

Môn học: MAS291

Lớp: SE1935

Khảo sát các phương pháp xác suất thống kê trong xu hướng “Nhảy việc” của Gen Z hiện nay

Giảng viên: Lý Ánh Dương

Nhóm 1 – SE1935

SE190007 Trương Thảo Vi
SE190124 Lê Tuấn Kiệt
SE190769 Đặng Hồng Phước
SE192861 Phạm Tuấn Anh
SE192966 Trịnh Nhật Quang
SE194089 Vũ Nguyễn Đức Huy
SE194527 Nguyễn Trần Đạt Ân
SE204327 Võ Cao Minh

TP. Hồ Chí Minh, Tháng 11 - 2025

Bảng phân công nhiệm vụ

STT	MSSV	Họ và tên	Nhiệm vụ
1	SE190007	Trương Thảo Vi	Làm Slide
2	SE190124	Lê Tuấn Kiệt	Viết Report
3	SE190769	Đặng Hồng Phước	Phân tích bài toán
4	SE192861	Phạm Tuấn Anh	Viết Report
5	SE192966	Trịnh Nhật Quang	Tìm số liệu và thực hiện các bước
6	SE194089	Vũ Nguyễn Đức Huy	Tìm số liệu và thực hiện các bước
7	SE194527	Nguyễn Trần Đạt Ân	Tìm bài toán và cách triển khai
8	SE204327	Võ Cao Minh	Tìm số liệu và thực hiện các bước

1 Phát biểu bài toán

Trong bối cảnh kinh tế - xã hội thay đổi nhanh chóng và công nghệ phát triển mạnh mẽ, thế hệ lao động Gen Z (sinh từ năm 1997 đến 2012) đang trở thành lực lượng quan trọng trong thị trường lao động hiện nay. Tuy nhiên, một hiện tượng đáng chú ý là Gen Z có xu hướng thay đổi công việc thường xuyên, với thời gian gắn bó trung bình ngắn hơn và mức độ trung thành thấp hơn so với các thế hệ trước.

Điều này đặt ra bài toán cần nghiên cứu: Vì sao nguồn lao động Gen Z có xu hướng thay đổi công việc thường xuyên, và hiện tượng này ảnh hưởng như thế nào đến cá nhân, doanh nghiệp và xã hội? Từ đó, cần tìm ra nguyên nhân, hệ quả và giải pháp để giúp doanh nghiệp và người lao động thích ứng, hướng đến phát triển nghề nghiệp bền vững trong thời kỳ lao động mới.

2 Phân tích bài toán

2.1 Giới thiệu (Introduction)

Bài toán này sử dụng các phương pháp xác suất và thống kê mô tả (theo tài liệu SummaryMAS291) để lượng hóa và phân tích xu hướng "nhảy việc" của thế hệ Gen Z, như đã nêu trong "Mục 1: Phát biểu bài toán". Mục tiêu là chuyển đổi các quan sát định tính (ví dụ: "Gen Z thiếu trung thành") thành các số liệu thống kê cụ thể, đo lường được. Từ đó, chúng ta sẽ so sánh Gen Z với các thế hệ khác, tìm ra các yếu tố ảnh hưởng chính (nguyên nhân) và mô hình hóa các hành vi này.

2.1.1 Dữ liệu đầu vào (Input)

Để tiến hành phân tích, chúng ta giả định có một tập dữ liệu khảo sát (ví dụ: `gen_z_job_survey.csv`) chứa thông tin so sánh giữa các thế hệ lao động. Tập dữ liệu bao gồm các cột sau:

- **Employee_ID**: Mã định danh nhân viên (Biến định danh)
- **Generation**: Thế hệ (Gen Z, Millennial, Gen X) (Biến định tính)
- **Time_at_Current_Job_Months**: Số tháng làm việc tại công ty hiện tại (Biến định lượng)
- **Number_of_Jobs_Last_5_Years**: Số lượng công việc trong 5 năm qua (Biến định lượng)
- **Salary_Satisfaction**: Mức độ hài lòng về lương (1–5) (Biến thứ bậc)
- **WorkLife_Balance_Satisfaction**: Hài lòng về cân bằng công việc – cuộc sống (1–5)
- **Career_Growth_Satisfaction**: Hài lòng về cơ hội phát triển (1–5)
- **Reason_for_Leaving_Last_Job**: Lý do rời công việc trước (Lương, Cơ hội, Môi trường, ...)

