TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐHQG _ TPHCM

KHOA CÔNG NGHỆ THÔNG TIN _ CLC

BẢN BÁO CÁO
PROJECT 03

Môn: *Toán ứng dụng và thống kê cho công nghệ thông tin*
Lớp: *21CLC05*
Mã môn: *MTH00057*
Giảng viên thực hành: *Phan Thị Phương Uyên*
Nguyễn Văn Quang Huy

HỌ TÊN: VÕ ANH QUÂN
MSSV : 21127147

MỤC LỤC

1. GIỚI THIỆU

Thông tin cá nhân

Họ tên: *Võ Anh Quân*

Mã số sinh viên: *21127147*

Lớp: *21CLC05*

Nội dung đồ án

Đồ án 03: Linear Regression

Mục tiêu của đồ án là tìm hiểu các yếu tố quyết định mức lương và việc làm của các kỹ sư ngay sau khi tốt nghiệp. Các yếu tố như điểm số ở các cấp/trường đại học, kỹ năng của ứng viên, sự liên kết giữa trường đại học và các khu công nghiệp/công ty công nghệ, bằng cấp của sinh viên và điều kiện thị trường cho các ngành công nghiệp cụ thể sẽ ảnh hưởng đến điều này. Bộ dữ liệu được sử dụng trong đồ án này thu thập tại Ấn Độ, nơi có hơn 6000 cơ sở đào tạo kỹ thuật công nghệ với khoảng 2,9 triệu sinh viên đang học tập. Mỗi năm, trung bình có 1,5 triệu sinh viên tốt nghiệp chuyên ngành Công nghệ/Kỹ thuật, tuy nhiên do thiếu kỹ năng cần thiết, ít hơn 20% trong số họ có việc làm phù hợp với chuyên môn của mình. Bộ dữ liệu này không chỉ giúp xây dựng công cụ dự đoán mức lương mà còn cung cấp thông tin về các yếu tố ảnh hưởng đến mức lương và chức danh công việc trên thị trường lao động. Sinh viên sẽ được khám phá những thông tin này trong phạm vi đồ án.

Dữ liệu ban đầu bao gồm 2998 dòng và 34 cột.
Sau khi được lượt bỏ những cột có giá trị chuỗi hoặc liên quan đến định danh và năm, bộ dữ liệu mới sẽ có 2998 mẫu dữ liệu và 23 đặc trưng.

Nội dung sơ lược về phần bài làm

Bài làm trình bày những yêu cầu sau:

Xây dựng mô hình dự đoán mức lương của kỹ sư sử dụng mô hình hồi quy tuyến tính theo các yêu cầu sau:

- **Yêu cầu 1a:**
    - Sử dụng 11 đặc trưng đầu tiên đề bài cung cấp bao gồm: Gender, 10percentage, 12percentage, CollegeTier, Degree, collegeGPA, CollegeCityTier, English, Logical, Quant, Domain.

- Thể hiện công thức cho mô hình hồi quy tuyến tính.
- Báo cáo 1 kết quả trên tập test.csv cho mô hình vừa huấn luyện được.

- **Yêu cầu 1b:**
  - Thử nghiệm lần lượt trên các đặc trưng tính cách gồm: conscientiousness, agreeableness, extraversion, neuroticism, openness_to_experience.
  - Sử dụng k-fold CrossValidation (với k = 5) đề tìm đặc trưng tốt nhất trong các đặc trưng tính cách.
  - Báo cáo 5 kết quả trung bình tương ứng cho 5 mô hình từ 5-fold Cross Validation.
  - Thể hiện công thức cho mô hình hồi quy theo đặc trưng tốt nhất.
  - Báo cáo 1 kết quả trên tập test.csv cho mô hình với đặc trưng tốt nhất tìm được.
- **Yêu cầu 1c:**
  - Thử nghiệm lần lượt trên các đặc trưng kỹ năng gồm: English, Logical, Quant.
  - Sử dụng k-fold CrossValidation (với k = 5) đề tìm đặc trưng tốt nhất trong các đặc trưng kỹ năng.
  - Báo cáo 3 kết quả trung bình tương ứng cho 3 mô hình từ 3-fold Cross Validation.
  - Thể hiện công thức cho mô hình hồi quy theo đặc trưng tốt nhất.
  - Báo cáo 1 kết quả trên tập test.csv cho mô hình với đặc trưng tốt nhất tìm được.
- **Yêu cầu 1d:**
  - Xây dựng 4 mô hình khác nhau.
  - Sử dụng phương pháp k-fold Cross Validation (k = 5) để tìm ra mô hình tốt nhất trong 4 mô hình mà sinh viên xây dựng.

- Báo cáo 4 kết quả trung bình tương ứng cho 4 mô hình từ 5-fold Cross Validation.

2. CÁC THƯ VIỆN SỬ DỤNG VÀ LÝ DO SỬ DỤNG TRONG BÀI LÀM

Thư viện pandas:

- Thư viện pandas được sử dụng đọc dữ liệu từ 2 file test.csv và train.csv.
- Biến đổi các kết quả (VD: weight, MAE) về dạng dataframe để in kết quả được rõ ràng.

Thư viện numpy:

- Thư viện numpy được sử dụng để chuyển các dữ liệu từ dạng dataframe về numpy thông qua hàm to_numpy() để có thể hậu xử lý các dữ liệu đó.
- Sử dụng các hàm tính toán của numpy như np.square().

Thư viện sklearn:

- Sử dụng hàm shuffle từ thư viện sklearn để xáo trộn bộ dữ liệu từ tập train, áp dụng cho các yêu cầu sử dụng k-fold Cross Validation.

Thư viện matplotlib. Module pyplot:

- Thư viện được sử dụng để plot dữ liệu thành các biểu đồ điểm của từng đặc trưng, phục vụ cho việc tìm ra mô hình tốt nhất trong yêu cầu 1d.

Thư viện seaborn:

- Thư viện seaborn được sử dụng để tạo ma trận mức độ tương quan giữa các đặc trưng với nhau và giữa các đặc trưng với mức lương, phục vụ cho việc tìm ra mô hình tốt nhất thông qua giá trị tương quan.

