**Final test**

*Chọn 1 trong 4 câu để thực hiện project*

<sonhom>\_<hoten>\_baicuoikhoa\_<so de>.ipynb

3\_ledinhhieu\_baicuoikhoa\_4.ipynb

Deadline 23h59 ngày 15/10

**Câu 1:** Bạn có thể dự đoán liệu một bệnh nhân có nên được điều trị bệnh tâm thần của mình hay không theo các giá trị thu được trong tập dữ liệu?

Quy trình như sau:

1. Library and data loading
2. Data cleaning
3. Encoding data
4. Covariance Matrix. Variability comparison between categories of variables
5. Some charts to see data relationship
6. Scaling and fitting
7. Tuning
8. Evaluating models
9. Logistic Regression
10. KNeighbors Classifier
11. Decision Tree Classifier
12. Random Forests
13. Bagging
14. Boosting
15. Success method plot
16. Creating predictions on test set
17. Submission
18. Conclusions

**Data:** [**link**](https://drive.google.com/file/d/1if36X5c6nYxzEJybh0JzvDxXYJzkoeTy/view?usp=sharing)

This dataset contains the following data:

* Timestamp
* Age
* Gender
* Country
* state: If you live in the United States, which state or territory do you live in?
* self\_employed: Are you self-employed?
* family\_history: Do you have a family history of mental illness?
* treatment: Have you sought treatment for a mental health condition?
* work\_interfere: If you have a mental health condition, do you feel that it interferes with your work?
* no\_employees: How many employees does your company or organization have?
* remote\_work: Do you work remotely (outside of an office) at least 50% of the time?
* tech\_company: Is your employer primarily a tech company/organization?
* benefits: Does your employer provide mental health benefits?
* care\_options: Do you know the options for mental health care your employer provides?
* wellness\_program: Has your employer ever discussed mental health as part of an employee wellness program?
* seek\_help: Does your employer provide resources to learn more about mental health issues and how to seek help?
* anonymity: Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
* leave: How easy is it for you to take medical leave for a mental health condition?
* mental*health*consequence: Do you think that discussing a mental health issue with your employer would have negative consequences?
* phys*health*consequence: Do you think that discussing a physical health issue with your employer would have negative consequences?
* coworkers: Would you be willing to discuss a mental health issue with your coworkers?
* supervisor: Would you be willing to discuss a mental health issue with your direct supervisor(s)?
* mental*health*interview: Would you bring up a mental health issue with a potential employer in an interview?
* phys*health*interview: Would you bring up a physical health issue with a potential employer in an interview?
* mental*vs*physical: Do you feel that your employer takes mental health as seriously as physical health?
* obs\_consequence: Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
* comments: Any additional notes or comment

**Câu 2:** Chúng tôi hướng tới việc quản lý Khách hàng dựa trên RFM để công ty có thể nhắm mục tiêu khách hàng của mình một cách hiệu quả.

Quy trình như sau:

Step 1: Reading and Understanding the Data

Step 2: Data Cleansing

Step 3: Data Preparation

Step 4: Model Building K-Means Clustering vs Hierarchical Clustering

Step 5: Final Analysis

**Data:** [**link**](https://drive.google.com/file/d/1buvTsZsnnQ5diefkus0sAa5Jk0qpiOOQ/view?usp=sharing)

**Câu 3:** Sử dụng các nguồn dữ liệu là các hoạt động ném bom trên không và điều kiện thời tiết trong chiến tranh thế giới 2 để dự đoán khi nào các hoạt động ném bom được thực hiện.

* Chúng ta sẽ bắt đầu với mô tả và làm sạch dữ liệu, sau đó chúng tôi sẽ trực quan hóa dữ liệu của mình để hiểu rõ hơn.
* Các quá trình này có thể được gọi là EDA (Phân tích Dữ liệu Khám phá).
* Sau đó, chúng tôi sẽ tập trung vào dự đoán chuỗi thời gian để dự đoán khi nào các hoạt động ném bom được thực hiện.
* Đối với dự đoán chuỗi thời gian, chúng ta có thể sử dụng phương pháp ARIMA sẽ là một hướng dẫn.

Data:

* **Aerial bombing Data description:** [**operation.csv**](https://drive.google.com/file/d/1fSJvVXagMTOuA4WLtN931vDjNeID6QQZ/view?usp=sharing)
  + Mission Date: Date of mission
  + Theater of Operations: Region in which active military operations are in progress; "the army was in the field awaiting action"; Example: "he served in the Vietnam theater for three years"
  + Country: Country that makes mission or operation like USA
  + Air Force: Name or id of air force unity like 5AF
  + Aircraft Series: Model or type of aircraft like B24
  + Callsign: Before bomb attack, message, code, announcement, or tune that is broadcast by radio.
  + Takeoff Base: Takeoff airport name like Ponte Olivo Airfield
  + Takeoff Location: takeoff region Sicily
  + Takeoff Latitude: Latitude of takeoff region
  + Takeoff Longitude: Longitude of takeoff region
  + Target Country: Target country like Germany
  + Target City: Target city like Berlin
  + Target Type: Type of target like city area
  + Target Industry: Target industry like town or urban
  + Target Priority: Target priority like 1 (most)
  + Target Latitude: Latitude of target
  + Target Longitude: Longitude of target
* **Weather Condition data description:**
  + [**Weather station location:**](https://drive.google.com/file/d/1qdz5uyYx95J59V9VKyYjkn7fyHO1EmE-/view?usp=sharing) 
    - WBAN: Weather station number
    - NAME: weather station name
    - STATE/COUNTRY ID: acronym of countries
    - Latitude: Latitude of weather station
    - Longitude: Longitude of weather station
  + [**Weather:**](https://drive.google.com/file/d/1q4e8SThcrApGGeQDw5g9FJxKtylotD_V/view?usp=sharing)
    - STA: weather station number (WBAN)
    - Date: Date of temperature measurement
    - MeanTemp: Mean temperature

**Câu 4:** Đề mở, tự crawl dữ liệu và chọn đề tài.

Thang điểm:

* Crawl dữ liệu: 10%
* EDA, visualize, preprocessing: 30%
* Feature extraction: 30%
* Training and evaluation: 20%
* Result: 10%