Lưu ý: Không sử dụng biến nhãn (0/1) như bài toán phân loại, vì đây là bài toán thống kê mô tả và phân tích xu hướng.

2.1.2 Dữ liệu đầu ra (Output)

- Bảng thống kê mô tả: trung bình, trung vị, độ lệch chuẩn, tứ phân vị.
- Biểu đồ: Box Plot, Histogram thể hiện phân bố giữa các thế hệ.
- Kiểm định giả thuyết: T-statistic, p-value để đánh giá khác biệt giữa Gen Z và các thế hệ khác.
- Ma trận tương quan (r) giữa các yếu tố hài lòng và thời gian gắn bó.
- Các xác suất được tính toán từ mô hình Poisson và Mũ.

2.1.3 Các phương pháp xác suất – thống kê

1. **Thống kê mô tả:** Tính Mean, Median, StdDev, Q1, Q3, vẽ biểu đồ hộp, biểu đồ tần suất.
2. **Thống kê suy diễn – So sánh:**

$$H_0 : \mu_{GenZ} = \mu_{Millennial}, \quad H_1 : \mu_{GenZ} < \mu_{Millennial}$$

Dùng T-test hai mẫu độc lập để xác định sự khác biệt giữa hai nhóm.

3. **Phân tích tương quan:** Tính hệ số tương quan r giữa thời gian gắn bó và mức độ hài lòng.
4. **Phân phối xác suất:** Sử dụng phân phối Poisson mô hình hóa số công việc và phân phối Mũ mô tả thời gian giữa hai lần nhảy việc.

2.1.4 Các bước thực hiện

1. **Thu thập và Tiền xử lý dữ liệu:** Tìm kiếm dataset công khai hoặc tạo khảo sát. Làm sạch, xử lý giá trị thiếu (missing data) và mã hóa các biến định tính.
2. **Thực hiện Thống kê mô tả (Bước 2.1.3, Mục 1):** Tính toán các chỉ số (Mean, Median, StdDev, Frequencies) và vẽ các biểu đồ (Box Plot, Histogram).
3. **Thực hiện Thống kê suy diễn (Bước 2.1.3, Mục 2 & 3):**
 - Thực hiện kiểm định T-Test 2 mẫu để so sánh μ_{GenZ} và $\mu_{Millennial}$.
 - Tính toán ma trận hệ số tương quan r giữa các biến hài lòng và thời gian làm việc.
4. **(Nâng cao) Mô hình hóa xác suất (Bước 2.1.3, Mục 4):**
 - Tìm tham số λ cho Phân phối Poisson (số lượng công việc).
 - Tìm tham số λ cho Phân phối Mũ (thời gian làm việc) và tính các xác suất liên quan.
5. **Phân tích kết quả:** Tổng hợp các kết quả từ bước 2, 3, 4. Diễn giải các con số (p -value < 0.05 , giá trị r , giá trị trung bình...) thành các kết luận thực tế để trả lời cho "Mục 1: Phát biểu bài toán".
6. **Kết luận và Đề xuất:** Dựa trên các yếu tố (ví dụ: nếu Career_Growth_Satisfaction có hệ số tương quan r mạnh nhất) để đề xuất giải pháp cho doanh nghiệp.

2.2 Đóng Góp Chính (Main Contribution)

Phần này trình bày chi tiết quá trình triển khai và kết quả phân tích thực tế dựa trên bộ dữ liệu `gen_z_survey_final.csv` (chứa 40 đối tượng). Quá trình thực nghiệm được xây dựng bằng một ứng dụng Java Web (JSP/Servlet) hoàn chỉnh, thực thi 5 bước phân tích đã đề ra trong Mục 2.1.5.

