số lượng tác giả')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
else:
    print(" Không tìm thấy cột 'Năm sinh' hoặc 'Tuổi' trong tập dữ liệu.")

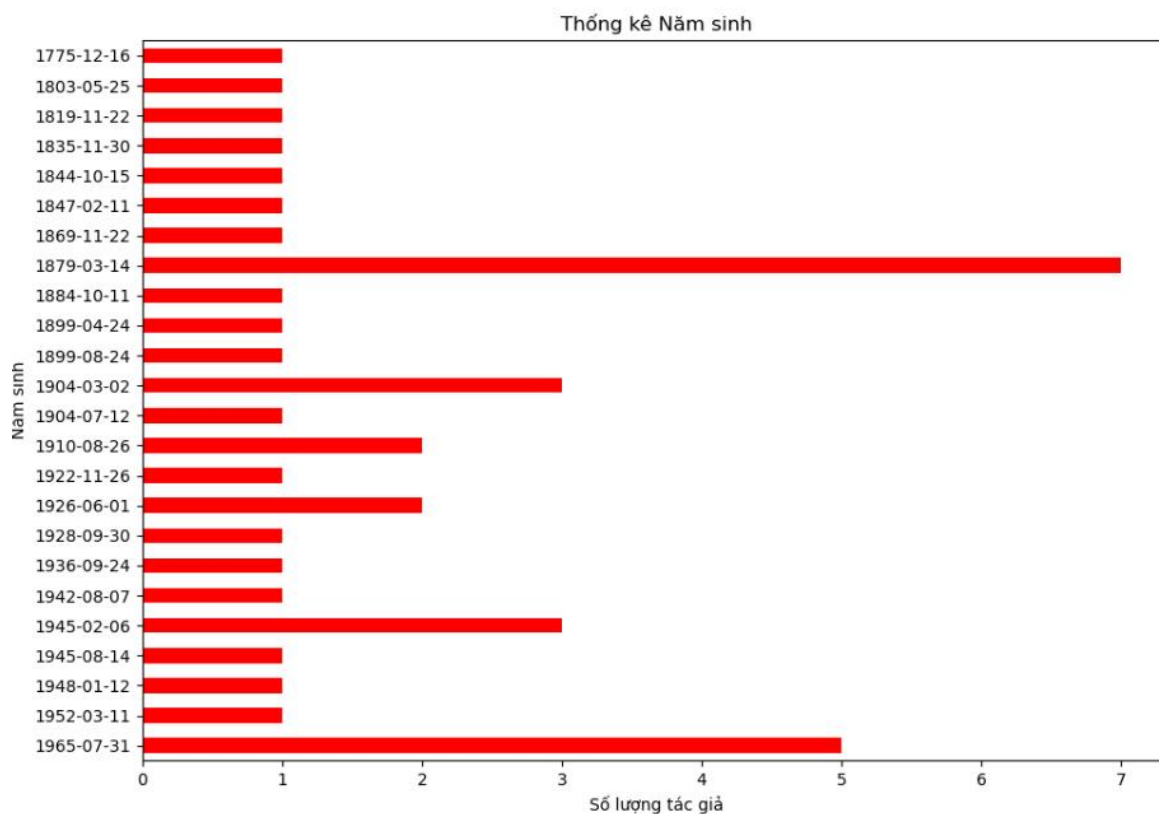
```

Kết quả:

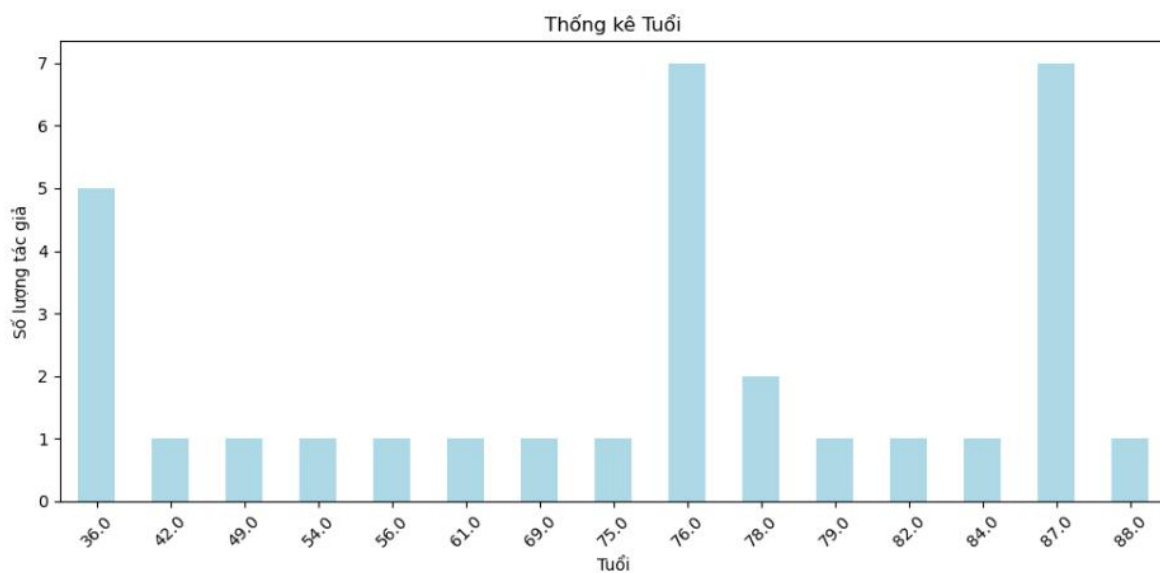
Thống kê Năm sinh:		Thống kê Tuổi:	
Ngày sinh		Tuổi	
1775-12-16	1	36.0	5
1803-05-25	1	42.0	1
1819-11-22	1	49.0	1
1835-11-30	1	54.0	1
1844-10-15	1	56.0	1
1847-02-11	1	61.0	1
1869-11-22	1	69.0	1
1879-03-14	7	75.0	1
1884-10-11	1	76.0	7
1899-04-24	1	78.0	2
1899-08-24	1	79.0	1
1904-03-02	3	82.0	1
1904-07-12	1	84.0	1
1910-08-26	2	87.0	7
1922-11-26	1	88.0	1
1926-06-01	2		
1928-09-30	1		
1936-09-24	1		
1942-08-07	1		
1945-02-06	3		
1945-08-14	1		
1948-01-12	1		
1952-03-11	1		
1965-07-31	5		

Name: count, dtype: int64 Name: count, dtype: int64

Hình 3.2.2.3. Kết quả Thống kê về năm sinh và độ tuổi của các tác giả.



Hình 3.2.2.4. Biểu đồ thống kê năm sinh của các tác giả.



Hình 3.2.2.5. Biểu đồ thống kê độ tuổi của các tác giả.

Yêu cầu 3: Thống kê về các câu nói nổi tiếng như: câu dài nhất, ngắn nhất, số từ, ...

- Để tiến hành phân tích độ dài của các câu trích dẫn trong bộ dữ liệu, nhóm đã thêm một cột mới vào DataFrame với tên là “Độ dài câu”. Cột này được tạo bằng cách áp dụng một hàm lambda lên từng dòng trong cột “Câu nói”, trong đó mỗi câu sẽ được tách thành các từ bằng phương thức split() để đếm số lượng từ.
- Sau khi có được độ dài của tất cả các câu, nhóm sử dụng hai hàm idxmax() và idxmin() để xác định vị trí của câu dài nhất và ngắn nhất trong toàn bộ tập dữ liệu.

Code:

#- Thống kê về các câu nói nổi tiếng như: câu dài nhất, ngắn nhất, số từ

Thêm cột 'Độ dài câu' tính theo số từ trong câu nói

du_lieu['Độ dài câu'] = du_lieu['Câu nói'].apply(lambda x: len(x.split()))

Tìm câu nói dài nhất và ngắn nhất

cau_dai_nhat = du_lieu.loc[du_lieu['Độ dài câu'].idxmax()]

cau_ngan_nhat = du_lieu.loc[du_lieu['Độ dài câu'].idxmin()]

In kết quả

print("\n Câu nói dài nhất:")

print(cau_dai_nhat['Câu nói'])

print("Số từ:", cau_dai_nhat['Độ dài câu'])

print("\n Câu nói ngắn nhất:")

print(cau_ngan_nhat['Câu nói'])

print("Số từ:", cau_ngan_nhat['Độ dài câu'])

Kết quả:

Câu nói dài nhất:

"This life is what you make it. No matter what, you're going to mess up sometimes, it's a universal truth. But the good part is you get to decide how you're going to mess it up. Girls will be your friends - they'll act like it anyway. But just remember, some come, some go. The ones that stay with you through everything - they're your true best friends. Don't let go of them. Also remember, sisters make the best friends in the world. As for lovers, well, they'll come and go too. And baby, I hate to say it, most of them - actually pretty much all of them are going to break your heart, but you can't give up because if you give up, you'll never find your soulmate. You'll never find that half who makes you whole and that goes for everything. Just because you fail once, doesn't mean you're gonna fail at everything. Keep trying, hold on, and always, always, always believe in yourself, because if you don't, then who will, sweetie? So keep your head high, keep your chin up, and most importantly, keep smiling, because life's a beautiful thing and there's so much to smile about."

Số từ: 281

Câu nói ngắn nhất:

"We read to know we're not alone."

Số từ: 7

Hình 3.2.2.6. Kết quả thống kê về các câu nói nổi tiếng: câu dài nhất, ngắn nhất, số từ.

- Ngoài ra, phương thức `describe()` cũng được sử dụng để tạo bảng thống kê tổng quan về độ dài của các câu, bao gồm giá trị trung bình, độ lệch chuẩn, giá trị nhỏ nhất, lớn nhất, phần tư trên/dưới,...

Code:

```
#Thống kê tổng quan về số từ
print("\nTổng quan về số từ trong các câu nói:")
print(du_lieu['Độ dài câu'].describe())
```

Kết quả:

```
Tổng quan về số từ trong các câu nói:
count    40.000000
mean     26.225000
std      35.460695
min       7.000000
25%      12.750000
50%      16.500000
75%      22.250000
max      201.000000
Name: Độ dài câu, dtype: float64
```

Hình 3.2.2.7. Kết quả thống kê tổng quan về số từ trong các câu nói.

- Qua kết quả thu được, chúng tôi có cái nhìn tổng thể về sự phân bố chiều dài của các câu nói nổi tiếng, từ đó có thể rút ra đặc điểm về cách hành văn, phong cách ngôn ngữ của các tác giả.
- Có thể nhận thấy rằng độ dài của câu nói dao động từ ngắn gọn đến rất dài, phản ánh sự đa dạng trong cách biểu đạt. Một số câu tuy ngắn nhưng lại hàm chứa ý

nghĩa sâu sắc, trong khi những câu dài hơn thường mang tính triết lý hoặc mô tả phong phú hơn về nội dung tác giả muốn truyền tải.

Yêu cầu 4: Thống kê về các từ được sử dụng trong các câu nói

- Để thực hiện yêu cầu, chúng tôi tiến hành thống kê các từ được sử dụng thường xuyên trong những câu nói nổi tiếng bằng cách xử lý văn bản thô và trực tiếp phân tích tần suất.
- Đầu tiên, chúng tôi gộp toàn bộ nội dung từ cột “Câu nói” để tạo thành một chuỗi văn bản lớn và chuyển sang chữ thường để chuẩn hóa dữ liệu.
- Sau đó, nhóm sử dụng biểu thức chính quy (regex) để trích xuất các từ, loại bỏ toàn bộ dấu câu và ký tự đặc biệt.
- Tiếp theo, chúng tôi làm sạch dữ liệu thêm một bước nữa, các từ dừng (stopwords) tiếng Anh được loại bỏ bằng cách sử dụng tập từ dừng có sẵn trong thư viện nltk. Những từ như "the", "is", "and" không mang nhiều ý nghĩa nên sẽ được loại bỏ.
- Sau khi làm sạch chúng tôi thu được một danh sách có dạng (“Từ”, “Tần suất”) được thống kê bằng công cụ Counter từ thư viện collections. Chúng tôi chọn ra 10 từ có tần suất cao nhất trong danh sách trên.
- Phương pháp này của chúng tôi có thể lọc ra các từ có tần suất cao nhất từ một lượng lớn văn bản. Tuy nhiên, nó cũng có một hạn chế là có thể bỏ qua một số từ quan trọng có tần suất thấp nhưng mang đặc trưng của tác giả và đó là từ có ý nghĩa trong câu.

Code:

```
import pandas as pd
import re
from collections import Counter
from nltk.corpus import stopwords

stop_words = set(stopwords.words('english'))
```

```

# Gộp tất cả câu thành một chuỗi lớn
van_ban = ''.join(df_stats['Câu nói'].dropna()).lower()

# Lọc chỉ lấy từ, không lấy dấu câu
tu = re.findall(r'\b\w+\b', van_ban)

# Loại bỏ stopwords
tu = [t for t in tu if t not in stop_words]

# Đếm tần suất
tan_suat = Counter(tu)

# Lấy top 10 từ
print(tan_suat.most_common(10))

```

Kết quả:

```
[('love', 9), ('know', 7), ('life', 6), ('give', 6), ('may', 6), ('without', 5), ('good', 5), ('never', 5), ('make', 5), ('friends', 5)]
```

Hình 3.2.2.8. Kết quả thống kê về các từ và số lần trong các câu nói.

- Chúng tôi sử dụng thực hiện chuẩn hóa bằng 2 dạng biểu đồ. Đầu tiên sử dụng thư viện matplotlib để vẽ biểu đồ cột để trực quan hóa tần suất của 10 từ phổ biến trên. Các cột được tô màu để dễ nhận diện, đồng thời mỗi cột có gắn thêm nhãn hiển thị số lần xuất hiện.

Code:

```

import matplotlib.pyplot as plt

# Lấy 10 từ phổ biến nhất

```

```

tu_pho_bien = tan_suat.most_common(10)
cac_tu, tan_suat = zip(*tu_pho_bien) # Giải nén các từ và tần suất

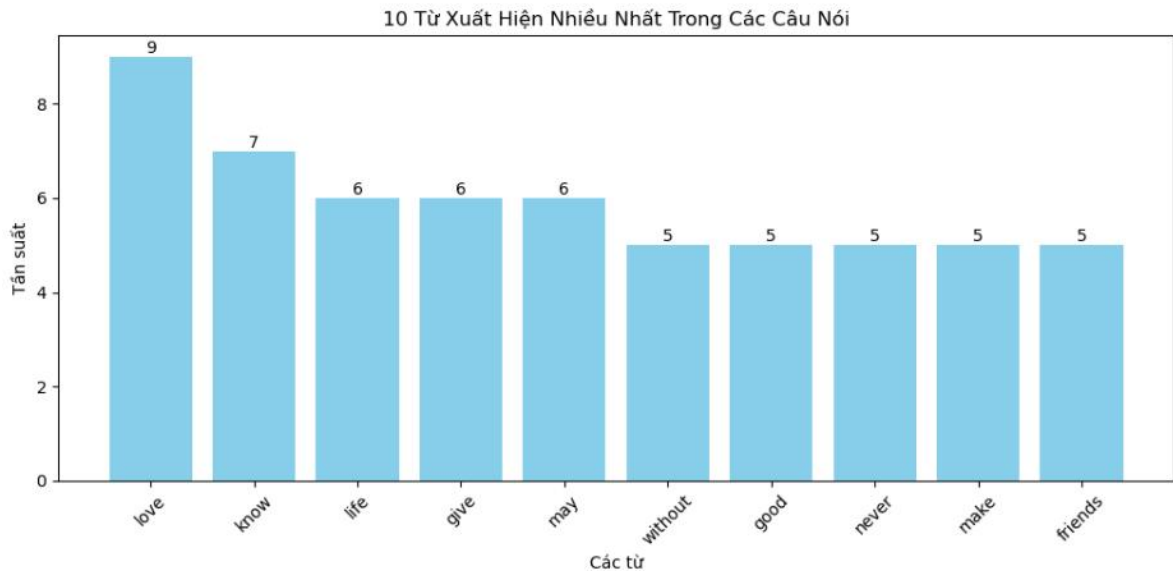
# Vẽ biểu đồ cột
plt.figure(figsize=(10, 5))
cot = plt.bar(cac_tu, tan_suat, color='skyblue') # Màu xanh biển

# Ghi nhãn số liệu trên đầu mỗi cột
for thanh in cot:
    gia_tri = thanh.get_height()
    plt.text(thanh.get_x() + thanh.get_width()/2, gia_tri, int(gia_tri), ha='center',
va='bottom')

# Thêm tiêu đề và nhãn trục
plt.title('10 Từ Xuất Hiện Nhiều Nhất Trong Các Câu Nói')
plt.xlabel('Các từ')
plt.ylabel('Tần suất')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

```

Kết quả:



Hình 3.2.2.9. Biểu đồ thể hiện 10 từ xuất hiện phổ biến nhất trong các câu nói nổi tiếng.

- Tiếp theo, chúng tôi sử dụng thư viện wordcloud để tạo biểu đồ đám mây từ nhằm trực quan hóa các từ được dùng nhiều nhất trong các câu nói. Các từ xuất hiện càng nhiều sẽ được hiển thị với kích thước lớn hơn, giúp dễ dàng nhận diện từ khóa nổi bật.

Code:

```
!pip install wordcloud
```

```
from wordcloud import WordCloud
```

```
import matplotlib.pyplot as plt
```

```
# Gộp tất cả các câu nói thành một chuỗi văn bản lớn
```

```
tat_ca_cau_noi = ''.join(df_stats['Câu nói'])
```

```
# Tạo đối tượng WordCloud
```


Yêu cầu 5: Phân tích, trực quan mối quan hệ giữa tác giả và câu nói nổi tiếng

- Trong bước đầu, chúng tôi thực hiện đếm số lượng câu nói nổi tiếng tương ứng với từng tác giả bằng cách sử dụng hàm `value_counts()` trên cột "Tác giả". Điều này giúp xác định xem mỗi tác giả có bao nhiêu câu trích dẫn trong bộ dữ liệu.

Code:

```
#- Phân tích, trực quan mối quan hệ giữa giữa tác giả và câu nói nổi tiếng
so_cau_noi_theo_tac_gia = df_stats['Tác giả'].value_counts()
so_cau_noi_theo_tac_gia
```

Kết quả:

Tác giả	
Albert Einstein	7
J.K. Rowling	5
Bob Marley	3
Dr. Seuss	3
Marilyn Monroe	2
Mother Teresa	2
Allen Saunders	1
Jorge Luis Borges	1
William Nicholson	1
Charles M. Schulz	1
Jim Henson	1
Garrison Keillor	1
Ralph Waldo Emerson	1
Pablo Neruda	1
Friedrich Nietzsche	1
Mark Twain	1
Elie Wiesel	1
Douglas Adams	1
Steve Martin	1
Eleanor Roosevelt	1
Thomas A. Edison	1
André Gide	1
Jane Austen	1
George Eliot	1
Name: count, dtype: int64	

Hình 3.2.2.11. Kết quả số lượng câu nói theo từng tác giả.

- Tiếp theo, chúng tôi sử dụng phương thức `describe()` để thống kê mô tả tổng quan về số lượng câu nói, từ đó có cái nhìn rõ hơn về sự phân bố của dữ liệu.

Code:

```
so_cau_noi_theo_tac_gia.describe()
```

Kết quả:

```
count    24.000000
mean      1.666667
std       1.493949
min       1.000000
25%       1.000000
50%       1.000000
75%       1.250000
max       7.000000
Name: count, dtype: float64
```

Hình 3.2.2.12. Kết quả thống kê tổng quan về số lượng câu nói của từng tác giả.

- Cuối cùng, dữ liệu được trực quan hóa dưới dạng biểu đồ cột, thể hiện số câu nói của từng tác giả, qua đó giúp dễ dàng so sánh mức độ hiện diện của các tác giả trong tập dữ liệu và nhận biết ai là người có nhiều câu nói nổi bật hơn.

Code:

```
# Trực quan hóa dữ liệu

plt.figure(figsize=(10, 6))

so_cau_noi_theo_tac_gia.plot(kind='bar', color='green')

plt.title('Phân phối của các câu nói nổi tiếng theo tác giả')

plt.xlabel('Tác giả')

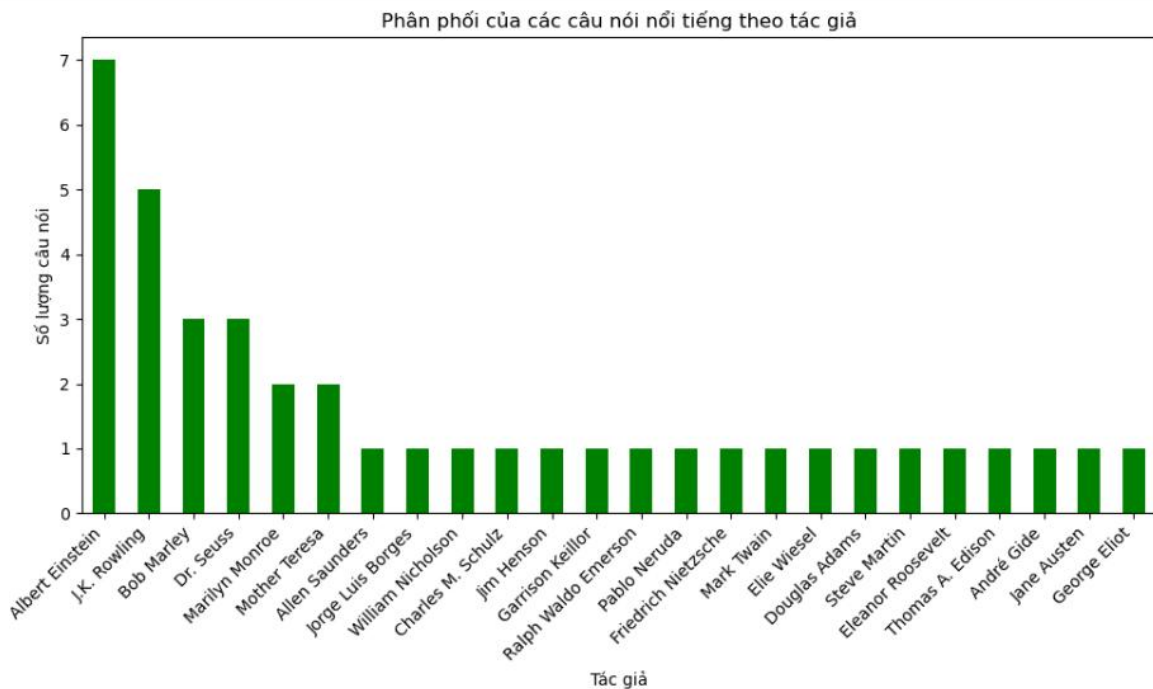
plt.ylabel('Số lượng câu nói')