Thư viện copy:

- Sử dụng để deep copy dữ liệu từ tập các tập X_train, X_test, y_train, y_test, đảm bảo việc xáo trộn hay xử lý trên các dữ liệu đã copy sẽ không ảnh hưởng đến dữ liệu gốc.

3. CÁC HÀM SỬ DỤNG VÀ GIẢI THÍCH CÁC THAM SỐ

Từ thư viện numpy:

- Hàm zeros(): để tạo một array các giá trị 0 với kích thước được truyền vào.
- Hàm concatenate(): để nối 2 array nhau (axis=0 trong bài này), sử dụng cho việc tạo các tập train trong quá trình k-fold CrossValidation.

Từ thư viện matplotlib:

- plt.figure: để tạo biểu đồ chính có kích thước là figsize=(15, 12)
- plt.subplot: để tạo các biểu đồ điểm phụ tương ứng với từng đặc trưng.
- plt.scatter: để plot dữ liệu thành các điểm với trục x là các đặc trưng, trục y là mức lương. (marker='o' để sử dụng chấm tròn cho các điểm).
- plt.title, plt.tight_layout, plt.show: đặt tên cho từng biểu đồ con và sắp xếp chúng dưới 1 khung sao cho khung bị đè lên nhau, sau đó thể hiện biểu đồ ra bằng hàm show().

Từ thư viện sklearn:

- Hàm shuffle: với các tham số được dùng trong bài:
    - arrays(tuple): mảng cần xáo trộn dưới dạng tuple.
    - Random_state: Xác định seed để tạo số ngẫu nhiên cho việc hoán đổi. Trong bài sử dụng (random_state=42) cho quá trình Cross Validation.

Từ thư viện seaborn:

- Hàm heatmap: với các tham số được sử dụng trong bài:
    - Data: dữ liệu đầu vào cho biểu đồ heatmap. Ở đây là tập train

- Annot : Xác định liệu giá trị của từng ô trong biểu đồ có nên hiển thị bằng văn bản hay không. (ở đây là True)
- Cmap : Màu sắc được sử dụng trong biểu đồ heatmap (trong bài là cmap="YlGnBu")
- Fmt : Định dạng chuỗi được sử dụng để hiển thị các giá trị trong ô. (trong bài là fmt=".2f" để lấy 2 số sau dấu phẩy của các giá trị tương quan).

Từ thư viện copy :

- Hàm copy.deepcopy() : tham số truyền vào là một đối tượng cần sao chép (ở đây là một dataframe).

Các hàm cài đặt sẵn :

- Class OLSLinearRegression.
- Hàm mae(y, y_hat) : để tính giá trị MAE theo công thức.

Các hàm tự cài đặt :

Hàm find_best_feature_with_mae(X_train_specific, k):

- Chuẩn bị một mảng mae_list bao gồm các giá trị 0 để lưu trữ các giá trị MAE. Mảng mae_res để lưu trữ giá trị MAE cùng với tên đặc trưng tương ứng.
- Hàm find_best_feature_with_mae(X_train_specific, k) được sử dụng cho quá trình Cross Validation với tham số X_train_specific là tập X_train sau khi đã được lấy các cột cần tính, k là số k-fold (k = 5).
- Hàm có 2 vòng lặp chính (gọi là i và j)
- Ở vòng lặp i: lần lượt tính vị trí bắt đầu và kết thúc trong tập X_train, để chia ra 1 tập test (X_train_fold, y_train_fold) và 4 tập còn lại là tập train (X_remain, y_remain).
- Ở vòng lặp j: lần lượt lặp qua các đặc trưng, lấy ra từng đặc trưng và huấn luyện với tập X_train_fold, y_train đã có bằng hàm fit() từ class OLSLinearRegression. Tiếp đến, tính các giá trị mae của đặc trưng thứ j và cộng vào một mảng.

- Lặp qua mảng mae_list và tính trung bình MAE của từng đặc trưng và tiếp tục lưu vào mảng mae_res để lưu giá trị MAE trung bình ứng với tên đặc trưng.

4. KẾT QUẢ VÀ NHẬN XÉT

Kết quả cho yêu cầu 1a :

$$MAE_{first11features} = 104863{,}7775$$

Kết quả và nhận xét cục bộ cho yêu cầu 1b

Giá trị MAE trung bình của các đặc trưng sau khi Cross Validation như sau :

Đặc trưng tính cách	Giá trị MAE trung bình
Conscientiousness	306311.589183
Agreeableness	300904.339371
Extraversion	307023.946425
Neuroticism	299572.187800
Openness_to_experience	302950.200467

Giá trị MAE của Neuroticism trên toàn bộ tập test:

$$MAE_{Neuroticism} = 291019.6932$$

Nhận xét:

- Từ kết quả MAE trung bình của từng đặc trưng được ghi nhận ở bảng trên, có thể thấy được 2 đặc trưng tính cách ảnh hưởng nhất đến mức lương là neuroticism và agreeable trong khi 2 đặc trưng ít tương quan nhất đến lương là extraversion và conscientiousness.
- Điều này có thể được hiểu rằng những người có tính cách hòa đồng, dễ hợp tác (agreeableness) và mức độ cảm xúc, căng thẳng của người đó (neuroticism) có sự tương quan cao, cần thiết đặc biệt trong lĩnh vực engineering. Dẫn đến việc những người có 2 đặc trưng tính cách này thường có lương cao hơn.

- Extraversion (năng động, hướng ngoại) và conscientiousness (cẩn thận, trách nhiệm) là 2 tính cách tuy tích cực nhưng lại không thực sự ảnh đến mức lương trong môi trường làm việc thuộc lĩnh vực engineering.
- Openness_to_experience có giá trị MAE trung bình là 302950.200467, nằm ở khoảng giữa so với 4 tính cách còn lại, đặc trưng này nói lên tính cách biết đón nhận những trải nghiệm, kinh nghiệm mới trong quá trình làm việc. Tính cách này cũng có ảnh hưởng nhất định với mức lương của kĩ sư.