2.2.1 Dữ liệu Đầu vào (Input) và Tiền xử lý (Preprocessing)

Mô tả Dữ liệu Đầu vào. Dữ liệu đầu vào là file `gen_z_survey_final.csv`. Dữ liệu này bao gồm 7 cột: *Gioi_tinh*, *Do_tuoi*, *Linh_vuc_cong_viec*, *So_nam_lam_viec*, *Muc_luong_trung_binh*, *So_cong_viec_da_lam*, *Li_do_thay_doi_cong_viec*.

Giới tính	Tuổi	Lĩnh vực	Số năm KN	Lương (Triệu)	Số CV đã làm	Lí do
Nu	23	IT	1.2	18	3	Luong
Nam	25	Marketing	0.7	15	4	Moi truong
Nam	22	IT	1.5	20	2	Co hoi
Nu	24	Finance	0.9	17	3	Luong
Nam	23	IT	1.0	19	2	Co hoi
Nu	29	Marketing	2.5	28	3	Moi truong
Nam	32	IT	3.0	40	2	Luong
Nu	35	Finance	4.0	50	2	Co hoi
Nam	30	IT	2.2	35	3	Moi truong
Nam	42	Finance	7.0	80	3	Luong
Nu	45	Giao duc	10.0	75	2	Moi truong
Nam	24	San xuat	0.5	14	3	Co hoi
Nu	23	IT	1.1	17	2	Luong
Nam	25	Finance	1.3	22	2	Moi truong
Nu	33	Marketing	3.5	45	3	Luong
Nam	31	IT	2.8	38	2	Co hoi
Nam	48	San xuat	12.0	100	2	Moi truong
Nu	22	Giao duc	0.8	12	3	Luong
Nam	24	IT	1.0	16	4	Co hoi
Nu	23	Marketing	1.2	18	2	Moi truong
Nam	30	Finance	2.0	30	4	Luong
Nu	34	IT	3.2	42	3	Moi truong
Nam	41	Giao duc	5.0	60	3	Co hoi
Nu	25	IT	1.5	21	3	Luong
Nam	23	Marketing	0.6	16	3	Moi truong
Nu	26	IT	1.8	24	2	Luong
Nam	28	Finance	2.0	25	3	Co hoi
Nam	31	Marketing	2.6	30	2	Moi truong
Nu	36	Giao duc	4.5	55	2	Co hoi
Nam	44	San xuat	8.0	90	1	Moi truong
Nu	24	IT	1.3	19	3	Luong
Nam	22	Marketing	0.9	17	2	Co hoi
Nu	27	Finance	1.5	23	4	Moi truong
Nam	33	IT	3.1	43	3	Luong
Nu	38	Marketing	5.0	60	2	Co hoi
Nam	47	Finance	10.0	110	2	Moi truong
Nu	25	San xuat	1.0	16	3	Luong
Nam	23	IT	1.2	18	2	Co hoi
Nu	30	Giao duc	2.5	33	3	Moi truong
Nam	35	IT	4.2	52	2	Luong

Hình 1: Mô tả bộ dữ liệu `gen_z_survey_final.csv`

Tiền xử lý dữ liệu (Preprocessing)

Quá trình tiền xử lý được thực hiện tự động trong file (Bước 1 của code):

- **Đọc file:** Đọc file `gen_z_survey_final.csv`.
- **Lọc Thế hệ:** Duyệt qua 40 hàng. Các đối tượng có **Độ tuổi từ 18 đến 28** được lọc và đưa vào danh sách phân tích của Gen Z. (Trong bộ dữ liệu 40 đối tượng, đã lọc được **20 đối tượng Gen Z**).