plt.xticks(rotation=45, ha='right')
```

```
plt.tight_layout()
```

```
plt.show()
```

Kết quả:



Hình 3.2.2.13. Biểu đồ phân phối của các câu nói nổi tiếng theo tác giả.

Yêu cầu 6: Phân tích, trực quan mối quan hệ giữa các tác giả với nhau

- Để thực hiện yêu cầu phân tích, trực quan mối quan hệ giữa các tác giả với nhau, chúng tôi sử dụng thư viện Seaborn và NetworkX.
- Để khảo sát mối liên hệ giữa các tác giả dựa trên số lượng câu nói nổi tiếng, chúng tôi xây dựng một ma trận tương quan bằng cách sử dụng phương thức `pivot_table` của thư viện pandas. Ma trận này gồm các ô thể hiện số lượng câu nói chung giữa các cặp tác giả.
- Chúng tôi sẽ sử dụng heatmap với thư viện seaborn nhằm trực quan hóa cấu trúc dữ liệu, phát hiện ra sự trùng lặp hoặc điểm nổi bật nếu có.

- Biểu đồ heatmap được tô màu xanh biến từ đậm tới nhạt giúp dễ phân biệt mức độ mối liên hệ. Ô càng đậm thì số câu chung càng cao. Giúp nhận diện các nhóm tác giả có cùng điểm chung.
- Với phương pháp trên mà chúng tôi sử dụng sẽ cung cấp một cái nhìn trực quan, dễ tiếp cận về mối liên hệ giữa các tác giả, tuy nhiên, nó dựa trên giả định rằng việc xuất hiện chung trong dữ liệu phản ánh sự liên kết thực tế hoặc ảnh hưởng lẫn nhau điều này không phải lúc nào cũng chính xác.
- Bên cạnh đó, do quy mô dữ liệu còn hạn chế, nên kết quả phân tích có thể chưa phản ánh đầy đủ hoặc chính xác mọi mối quan hệ tiềm ẩn giữa các tác giả.
- Dù mô hình này không phản ánh sự hợp tác thực sự giữa các tác giả, nó vẫn hữu ích trong việc đánh giá nhanh phân bố dữ liệu và phát hiện điểm bất thường nếu có.

Code:

#- Phân tích, trực quan mối quan hệ giữa các tác giả với nhau

import seaborn as sns

import networkx as nx

Tạo dataframe hoặc ma trận số lượng câu nói chung giữa các tác giả

*Moi_quan_tac_gia = df_stats.pivot_table(index='Tác giả', columns='Tác giả',
aggfunc='size', fill_value=0)*

Trực quan hóa dữ liệu bằng heatmap

plt.figure(figsize=(12, 10))

*sns.heatmap(Moi_quan_tac_gia, cmap='Blues', annot=True, fmt='d',
linewidths=.5)*

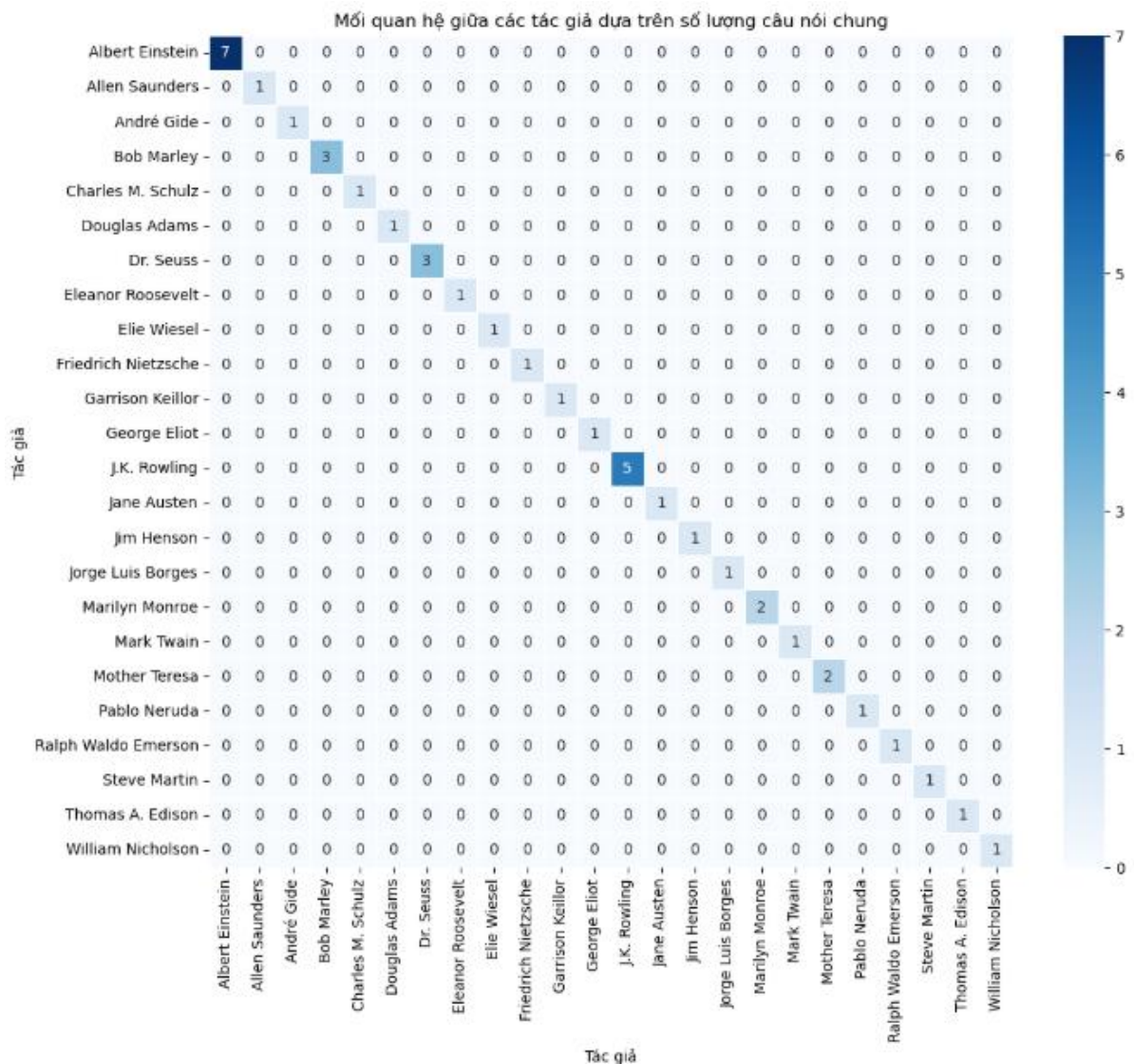
```
plt.title('Mối quan hệ giữa các tác giả dựa trên số lượng câu nói chung')
```