Kết quả và nhận xét cục bộ cho yêu cầu 1c:

Giá trị MAE trung bình của các đặc trưng kĩ năng sau quá trình 5-fold Cross Validation được ghi nhận như sau:

Đặc trưng kĩ năng	Giá trị MAE trung bình
English	121921.863413
Logical	120270.240819
Quant	118122.118030

Giá trị MAE của Quant khi huấn luyện trên toàn bộ tập test:

$$MAE_{Quant} = 106819.5776$$

Nhận xét:

- Từ các giá trị MAE trung bình ghi nhận được của các kĩ năng, có thể thấy được MAE của Quant là thấp nhất, thấp thứ 2 là Logical và cuối cùng là English. Các giá trị MAE không có sự chênh lệch rõ ràng.
- Các kĩ năng English (tiếng anh), Logical (khả năng logic) và Quant (khả năng định lượng) đều có vai trò quan trọng và sự tương quan nhất định trong việc dự đoán mức lương. Một kĩ sư có khả năng giao tiếp bằng tiếng anh, khả năng suy luận logic và khả năng định lượng tốt sẽ có xu hướng nhận được lương cao hơn.

Kết quả và nhận xét cục bộ cho yêu cầu 1d:

Công thức của từng mô hình

Mô hình 1:

$$Salary = w1 \times ComputerScience + w2 \times collegeGPA + w3 \times ComputerProgramming + w4 \times Domain + w5 \times CollegeTier + w6 \times 10percentage$$

Mô hình 2:

$$Salary = w1 \times (10percentage^2 + 12percentage^2) + w2 \times (agreeableness^2 + openessToExperience^2 + conscientiousness^2) + w3 \times (Logical^2 + Quant^2) + w4 \times (Domain^2 + ComputerProgramming^2)$$

Mô hình 3:

$$Salary = w1 \times (10percentage^2 + 12percentage^2) + w2 \times ComputerScience + w3 \times ComputerProgramming + w4 \times Domain + w5 \times CollegeTier + w6 \times Quant + w7 \times CivilEngg + w8 \times ElectronicsAndSemicon + w9 \times ElectricalEngg$$

Giá trị MAE trung bình của từng mô hình sau quá trình 5-fold Cross Validation được ghi nhận như sau:

Mô hình	Giá trị MAE trung bình
Mô hình 1	113637.172238
Mô hình 2	113188.803484
Mô hình 3	111251.911142

Giá trị MAE của mô hình 3 (mô hình có MAE trung bình thấp nhất) khi huấn luyện trên toàn bộ tập test là:

$$MAE_{ThirdModel} = 99456.6383$$

Nhận xét:

- Từ các giá trị MAE đã ghi nhận, có thể thấy MAE trung bình của mô hình 3 là thấp nhất (111251.911142), trong khi mô hình 1 và 2 có giá trị MAE cao hơn và tương đương nhau.
- Mô hình 1 tuy sử dụng các đặc trưng có tương quan với lương nhưng không có sự biến đổi dữ liệu hay tạo ra các đặc trưng mới. Tương tự, mô hình 2 sử dụng các dữ liệu có sự biến đổi và tương quan cao với lương và các đặc trưng được tạo mới, cả 2 mô hình đầu tiên đều cho giá trị MAE ổn định và không quá lớn.
- Ngược lại, mô hình 3 chỉ sử dụng 2 đặc trưng 10percentage và 12percentage để biến đổi và tạo ra 1 đặc trưng mới, các đặc trưng được chọn còn lại được lọc trực tiếp từ tập train và đều có sự tương quan cao với lương, và mang tính thực tế hơn các đặc trưng khác trong tập train nên cho giá trị MAE thấp hơn (thấp nhất trong kể cả với các yêu cầu 1a, 1b, 1c)

Nhận xét toàn cục:

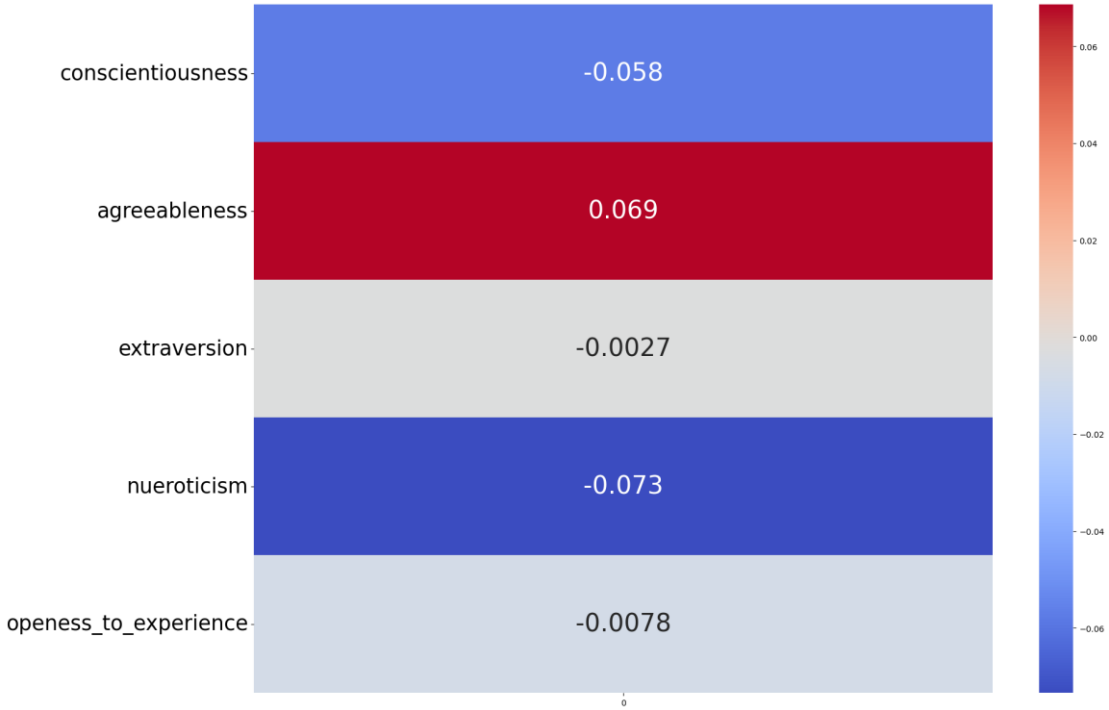
- Từ những kết quả MAE được nêu trên, có thể thấy mô hình tốt nhất là mô hình 3 ở yêu cầu 1d, với MAE = 99456.6383.
- Mô hình 3 sử dụng 8 đặc trưng có sẵn và một đặc trưng mới được tạo ra bằng cách lấy tổng bình phương 2 đặc trưng 10percentage và 12percentage.
- Mô hình có MAE thấp nhất là mô hình sử dụng 1 đặc trưng Neuroticism, có thể hiểu rằng Neuroticism là một khía cạnh của tính cách liên quan đến cảm xúc kể cả những cảm xúc tiêu cực như căng thẳng, lo âu, thiếu ổn định. Những kĩ sư có tính cách dễ biến đổi tâm lý, tâm trạng không ổn định dẫn đến làm việc thiếu hiệu quả, năng suất. Vì vậy, dù có chỉ số neuroticism cao hay thấp, mức lương của kỹ sư vẫn sẽ phụ thuộc vào những kỹ năng chuyên môn, kỹ năng giao tiếp, giải quyết vấn đề nhiều hơn.