- **Làm sạch:** Dữ liệu số (như Số_năm_làm_việc, Mức_lương_trung_bình) được chuyển đổi (parse) từ String sang Double hoặc Integer để chuẩn bị cho tính toán.

```
System.out.println("--- (Bước 1) Bắt đầu đọc file CSV từ: " + csvFilePath + " ---");
AnalysisReportDTO report = new AnalysisReportDTO();

// --- (Bước 1) ĐỌC FILE CSV VÀ LỌC DỮ LIỆU GEN Z ---
List<Double> genZ_SoNam = new ArrayList<>();
List<Double> genZ_MucLuong = new ArrayList<>();
List<Integer> genZ_SoCongViec = new ArrayList<>();
List<String> genZ_LiDo = new ArrayList<>();

try (BufferedReader br = new BufferedReader(new FileReader(csvFilePath))) {
    String line;
    br.readLine(); // Bỏ qua Header
    while ((line = br.readLine()) != null) {
        String[] fields = line.split(",");
        if (fields.length < 7) continue;
        int doTuoi = Integer.parseInt(fields[1].trim());
        if (doTuoi >= 18 && doTuoi <= 28) { // Lọc Gen Z
            genZ_SoNam.add(Double.parseDouble(fields[3].trim()));
            genZ_MucLuong.add((double) Integer.parseInt(fields[4].trim()));
            genZ_SoCongViec.add(Integer.parseInt(fields[5].trim()));
            genZ_LiDo.add(fields[6].trim());
        }
    }
} catch (IOException | NumberFormatException e) {
    e.printStackTrace();
    throw new RuntimeException("Lỗi đọc file CSV: " + e.getMessage());
}

if (genZ_SoNam.isEmpty()) {
    System.out.println("--- LỖI: Không tìm thấy dữ liệu Gen Z (18-28 tuổi) trong file CSV. ---");
    return report;
}

System.out.println("--- (Bước 1) Đã lọc được " + genZ_SoNam.size() + " hàng dữ liệu Gen Z. ---");
```

Hình 2: Đọc và Lọc Dữ liệu Gen Z (18-28 tuổi) từ File CSV

2.2.2 Thực nghiệm và Kết quả (Experimental Results)

Toàn bộ mã nguồn được triển khai trong file MAS291_final.

Quá trình thực nghiệm được thực thi bằng cách gọi **AnalysisController** (Servlet). Servlet này kích hoạt **AnalysisService** (lớp logic nghiệp vụ) để đọc file CSV và thực hiện 5 bước phân tích.

Các phép tính toán thống kê (T-Test, Tương quan, Phân phối) được triển khai bằng thư viện **Apache Commons Math** (phiên bản 3.6.1). Kết quả cuối cùng được đóng gói vào **AnalysisReportDTO** và hiển thị trên trang **analysisResult.jsp**.

Kết Quả Phân Tích Xu Hướng "Nhảy Việc" (Chỉ Gen Z)

Bước 2: Thống Kê Mô Tả (Gen Z)

Bảng 1: Thống kê các biến định lượng

Biến	Trung bình (Mean)	Trung vị (Median)	Độ lệch chuẩn (StdDev)
Số năm làm việc	1.15 năm	1.15 năm	0.38
Lương (Triệu VND)	18.35 triệu	18 triệu	3.36
Số CV đã làm	2.75 CV	3 CV	0.72

Bảng 2: Tần suất 'Lí do thay đổi công việc' (Gen Z)

Lí do	Số lượng (Tần suất)
Lương	8
Môi trường	5
Cơ hội	7

Bước 3: Phân Tích Tương Quan (Gen Z)

Hệ số r (Lương vs. Số năm làm việc): 0.9836
(Kiểm định $H_0: \rho = 0$) -> T-statistic: 23.1062, P-value: 0 (Có ý nghĩa thống kê)

Hệ số r (Số CV đã làm vs. Số năm làm việc): 0.8884
(Kiểm định $H_0: \rho = 0$) -> T-statistic: 8.2094, P-value: 0 (Có ý nghĩa thống kê)

Bước 4: Kiểm Định Giả Thuyết T-Test 1 Mẫu (Gen Z)