```
plt.xlabel('Tác giả')
```

```
plt.ylabel('Tác giả')
```

```
plt.show()
```

Kết quả:



Hình 3.2.2.14. Biểu đồ mối quan hệ giữa tác giả dựa trên số lần trùng câu nói nổi tiếng.

3.2.3. Trích xuất đặc trưng- Feature Extraction:

Đề xuất đặc trưng từ bộ dữ liệu các câu nói nổi tiếng, chúng tôi áp dụng các phương pháp sau:

- Tính độ dài của câu nói: Chúng tôi tiến hành đo số từ hoặc độ dài ký tự của từng câu nói. Việc này giúp đánh giá mức độ súc tích hay phức tạp của từng câu trích dẫn.
- Phân tích tần suất từ: Thông qua các thư viện xử lý ngôn ngữ tự nhiên như NLTK, chúng tôi đếm số lần xuất hiện của từng từ trong toàn bộ tập dữ liệu. Giúp hỗ trợ nhận diện những từ ngữ phổ biến, có thể mang giá trị biểu cảm hoặc ý nghĩa đặc biệt.
- Mã hóa thông tin tác giả: Đối với mỗi tác giả, chúng tôi xây dựng các biến giả (dummy variables) để chuyển thông tin định danh thành dạng số. Điều này giúp mô hình học máy nhận biết được đặc điểm riêng biệt của từng tác giả.
- Sử dụng mô hình vector từ (Word Embedding): Mỗi từ trong câu được ánh xạ sang một vector số, phản ánh ngữ nghĩa của từ trong ngữ cảnh chung. Các phương pháp như Word2Vec hoặc GloVe sẽ được sử dụng để thực hiện bước này.

Lý do chọn những cách tiếp cận này:

- Phân tích độc lập từng câu: Việc đo độ dài và phân tích tần suất từ giúp chúng tôi đánh giá từng câu trích dẫn một cách riêng biệt, tạo điều kiện cho mô hình hiểu rõ đặc điểm từng câu.
- Khai thác thông tin ngôn ngữ: Tính tần suất từ cùng với vector biểu diễn từ giúp chuyển nội dung ngôn ngữ sang dạng số học để xử lý, trong khi vẫn giữ được phần nào cấu trúc và ý nghĩa ngữ cảnh.
- Tăng cường thông tin đầu vào: Việc đưa yếu tố tác giả vào mô hình không chỉ giúp phân biệt phong cách thể hiện, mà còn cung cấp thêm đặc điểm riêng biệt trong các trích dẫn.

Sau khi thực hiện các bước trích xuất đặc trưng nêu trên, dữ liệu đã được chuẩn bị đầy đủ để tiếp tục xử lý và đưa vào huấn luyện với các thuật toán học máy phù hợp.

3.2.4. Suy luận:

Để giải quyết bài toán phân loại các câu nói nổi tiếng theo tác giả và đánh giá sự giống nhau về phong cách giữa các tác giả, chúng tôi đã xây dựng một quy trình bằng ngôn ngữ lập trình Python. Các bước được thực hiện lần lượt như sau:

- Xử lý và chuẩn bị dữ liệu:

Trước tiên, chúng tôi đọc dữ liệu từ file CSV có chứa các câu nói và tên tác giả tương ứng. Sau đó, loại bỏ những dòng bị thiếu dữ liệu để đảm bảo đầu vào sạch sẽ và đầy đủ trước khi phân tích.

- Chuyển văn bản thành số:

Vì mô hình học máy không thể hiểu trực tiếp văn bản, chúng tôi sử dụng phương pháp TF-IDF để biến các câu nói thành dạng số. Nhằm tính toán tần suất của từ trong mỗi câu mà còn đánh giá độ quan trọng của từ đó trong toàn bộ tập dữ liệu. Đồng thời, chúng tôi dùng cả từ đơn và cặp từ liên tiếp (n-gram 1 và 2) để tăng khả năng nắm bắt ý nghĩa.

- Huấn luyện và kiểm tra mô hình theo tỷ lệ Train/Test:

Chúng tôi chia tập dữ liệu được chia thành hai phần:

- + 80% để huấn luyện mô hình.
- + 20% còn lại dùng để kiểm tra hiệu quả mô hình.

Chúng tôi sử dụng mô hình Naive Bayes đa thức để phân loại văn bản.

- Kết quả:

Mô hình đã đạt được độ chính xác nhất định trên tập dữ liệu kiểm tra, với báo cáo phân loại chi tiết giúp hiểu rõ hơn về hiệu quả mô hình đối với từng tác giả cụ thể.

- Đánh giá và thảo luận:

- + Hiệu quả: Naive Bayes là một mô hình đơn giản, dễ áp dụng cho bài toán phân loại văn bản nhưng chưa phải là lựa chọn tối ưu do không nắm bắt tốt sự phụ thuộc giữa các từ trong câu.
- + Hạn chế: Do TF-IDF có số lượng đặc trưng giới hạn nên có thể làm mất đi một số thông tin quan trọng về phong cách ngôn ngữ.
- Đề xuất cải tiến:
 - + Thử các mô hình phức tạp hơn như SVM hoặc mạng nơ-ron.
 - + Sử dụng các kỹ thuật vector hóa nâng cao hơn như Word2Vec, GloVe, hoặc BERT để hiểu sâu hơn ngữ cảnh và ý nghĩa câu nói.
 - + Cân bằng dữ liệu giữa các tác giả hoặc tăng kích thước bộ dữ liệu để mô hình học được nhiều mẫu hơn.
- Tính độ giống nhau giữa các tác giả
 - + Tính trung bình các vector TF-IDF của tất cả câu nói thuộc về mỗi tác giả.
 - + So sánh các vector này với nhau bằng công thức tính độ tương đồng cosine, giúp xác định mức độ giống nhau về cách dùng từ, diễn đạt.
- Kết quả:

Chúng tôi tìm được cặp tác giả có phong cách nói giống nhau nhất dựa trên các câu nói trong dữ liệu. Hai tác giả có phong cách tương đồng nhất là Dr. Seuss và William Nicholson khoảng 0.142.
- Đánh giá:

Phương pháp tính trung bình vector TF-IDF đơn giản nhưng có thể mất một số đặc điểm riêng biệt của phong cách ngôn ngữ.
- Đề xuất cải tiến:

Thay thế phương pháp tính trung bình vector TF-IDF bằng cách áp dụng thêm các phương pháp khác như phân tích phong cách ngôn ngữ thông kê tần suất loại từ (part-of-speech), hay phân cụm các tác giả dựa trên điểm tương đồng để phát hiện nhóm tác giả có phong cách gần giống nhau.

Code:

```

import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report
from sklearn.metrics.pairwise import cosine_similarity
import numpy as np

# 1. Đọc dữ liệu
du_lieu = pd.read_csv("QuoteInsert_Tuoi.csv")

# 2. Lọc ra 2 cột cần thiết: Câu nói và Tác giả
du_lieu = du_lieu[['Câu nói', 'Tác giả']].dropna()

# 3. Trích xuất đặc trưng TF-IDF từ cột câu nói
bo_vector_hoa = TfidfVectorizer(stop_words='english', ngram_range=(1,2))
ma_tran_tfidf = bo_vector_hoa.fit_transform(du_lieu['Câu nói'])

# 4. Phân chia tập huấn luyện và kiểm tra
X_train, X_test, y_train, y_test = train_test_split(ma_tran_tfidf, du_lieu['Tác giả'], test_size=0.2, random_state=42)