- Mô hình ở yêu cầu 1c tuy chỉ sử dụng 1 đặc trưng Quant (khả năng định lượng) nhưng lại cho chỉ số MAE ấn tượng (106819.5776), tuy đây không phải là chỉ số MAE thấp nhất nhưng với việc chỉ sử dụng 1 đặc trưng thì MAE của mô hình này là tốt và thiết thực.

5. GIẢ THUYẾT / GIẢI THÍCH

Yêu cầu 1b: Đặc trưng tính cách tốt nhất là Neuroticism

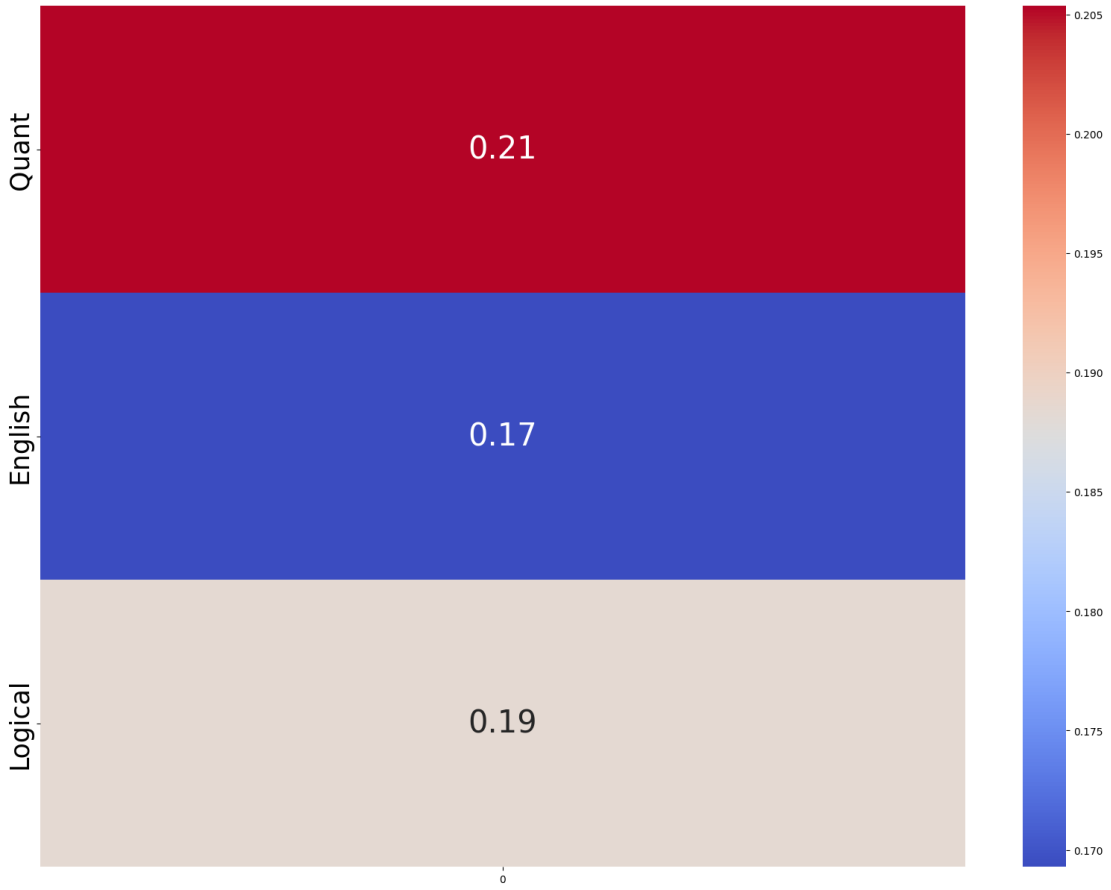
Ta khảo sát ma trận tương quan giữa 5 tính cách và mức lương dưới đây:



Có thể thấy được Neuroticism có giá trị tương quan là -0.073, tuy giá trị tương quan không cao (không đồng biến) nhưng lại có độ lớn giá trị tương quan (0.073) lớn nhất trong các đặc trưng còn lại. Tức là ảnh hưởng nhất đến mức lương.

Yêu cầu 1c: Đặc tính kỹ năng tốt nhất là Quant

Ta khảo sát ma trận tương quan giữa 3 đặc trưng kỹ năng và mức lương dưới đây:



Từ sơ đồ tương quan giữa 3 đặc trưng kĩ năng là English, Logical, Quant với mức lương, có thể thấy hệ số tương quan giữa Quant và mức lương là cao nhất với 0.21, tiếp sau đó là English (0.17) và Logical (0.19)