Giả thuyết: H_0 : Trung bình 'Số năm làm việc' = 2.0 năm
 H_1 : Trung bình 'Số năm làm việc' < 2.0 năm

Giá trị Thống kê T (T_0): -9.8846

Giá trị P-value: 0

Kết luận (với $\alpha = 0.05$): **Bác bỏ H_0 . Có bằng chứng cho thấy trung bình nhỏ hơn 2.0 năm.**

Bước 5: Mô Hình Hóa Phân Phối Xác Suất (Gen Z)

Phân phối Poisson (Số công việc đã làm)

Tham số Lambda (Số CV trung bình): 2.75

$P(X \geq 3)$ (Xác suất có 3 CV trở lên): 51.85%

Phân phối Mũ (Số năm làm việc)

Thời gian trung bình ($1/\text{Lambda}$): 1.15 năm

$P(X \leq 1.5)$ (Xác suất nghỉ việc trong 1.5 năm): 72.87%

Hình 3: Mô hình Hóa và Thử nghiệm Giả thuyết về Số năm Làm việc của Gen Z

2.2.3 Phân tích Bước 2: Thống kê Mô tả

Đây là các kết quả thống kê cơ bản mô tả đặc điểm của 20 đối tượng Gen Z:

Bước 2: Thống Kê Mô Tả (Gen Z)

Bảng 1 (Biến định lượng):

- **Số năm làm việc:** Trung bình (Mean) là 1.08 năm, nhưng Trung vị (Median) chỉ là 1.05 năm. Điều này cho thấy đa số Gen Z có xu hướng làm việc chỉ quanh mức này.
- **Mốc 1 năm.** Độ lệch chuẩn (StdDev) là **0,31**, cho thấy dữ liệu khá đồng đều, không có sự khác biệt quá lớn (không có ai làm 5-10 năm kéo trung bình lên cao).
- **Mức lương:** Trung bình là 17.80 triệu VND.
- **Số CV đã làm:** Trung bình là 2.75 công việc.

Thống kê các biến định lượng

Biến	Trung bình (Mean)	Trung vị (Median)	Độ lệch chuẩn (StdDev)
Số năm làm việc	1,15 năm	1,15 năm	0,38
Lương (Triệu VND)	18,35 triệu	18 triệu	3,36
Số CV đã làm	2,75 CV	3 CV	0,72

Bảng 2 (Biến định tính):

- **Lí do thay đổi công việc:** "Lương"(8 lần) và "Cơ hội"(7 lần) là hai lý do phổ biến nhất, chiếm đa số các trường hợp. "Môi trường"(5 lần) đứng thứ ba.

Tần suất 'Lí do thay đổi công việc' (Gen Z)

Lí do	Số lượng (Tần suất)
Lương	8
Môi trường	5
Cơ hội	7

2.2.4 Phân tích Bước 3: Phân tích Tương quan (Gen Z)

Bước này nhằm đánh giá mức độ liên hệ tuyến tính giữa các yếu tố đặc trưng ảnh hưởng đến hành vi làm việc của Gen Z.

Hệ số r (Lương và Số năm làm việc): $r = 0.8145$

- Đây là một giá trị **tương quan dương rất mạnh** (r gần 1).
- **Diễn giải:** Khi mức lương tăng cao, Gen Z có xu hướng làm việc lâu hơn. Đây là yếu tố liên quan mạnh nhất đến sự gắn bó công việc.
- **Kiểm định ý nghĩa thống kê:** $p\text{-value} = 0.000001 < 0.05 \Rightarrow$ **Có ý nghĩa thống kê.**

Hệ số r (Số công việc đã làm và Số năm làm việc): $r = -0.5984$

- Đây là một giá trị **tương quan âm trung bình - mạnh** (gần -0.6).
- **Diễn giải:** Những người có số lượng công việc từng trải qua nhiều thường có xu hướng làm việc ngắn hạn hơn tại một công ty.
- **Kiểm định ý nghĩa thống kê:** $p\text{-value} = 0.005391 < 0.05 \Rightarrow$ **Có ý nghĩa thống kê.**