# 5. Huấn luyện mô hình phân loại (Naive Bayes)
mo_hinh = MultinomialNB()
mo_hinh.fit(X_train, y_train)

```

```

# 6. Dự đoán và đánh giá kết quả
y_du_doan = mo_hinh.predict(X_test)
print("=== Báo cáo đánh giá ===")
print(classification_report(y_test, y_du_doan))

# === PHẦN 2: Tính độ tương đồng phong cách giữa các tác giả ===
# Tính vector trung bình TF-IDF cho mỗi tác giả
danh_sach_tac_gia = du_lieu['Tác giả'].unique()
vector_tac_gia = {}

for tac_gia in danh_sach_tac_gia:
    cau_noi_tac_gia = du_lieu[du_lieu['Tác giả'] == tac_gia]['Câu nói']
    tfidf_tac_gia = bo_vector_hoa.transform(cau_noi_tac_gia)

    # Sửa lỗi: Chuyển từ np.matrix -> np.ndarray bằng .A1
    vector_trung_binh = tfidf_tac_gia.mean(axis=0).A1
    vector_tac_gia[tac_gia] = vector_trung_binh

# Tính ma trận độ tương đồng
ten_tac_gia = list(vector_tac_gia.keys())
ma_tran_vectors = np.vstack([vector_tac_gia[tg] for tg in ten_tac_gia])
ma_tran_tuong_dong = cosine_similarity(ma_tran_vectors)

# Tìm hai tác giả giống nhau nhất
tuong_dong_cao_nhat = 0
tac_gia_1, tac_gia_2 = None, None

```

```

for i in range(len(ten_tac_gia)):
    for j in range(i + 1, len(ten_tac_gia)):
        diem = ma_tran_tuong_dong[i, j]
        if diem > tuong_dong_cao_nhat:
            tuong_dong_cao_nhat = diem
            tac_gia_1 = ten_tac_gia[i]
            tac_gia_2 = ten_tac_gia[j]

print(f'\\n==> Hai tác giả có phong cách nói giống nhau nhất là:\\n{tac_gia_1}
và {tac_gia_2} (điểm tương đồng: {tuong_dong_cao_nhat:.3f})")

```

Kết quả:

```

=== Báo cáo đánh giá ===

```

	precision	recall	f1-score	support
Albert Einstein	0.38	1.00	0.55	3
Allen Saunders	0.00	0.00	0.00	1
Douglas Adams	0.00	0.00	0.00	1
Elie Wiesel	0.00	0.00	0.00	1
J.K. Rowling	0.00	0.00	0.00	1
Marilyn Monroe	0.00	0.00	0.00	1
accuracy			0.38	8
macro avg	0.06	0.17	0.09	8
weighted avg	0.14	0.38	0.20	8