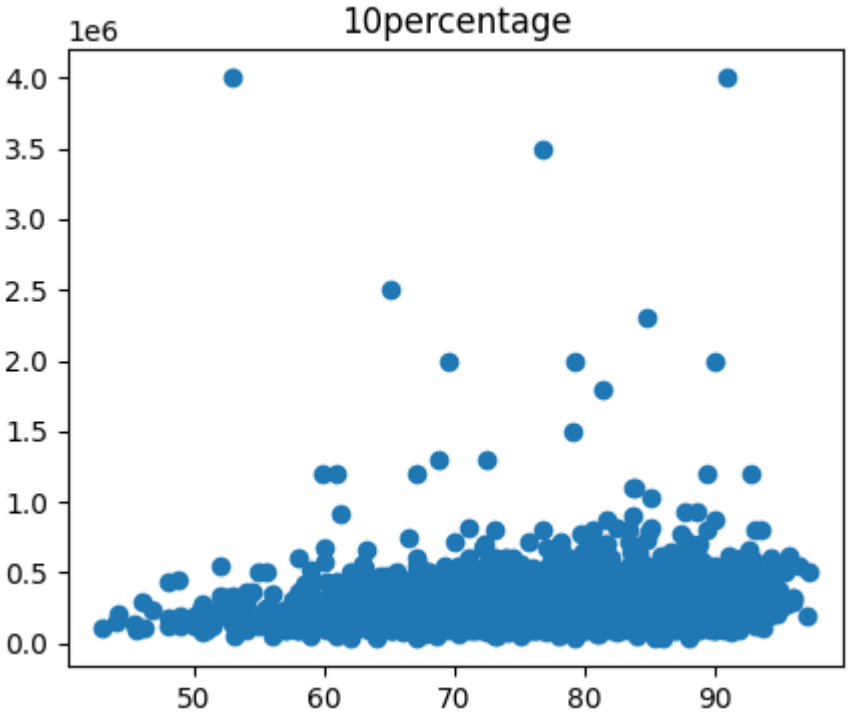
Yêu cầu 1d: Công thức của mô hình tốt nhất như sau

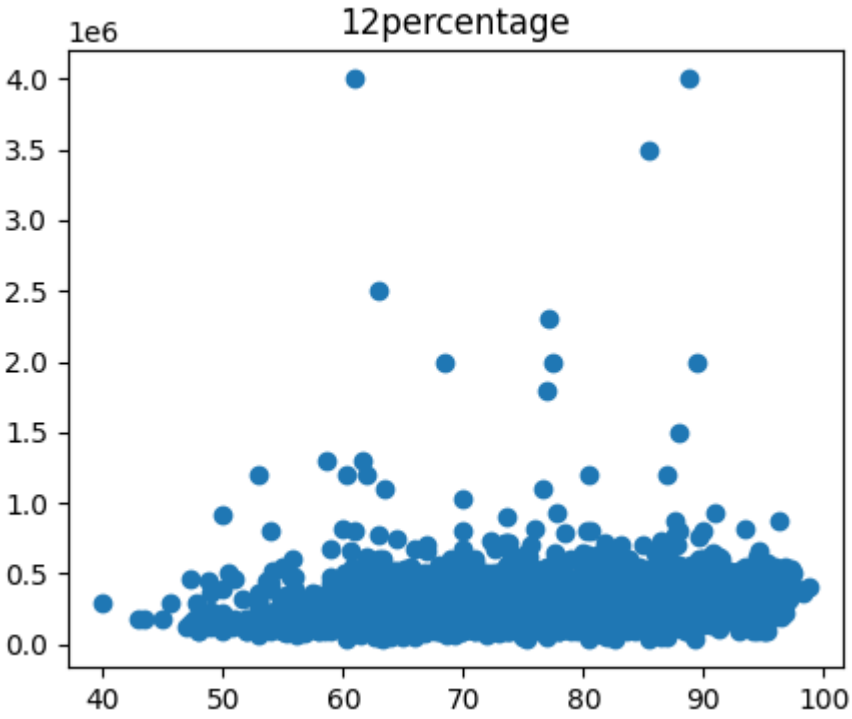
$$Salary = w1 \times (10percentage^2 + 12percentage^2) + w2 \times ComputerScience + w3 \times ComputerProgramming + w4 \times Domain + w5 \times CollegeTier + w6 \times Quant + w7 \times CivilEngg + w8 \times ElectronicsAndSemicon + w9 \times ElectricalEngg$$

Mô hình trên sử dụng 9 đặc trưng.

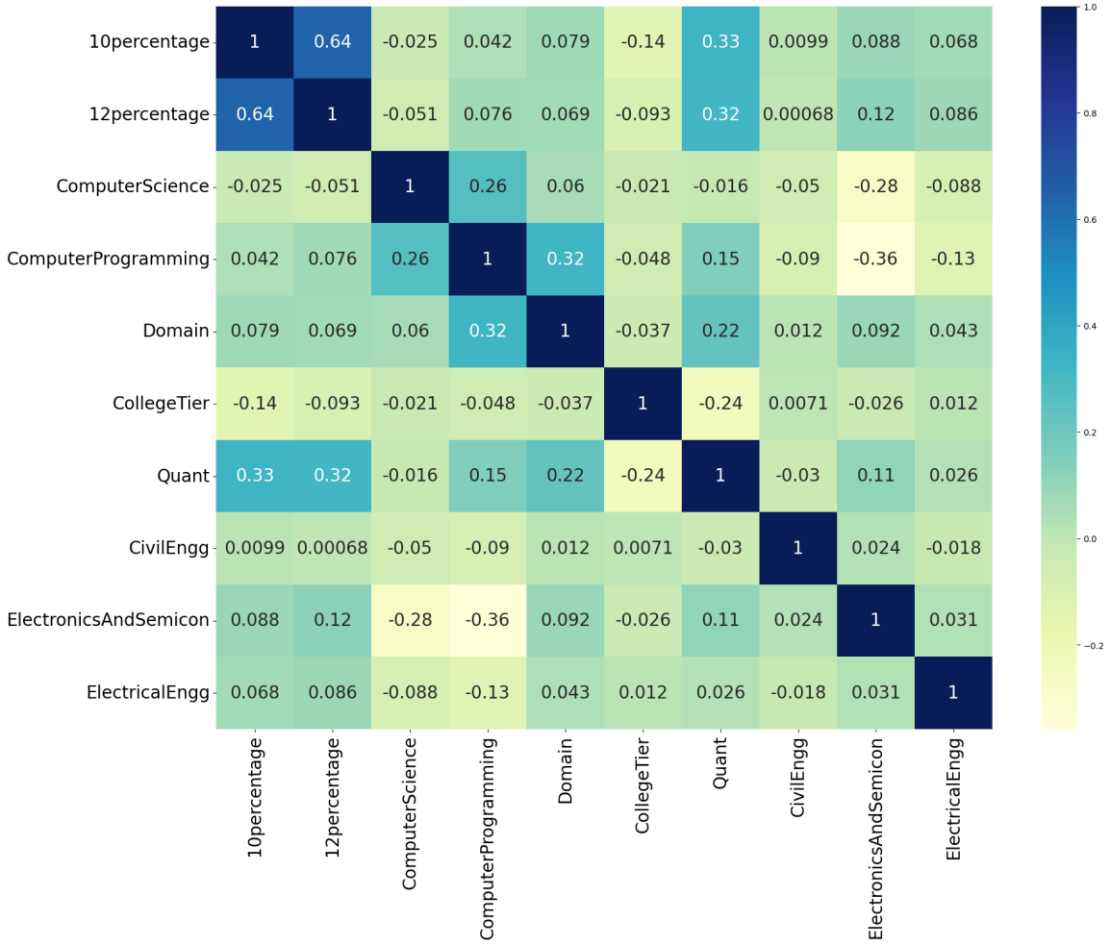
- Trong đó đặc trưng đầu tiên là tổng bình phương của 2 đặc trưng 10percentage và 12percentage.