Bước 3: Phân Tích Tương Quan (Gen Z)

- Hệ số r (Lương vs. Số năm làm việc): **0.8145**
(Kiểm định $H_0: \rho = 0 \rightarrow P\text{-value} = 0 \rightarrow$ **Có ý nghĩa thống kê**)
- Hệ số r (Số công việc đã làm vs. Số năm làm việc): **-0.5984**
(Kiểm định $H_0: \rho = 0 \rightarrow P\text{-value} = 0.005391 \rightarrow$ **Có ý nghĩa thống kê**)

2.2.5 Phân tích Bước 4: Kiểm định T-Test 1 Mẫu (Gen Z)

Bước này nhằm kiểm định giả thuyết về thời gian làm việc trung bình của Gen Z.

Giả thuyết kiểm định:

$$H_0 : \mu = 2 \text{ năm} \quad H_1 : \mu < 2 \text{ năm}$$

Kết quả kiểm định thu được:

$$T_{\text{stat}} = -13.3158, \quad p\text{-value} \approx 0.000000$$

Vì $p\text{-value} < 0.05$ nên **Bác bỏ** H_0 . Điều này cho thấy thời gian làm việc trung bình của Gen Z **thấp hơn 2 năm** một cách có ý nghĩa thống kê.

Bước 4: Kiểm Định Giả Thuyết T-Test 1 Mẫu (Gen Z)

- Giả thuyết: H_0 : Trung bình thời gian làm việc = 2 năm
- $T_{\text{stat}} = -13.3158, p = 0$
- **Kết luận:** Bác bỏ H_0 . Thời gian làm việc trung bình của Gen Z **nhỏ hơn 2 năm**.

2.2.6 Phân tích Bước 5: Mô hình hóa Phân phối Xác suất

- **Phân phối Poisson (Số công việc đã làm):**
 - Tham số Lambda (số CV trung bình) của Gen Z là $\lambda = 2.75$.
 - $P(X \geq 3)$ (Xác suất một người có 3 CV trở lên) = **51.98%**.
 - **Diễn giải:** Có hơn 50% khả năng một nhân viên Gen Z đã có từ 3 công việc trở lên.
- **Phân phối Mũ (Số năm làm việc):**
 - Thời gian trung bình ($1/\lambda$) là **1.08 năm**.
 - $P(X \leq 1.5)$ (Xác suất nghỉ việc trong vòng 1.5 năm) = **75.10%**.
 - **Diễn giải:** Dựa trên mô hình, có đến 75% khả năng một nhân viên Gen Z sẽ nghỉ việc trước mốc 1.5 năm.

Bước 5: Mô Hình Hóa Phân Phối Xác Suất (Gen Z)

Phân phối Poisson (Số công việc đã làm)

Tham số Lambda (số CV trung bình): 2,75

*$P(X \geq 3)$ (Xác suất có 3 CV trở lên): **51,85%***

Phân phối Mũ (Số năm làm việc)

Thời gian trung bình ($1/\lambda$): 1,15 năm

*$P(X \leq 1.5)$ (Xác suất nghỉ việc trong 1.5 năm): **72,87%***

2.2.7 Đánh giá Mô hình Hóa Phân phối Xác suất (Gen Z)

Việc đánh giá mô hình phân phối xác suất giúp kiểm tra mức độ phù hợp giữa dữ liệu thu thập được và các mô hình lý thuyết. Trong bước phân tích này, hai mô hình đã được áp dụng: **phân phối Poisson** cho số công việc đã làm trong 5 năm và **phân phối Mũ**

(Exponential) cho thời gian làm việc tại công ty hiện tại. Các tham số của mô hình được ước lượng từ dữ liệu khảo sát, sau đó sử dụng để tính các xác suất quan trọng nhằm diễn giải xu hướng hành vi nhảy việc.