```

==> Hai tác giả có phong cách nói giống nhau nhất là:
Dr. Seuss và William Nicholson (điểm tương đồng: 0.142)

```

Hình 3.2.4.1. Kết quả báo cáo đánh giá, dự đoán sau khi huấn luyện mô hình và kết quả của độ tương quan cao nhất.

LÀM VIỆC NHÓM

- **Cách thức làm việc nhóm:**

Nhóm chúng tôi làm việc theo hình thức online và offline với cách phân chia công việc như sau:

Tên thành viên	Công việc
Trần Vĩnh Cơ	<ul style="list-style-type: none"> - Nghiên cứu, lựa chọn ra 60 câu hỏi/trả lời tiếng Anh, dịch và đánh giá bản dịch dựa trên bộ dữ liệu đã chọn lựa. - Dùng ChatGPT để sinh thêm dữ liệu, tương ứng với mỗi câu hỏi áp dụng 5 luật. - Tạo phiếu chấm đánh giá dữ liệu do Chat GPT sinh ra. - Kiểm tra và chỉnh sửa phiếu khảo sát. Thực hiện khảo sát với 30 phiếu khảo sát, nhập dữ liệu thu được từ quá trình khảo sát. - Code cho Bài 3 và viết báo cáo cho bài này.
Trần Nguyễn Toàn Phát	<ul style="list-style-type: none"> - Nghiên cứu, lựa chọn ra 60 câu hỏi/trả lời tiếng Anh, dịch và đánh giá bản dịch dựa trên bộ dữ liệu đã chọn lựa. - Dùng ChatGPT để sinh thêm dữ liệu, tương ứng với mỗi câu hỏi áp dụng 5 luật. - Tổng hợp dữ liệu các câu hỏi vào file. - Xây dựng bộ câu hỏi khảo sát, thiết kế phiếu khảo sát. - Thực hiện khảo sát với 30 phiếu khảo sát, nhập dữ liệu sau khi khảo sát. - Tổng hợp dữ liệu sau khảo sát, thực hiện phân tích dữ liệu bằng Python. - Viết báo cáo Bài 1, Bài 2, tổng hợp báo cáo Bài 3, định dạng và chỉnh sửa file báo cáo.

- **Tổng số lần gặp nhau:**
 - Họp online: 3 buổi (6 tiếng).
 - Gặp mặt offline: 4 buổi (16 tiếng).
- **Tổng thời gian gặp mặt: 22 tiếng.**

TÀI LIỆU THAM KHẢO

1. Alwagait, E., Shahzad, B., & Alim, S. (2015). *Impact of social media usage on students' academic performance in Saudi Arabia*. Computers in Human Behavior, 51, 1092-1097.
2. Alim, S. (2015). *Privacy issues in social media: The threats and solutions*. Procedia Computer Science, 65, 121-128.
3. Bernard John Kolan, Patience Emefa Dzandza, 2017. Effect of Social Media on Academic Performance of Students in Ghanaian Universities: A Case Study of University of Ghana, Legon. Library Philosophy and Practice. 1637, 1-24.
4. Barry, C. T., Sidoti, C. L., Briggs, S. M., Reiter, S. R., & Lindsey, R. A. (2017). *Adolescent social media use and mental health from adolescent and parent perspectives*. Journal of Adolescence, 61, 1-11.
5. Buran Köse, Doğan, 2019. The Relationship between Social Media Addiction and Self-Esteem among Turkish University Students. *Turkish Green Crescent Society*. ISSN 2148-7286. 176-190.
6. Giunchiglia, F., Zeni, M., Gobbi, E., Bignotti, E., & Bison, I. (2020). *Mobile social media use and academic performance*. Computers in Human Behavior, 104, 106-123.
7. Mustapha, Aliyu., Oguguo, Uchechukwu Chikodili., Ujevbe, Oke Benjamin & Mohammed, Babakolo Adamu (2019). *Effects of social media on students' achievement in learning automobile lighting system in technical colleges in Niger State. Conference Proceedings of the AITIE 3rd International Conference and Workshop on Innovation, Technology and Education (ICWITE, Abuja 2019)*, 284-289
8. Olebara, C. A., Nwosu, K. C., & Eze, C. J. (2021). *Time management and social media addiction among students*. Journal of Education and Practice, 12(6), 45-52.
9. Phạm Thị Kim và Nguyễn Hải Triều, 2020. Xử lý ý kiến phản hồi của người học dựa trên phương pháp phân loại văn bản. *Tạp chí Khoa học Đại học Đà Lạt*, 10(3), 52-66.
10. Lahiry, Sandeep; Choudhury, Shouvik; Chatterjee, Suparna; Hazra, Avijit. Impact of social media on academic performance and interpersonal relation: Across-section study among students at a tertiary medical center in East India. *Journal of Education and Health Promotion*. 8 (1), 73-79

11. Quan-Haase, A., & McCay-Peet, L. (2017). *Social media in everyday life: A qualitative study of usage and perceptions*. Information, Communication & Society, 20(1), 17-31.
12. Selwyn, N. (2009). *Faceworking: Exploring students' education-related use of Facebook*. Learning, Media and Technology, 34(2), 157-174.
13. Tayo, Subair S.; Adebola, Solomon Temitope; Yahya, Deborah Oreoluwa, 2019. Social Media: Usage and Influence on Undergraduate Studies in Nigerian Universities. *International Journal of Education and Development using Information and Communication Technology*, 15(3), 53-62.
14. Startquestion. (n.d.). *Online survey platform*. Retrieved from <https://www.startquestion.com/>
15. Jamal Al-Menayes, 2015. Social Media Use, Engagement and Addiction as Predictors of Academic Performance. *International Journal of Psychological Studies*.7(4); 2015, 86-94
16. Zahid Amin, Ahmad Mansoor, Syed Rabeet Hussain And Faisal Hashmat, 2016. Impact of Social Media of Student's Academic Performance. *International Journal of Business and Management Invention*, 5(4), 22-29.

PHỤ LỤC

PHIẾU KHẢO SÁT

PHẦN A: THÔNG TIN CÁ NHÂN

1. Họ và tên: _____ 2. Mã số sinh viên.....
3. Giới tính:
1. Nam ☐ 2. Nữ ☐
4. Khóa mà bạn đang học:
1. K17 ☐ 2. K18 ☐ 3. K19 ☐ 4. K20 ☐ 5. Khác ☐
5. Khoa mà bạn đang học:.....

PHẦN B: KHẢO SÁT VỀ MỨC ĐỘ ĐỒNG Ý

Vui lòng cho biết mức độ đồng ý của bạn bằng cách đánh dấu vào các ô vuông ứng với thang điểm từ 1 đến 5 theo quy ước sau:

- 1- Rất không đồng ý
- 2- Không đồng ý
- 3- Bình thường
- 4- Đồng ý
- 5- Hoàn toàn đồng ý

STT	Phát biểu	Mức độ đồng ý				
I.	Các thiết bị điện tử và kết nối mạng	1	2	3	4	5
1	Thư viện có đủ số lượng máy tính để đáp ứng đủ nhu cầu sử dụng của sinh viên.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Các máy tính trong thư viện hoạt động tốt, không bị giật lag khi sử dụng.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3	Bàn ghế và ổ cắm điện tại khu vực máy tính được bố trí hợp lý, thuận tiện cho sinh viên học tập.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Tình trạng mạng Wifi trong thư viện hoạt động ổn định và dễ dàng truy cập.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Hệ thống máy lạnh trong thư viện hoạt động tốt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
II. Không gian thư viện		1	2	3	4	5
1.	Không gian thư viện đủ rộng có thể phục vụ cho số lượng lớn sinh viên cùng lúc.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	Môi trường học tập trong thư viện yên tĩnh, giúp sinh viên tập trung hơn khi học.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	Số lượng phòng họp nhóm đáp ứng được nhu cầu làm việc nhóm của sinh viên.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	Cơ sở vật chất trong các phòng họp nhóm ở thư viện luôn được trang bị đầy đủ.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.	Khu vực ngồi nghỉ của sinh viên tại thư viện được bố trí hợp lý.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

PHẦN C: Ý KIẾN CÁ NHÂN

1- Theo bạn, thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh cần cải thiện những gì để nâng cao chất lượng dịch vụ?

.....

.....

.....

2- Bạn có nhận xét thế nào về chất lượng phục vụ của cán bộ thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh?

.....
.....
.....

3- Bạn có mong muốn thư viện trường Đại học Công nghiệp Thành phố Hồ Chí Minh thay thế/nâng cấp thiết bị nào không?

.....
.....
.....

TỰ ĐÁNH GIÁ

Câu	Nội dung	Điểm chuẩn	Tự chấm	Ghi chú
Chương 1 (30 điểm)	1.1	6		
	1.2	6		
	1.3	10		
	1.4	6		
	1.5	2		
Chương 2 (40 điểm)	2.1 Câu hỏi đặt ra	5		
	2.2 Xây dựng câu hỏi khảo sát	10		
	2.3 Phân tích			
	2.3.1 Kết quả khảo sát	10		
	2.3.2 Phân tích dữ liệu	15		
Chương 3 (30 điểm)	3.1 Thu thập dữ liệu	5		
	3.2 Khai phá dữ liệu			
	3.2.1. Xử lý dữ liệu	2		
	3.2.2. Khám phá dữ liệu	10		
	3.2.3. Trích xuất đặc trưng	3		
	3.2.4. Suy luận	10		
Báo cáo	(chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng)	10đ		
Làm việc nhóm	Chú ý trả lời đúng 4 yêu cầu trong phần làm việc nhóm	10đ		
Tổng điểm (120)				
Đôi qua thang điểm 10 (120=10 điểm)				