- Dựa vào biểu 2 đồ điểm của lần lượt 10percentage và 12percentage với salary, có thể thấy khi lấy bình phương thì các điểm sẽ có hình dạng gần với đường hồi quy hơn.





- Ngoài ra, giữa 2 đặc trưng 10percentage và 12percentage cũng có sự tương quan khá cao, được thể hiện bằng hệ số tương quan và biểu đồ điểm sau đây:



- Các đặc trưng về kĩ năng như ComputerProgramming, ComputerScience, CivilEngg, ElectricalEngg cũng thuộc hàng top những đặc trưng ảnh hưởng nhất đến mức lương. Trong 1 bài viết học thuật về những lĩnh vực ảnh hưởng nhất đến mức lương **[1]** có khảo sát các về mức lương như sau:

degree	average income
Computer Science	$256,539
Computer and Information Sciences	$253,305
Computer Science	$247,552
Computer and Information Sciences	$246,946
Computer Engineering	$227,172
Computer Science	$218,525
Finance and Financial Management Services	$206,646
Computer and Information Sciences	$203,685
Electrical, Electronics and Communications Engineering	$202,911
Computer Science	$200,950

- Đặc trưng về ElectronicsAndSemicon cũng có sự liên hệ nhất định đến mức lương. Trong một báo cáo về ngành Semiconductor Industry (ngành công nghiệp chất bán dẫn) **[2]**, lương trung bình của kĩ sư làm về lĩnh vực này dao động từ $51.000 đến $110.000 một năm. Trong năm 2021, có đến 34.000 công việc về lĩnh vực này bị thiếu nhân sự trên nhiều công ty.
- Đối với đặc trưng CollegeTier, những người tốt nghiệp MBA từ các trường đại học hạng 1 có mức lương khởi điểm gấp bốn lần so với những người tốt nghiệp MBA từ các trường đại học hạng 3, điều này được đề cập trong bài viết học thuật về những sinh viên tốt nghiệp từ những trường hạng 1 so với các trường hạng 3 **[3].**

6. QUÁ TRÌNH XÂY DỰNG m MÔ HÌNH

Mô hình 1:

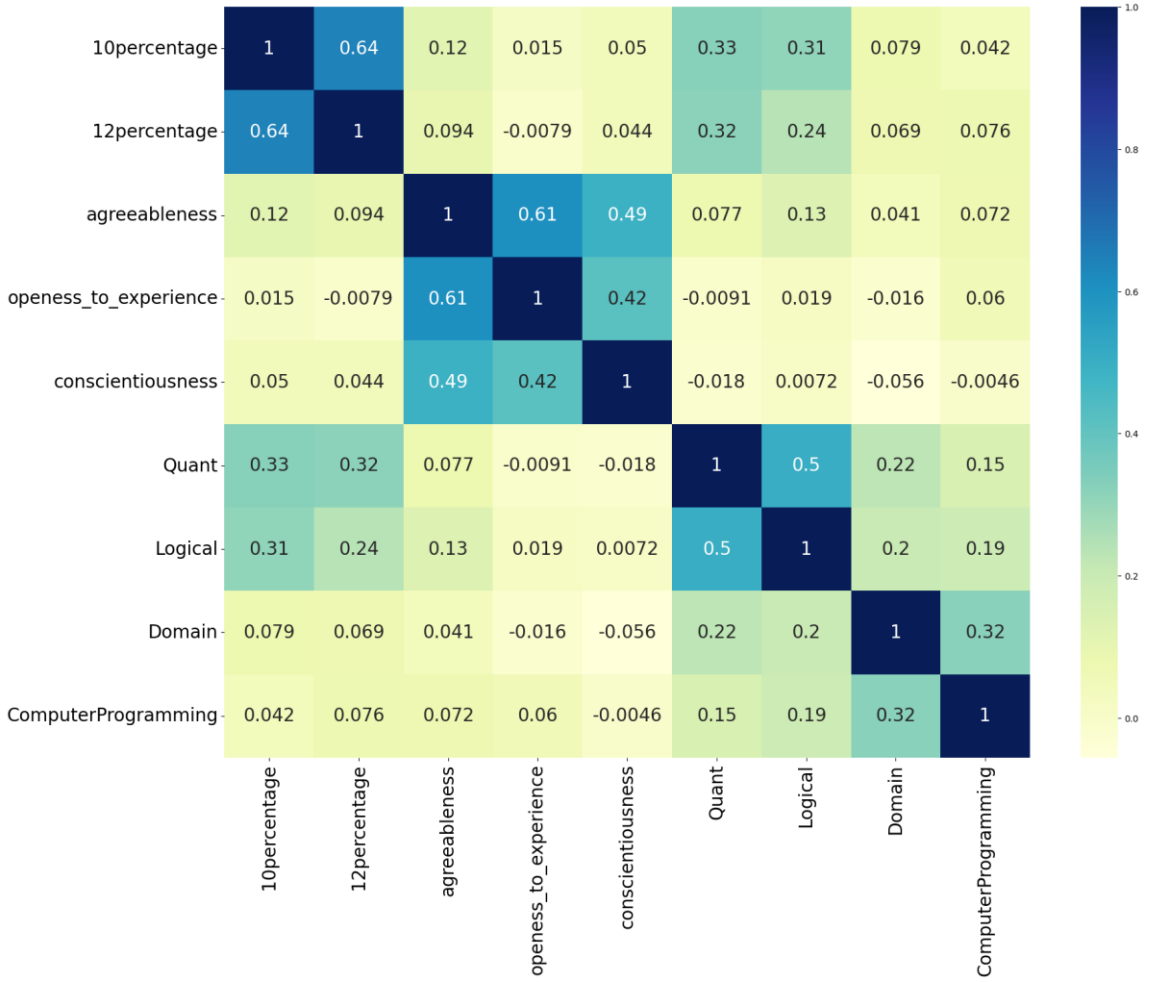
$$Salary = w1 \times ComputerScience + w2 \times collegeGPA + w3 \times ComputerProgramming + w4 \times Domain + w5 \times CollegeTier + w6 \times 10percentage$$