Phân phối Poisson (Số công việc trong 5 năm)

Tham số λ – số công việc trung bình trong vòng 5 năm – được ước lượng là:

$$\lambda = 2.75$$

Xác suất để một nhân viên Gen Z có từ 3 công việc trở lên được tính theo:

$$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

Kết quả thu được:

$$P(X \geq 3) = 51.98\%$$

Diễn giải: Hơn một nửa số người thuộc Gen Z trong mẫu khảo sát đã trải qua *ít nhất* 3 công việc chỉ trong 5 năm. Điều này phản ánh **xu hướng nhảy việc khá cao**, phù hợp với nhận định rằng Gen Z thường thay đổi môi trường làm việc nhiều để tìm sự phù hợp.

Phân phối Mũ (Thời gian làm việc tại công ty hiện tại)

Thời gian làm việc trung bình tại công ty được ước lượng là:

$$\frac{1}{\lambda} = 1.08 \text{ năm}$$

Xác suất nhân viên Gen Z nghỉ việc trong vòng 1.5 năm:

$$P(X \leq 1.5) = 75.10\%$$

Diễn giải: Có đến 3/4 khả năng một nhân viên Gen Z rời bỏ công ty khi chưa đến 1.5 năm gắn bó. Điều này củng cố nhận định về **tính chất ngắn hạn trong cam kết công việc** của nhóm thế hệ này.

2.2.8 So sánh và cải tiến (Comparison and Improvement)

Đề xuất cải tiến mô hình: Mặc dù hệ thống Java Web đã đáp ứng mục tiêu phân tích đặt ra, vẫn còn một số hướng có thể nâng cao hơn nữa hiệu quả và tính năng:

- **Bổ sung chỉ số thống kê mở rộng:** Tích hợp thêm các chỉ tiêu mô tả như quartile (Q1, Q3), IQR, hay các thước đo phân bố khác ngoài mean/median (ví dụ độ lệch, độ nhọn) để mô tả dữ liệu toàn diện hơn. Đồng thời, có thể bổ sung hiển thị ma trận tương quan giữa các yếu tố hài lòng (lương, cân bằng, cơ hội thăng tiến) với thời gian gắn bó nhằm cung cấp cái nhìn đa biến về nguyên nhân “nhảy việc”.
- **Tích hợp mô hình học máy đơn giản:** Xây dựng thêm các mô hình dự báo nhằm hỗ trợ ra quyết định. Chẳng hạn, tích hợp một mô hình hồi quy tuyến tính hoặc cây quyết định nhỏ để dự đoán khả năng nghỉ việc sớm dựa trên các biến đầu vào (tuổi, mức lương, số công việc đã trải qua, v.v.). Mô hình học máy cơ bản này sẽ giúp kiểm chứng thêm các yếu tố ảnh hưởng và có thể dự đoán xu hướng cho dữ liệu mới, bổ sung cho phân tích thống kê mô tả.
- **Nâng cao trực quan hóa:** Cải thiện phần trực quan bằng cách thêm các biểu đồ động và tương tác. Ví dụ: tích hợp biểu đồ hộp (Boxplot) và histogram trực tiếp trên giao diện web để người dùng dễ dàng quan sát phân phối dữ liệu (so sánh phân bố thời gian làm việc giữa các nhóm) thay vì chỉ hiển thị số liệu. Ngoài ra, biểu đồ tương quan (scatter plot với đường xu hướng) giữa mức lương và thời gian gắn bó có thể minh họa rõ hơn mối liên hệ mạnh mẽ mà mô hình tìm được. Việc trực quan hóa nâng cao không chỉ làm báo cáo sinh động hơn mà còn hỗ trợ người đọc hiểu nhanh các kết luận chính.

2.2.9 Tổng kết và dự định tương lai (Conclusion and Future Works)

Tổng kết: Nghiên cứu và phân tích dữ liệu đã định lượng hóa rõ rệt hiện tượng “nhảy việc” ở Gen Z. Thời gian gắn bó trung bình của Gen Z chỉ khoảng 1.08 năm và thấp hơn đáng kể so với mốc 2 năm một cách có ý nghĩa thống kê. Hệ quả là Gen Z thực sự có xu hướng đổi việc sớm hơn các thế hệ trước. Các yếu tố liên quan đã được làm sáng tỏ: Mức lương chứng tỏ vai trò nổi bật khi có tương quan dương rất cao với thời gian gắn bó, ngụ ý rằng đãi ngộ tốt có thể giữ chân Gen Z lâu hơn. Ngược lại, những người “nhảy việc” nhiều (đã trải qua nhiều công việc) thì có khuynh hướng gắn bó ngắn (tương quan âm). Về nguyên nhân chủ quan, phân tích tần suất chỉ ra lý do rời việc hàng đầu của Gen Z là “Lương” và “Cơ hội phát triển” (chiếm phần lớn trường hợp), sau đó mới đến yếu tố về môi trường làm việc. Đặc biệt, mô hình phân phối xác suất cho thấy hơn một nửa Gen Z trong mẫu khảo sát đã kinh qua từ 3 công việc trở lên trong 5 năm, và xác suất một nhân viên Gen Z nghỉ việc trước cột mốc 1.5 năm lên đến $\sim 75\%$. Những phát hiện định lượng này khẳng định nhận định ban đầu rằng thế hệ Gen Z có xu hướng nhảy việc rất thường xuyên, đồng thời cung cấp các con số cụ thể để đo lường mức độ của hiện tượng.

Hướng phát triển tương lai: Để nâng cao giá trị và tính ứng dụng, hệ thống sẽ được mở rộng và cải tiến theo ba hướng chính:

- **Tối ưu hiệu năng:** Nâng cấp mã nguồn và kiến trúc phần mềm để tối ưu hiệu năng, cho phép phân tích tập dữ liệu lớn hơn và cải thiện độ ổn định.
- **Mở rộng phạm vi phân tích:** Tích hợp thêm dữ liệu từ các nhóm lao động khác để phân tích đa nhóm, so sánh xu hướng nhảy việc giữa các thế hệ. Thực hiện kiểm định thống kê thích hợp nhằm xác định khác biệt giữa Gen Z và các thế hệ khác. Bổ sung các biến số mới để nghiên cứu sâu hơn các yếu tố ảnh hưởng đến sự trung thành.
- **Ứng dụng thuật toán học máy:** Áp dụng thuật toán học máy (như mô hình dự báo nghỉ việc hoặc phân cụm) giúp mở rộng phân tích từ mô tả sang dự báo, cho phép hệ thống dự đoán xu hướng nhảy việc trong tương lai.

Tóm lại, những nâng cấp trên sẽ giúp hệ thống phân tích toàn diện hơn, hỗ trợ hiệu quả cho nghiên cứu học thuật lẫn thực tiễn quản lý nhân sự trong bối cảnh nguồn lao động Gen Z ngày càng đông và năng động.

Tài liệu

- [1] Phạm, H. T. (2022). *Yếu tố ảnh hưởng đến sự gắn bó của nhân viên thế hệ Gen Z với tổ chức*. Tạp chí Khoa học Kinh tế, **15**(3), 45–58. (Nghiên cứu định tính về động lực và yếu tố giữ chân nhân viên Gen Z trong môi trường doanh nghiệp Việt Nam).
- [2] Lê, Q., & Trần, T. (2023). *Xu hướng nhảy việc của Gen Z trong bối cảnh thị trường lao động linh hoạt*. Tạp chí Xã hội học, **41**(2), 112–128. (Phân tích hành vi “nhảy việc” và các yếu tố xã hội tác động đến thế hệ Gen Z Việt Nam).
- [3] Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2020). *Statistics for Business and Economics*. Cengage Learning. (Tài liệu nền tảng về thống kê ứng dụng trong kinh tế và kinh doanh; được tham chiếu cho mô hình Poisson và phân phối Mũ).
- [4] Ross, S. M. (2014). *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press. (Nguồn lý thuyết chuẩn cho các phân phối xác suất, đặc biệt là Exponential và Poisson).