- Các kĩ năng như computer science, computer programming đều có sự ảnh hưởng lớn đến lĩnh vực kĩ thuật và công nghệ thông tin.
- Đối với Computer Science, cục Thống kê Lao động Hoa Kỳ dự đoán rằng kinh tế sẽ thêm 682,800 công việc mới vào lĩnh vực máy tính và công nghệ thông tin từ năm 2021 đến năm 2031. Điều này tương đương với tăng trưởng 15% trên mức lương cho lĩnh vực này - nhanh hơn nhiều so với tốc độ tăng trưởng trung bình dự kiến là 5% cho tất cả các công việc trên toàn quốc. **[4]**
- Đối với Computer Programming, lương trung bình của các lập trình viên hằng năm là $98.936 được thống kê vào năm 2022. **[4]**
- Khi chỉ số GPA tăng lên 1 đơn vị, mức lương khởi điểm hàng tháng tăng trung bình là 29.6 phần trăm, và mức lương trong năm khảo sát sau 3-5 năm tốt nghiệp (mức lương hiện tại) tăng lên 25 phần trăm.**[5]**

Mô hình 2:

$$Salary = w1 \times (10percentage^2 + 12percentage^2) + w2 \times (agreeableness^2 + openessToExperience^2 + conscientiousness^2) + w3 \times (Logical^2 + Quant^2) + w4 \times (Domain^2 + ComputerProgramming^2)$$

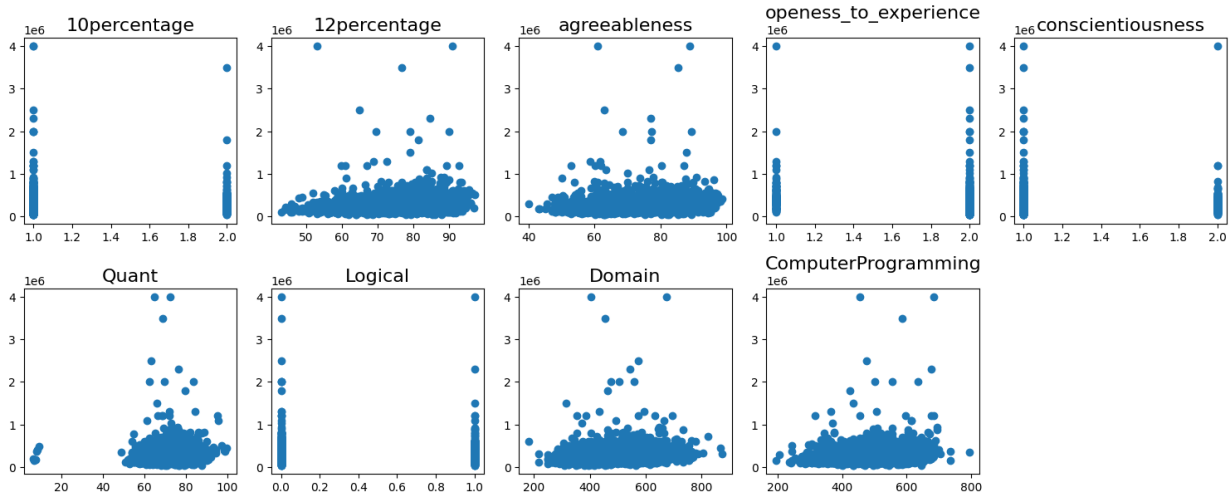
Trong mô hình này, ta xem xét ma trận tương quan giữa các đặc trưng với nhau:



- Từ ma trận tương quan trên, có thể thấy 10percentage và 12percentage có sự tương quan khá cao (0.64). Tương tự với hệ số tương quan của các đặc trưng tính cách (agreeableness, openness to experience, conscientiousness). Hệ số tương quan của Logical và Quant là 0.5, có thể được sử dụng để tạo ra mô hình tốt. Domain và ComputerProgramming cũng có sự liên hệ nhất định với nhau (hệ số tương quan = 0.32).

- Mô hình 2 sử dụng phép bình phương để biến đổi đặc trưng và tạo ra đặc trưng mới bằng cách lấy tổng bình phương của 2 cạnh đối diện.
- Các đặc trưng trong mô hình 2 chủ yếu được chọn từ các hệ số tương quan và biểu đồ điểm của từng đặc trưng với mức lương.

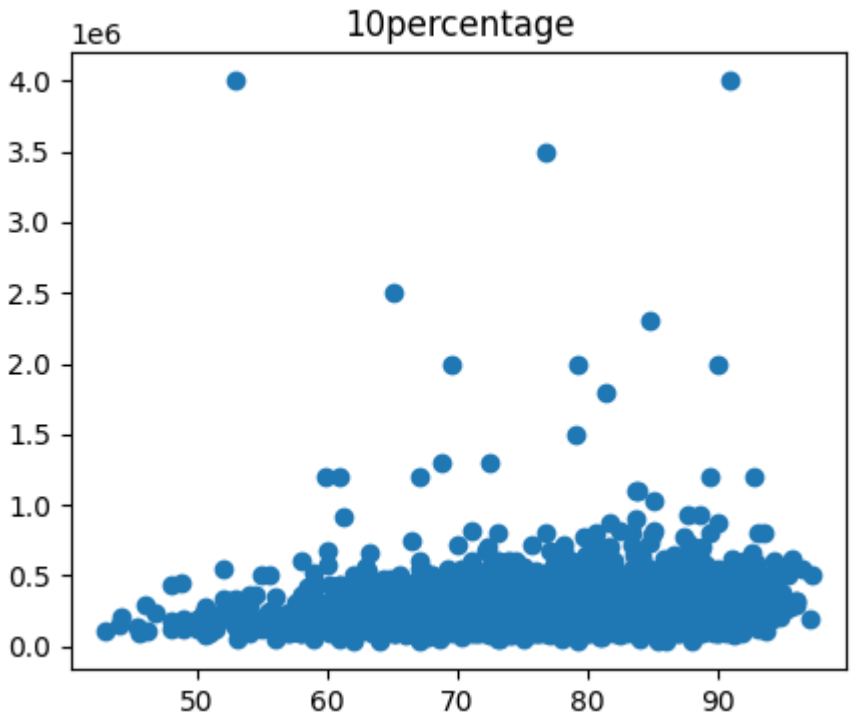
- Các biểu đồ điểm của các đặc trưng kể trên khi lấy bình phương thì các điểm sẽ phân phối về một đường tuy không giống nhưng gần với đường hồi quy.

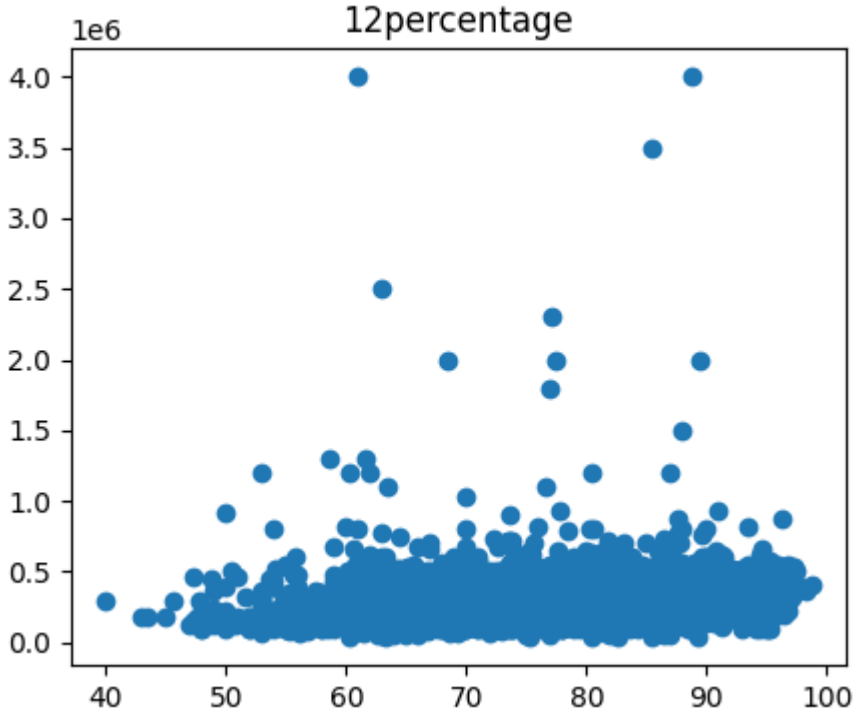


Mô hình 3:

$$Salary = w1 \times (10percentage^2 + 12percentage^2) + w2 \times ComputerScience + w3 \times ComputerProgramming + w4 \times Domain + w5 \times CollegeTier + w6 \times Quant + w7 \times CivilEngg + w8 \times ElectronicsAndSemicon + w9 \times ElectricalEngg$$

- Mô hình 3 được tạo ra với đặc trưng thứ nhất là tổng bình phương của 2 đặc trưng (10percentage và 12percentage) vì biểu đồ điểm của 2 đặc trưng này với lương có dạng gần với đường hồi quy hơn khi bình phương. Lấy tổng 2 đặc trưng này vì hệ số tương quan giữa chúng khá cao (0.64).





- Các đặc trưng khác được chọn lọc dựa vào tính thực tế như sau:
    - ComputerProgramming và ComputerScience là 2 kĩ năng mạnh và được ưu tiên đặc biệt trong ngành engineer.
    - Khả năng định lượng (Quant) là đặc trưng cho ra MAE khá tốt (MAE = 106819.5776 ở yêu cầu 1b) mặc dù mô hình chỉ sử

dụng duy nhất đặc trưng Quant. Đồng thời Quant cũng cho hệ số tương quan khá ổn định với mức lương.

- Những người tốt nghiệp từ trường đại học danh tiếng (CollegeTier) thường có xu hướng nhận công việc với lương cao hơn. **[4]**
- Các kĩ năng engineer còn lại (Civil Engineer, Electrical Engineer, Semiconductor…) cũng đóng vai trò quan trọng với mức lương nhất là trong ngành kĩ sư.

7. REFERENCES

[1]: https://www.cbsnews.com/news/college-major-top-and-lowest-earning-majors-impact-on-income-pay/?fbclid=IwAR1Q9V2f6QwiWocpQVVOX9WlTH1qc0_PTrfZpP0EGCbIaQyOhTmD9rp-eZk

[2]: https://www.theforage.com/blog/careers/semiconductors-good-career-path

[3]: https://economictimes.indiatimes.com/jobs/tier-1-college-tech-graduates-paid-three-times-more-than-tier-3/articleshow/54203223.cms?from=mdr&fbclid=IwAR3eECdSc6fTPv-5GyaZ22FiDOS56TLP0hFHKBKkr8ChLYR6FtWowWTlTCw

[4]: https://www.forbes.com/advisor/education/computer-science-vs-computer-engineering/?fbclid=IwAR0KBYxzDBT6rvfwJX7RzwWpWD8qNITrTaWmnjAxT-qJwr-AYcS98vlhBDw#:~:text=It%20depends%20on%20your%20title,Computer%20hardware%20engineers%20earned%20%24128%2C170

[5]:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9004755/?fbclid=IwAR0wKdlSn45zFj-TOcJKxNxDgmu-5HsWz5Yb6bapzP7MBq-Te2FOkWhK6zQ