According to World Health Organisation (WHO), every year around 17.9 million deaths are due to cardiovascular diseases (CVDs) predisposing CVD becoming the leading cause of death globally. CVDs are a group of disorders of the heart and blood vessels, if left untreated it may cause heart attack. Heart attack occurs due to the presence of obstruction of blood flow into the heart. The presence of blockage may be due to the accumulation of fat, cholesterol, and other substances. Despite treatment has improved over the years and most CVD's pathophysiology have been elucidated, heart attack can still be fatal.

Thus, clinicians believe that prevention of heart attack is always better than curing it. After many years of research, scientists and clinicians discovered that, the probability of one's getting heart attack can be determined by analysing the patient's age, gender, exercise induced angina, number of major vessels, chest pain indication, resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, and maximum heart rate achieved.

Therefore, your task as a data analyst cum machine learning engineer, you need to develop an app to predict the chance of a person having heart attack.

Below are the criteria of the project:

1) Develop a model using **only one machine learning** approach (knn, random forest, regression, etc).
2) Prepare two scripts. One for model training, the other script will be used for model deployment and app development.
3) **Perform necessary steps** during EDA and **justify** the taken steps. You may write your justification as comments in your script.
4) Achieve model with validation accuracy of more than 70%.
5) Deploy the model and preform prediction on a web app developed using **Streamlit**

Test case

So once your app is developed, the clinicians decided to test the app you developed. The table below contains the patient's condition. Key in the data into your model to see if your model able make prediction same as the true output. Then compute the accuracy in percentage (%).

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	True output
65	1	3	142	220	1	0	158	0	2.3	1	0	1	1
61	1	0	140	207	0	0	138	1	1.9	2	1	3	0
45	0	1	128	204	0	0	172	0	1.4	2	0	2	1
40	0	1	125	307	0	1	162	0	0	2	0	2	1
48	1	2	132	254	0	1	180	0	0	2	0	2	1
41	1	0	108	165	0	0	115	1	2	1	0	3	0
36	0	2	121	214	0	1	168	0	0	2	0	2	1
45	1	0	111	198	0	0	176	0	0	2	1	2	0
57	1	0	155	271	0	0	112	1	0.8	2	0	3	0
69	1	2	179	273	1	0	151	1	1.6	1	0	3	0

Files to be submitted and uploaded to GitHub and LMS (submission link will be given on the assessment day):

1) Both training and deployment scripts (GitHub and LMS)
2) Dataset (.csv file) (GitHub and LMS)
3) Saved model and scalers (if any) in .pkl file format. (GitHub and LMS)
4) A screenshot of your developed app should be saved as .png file format and zipped in a folder together with the rest of the files for LMS submission. Also include the screenshot in README.md and display on your GitHub repo (GitHub and LMS)
5) Performance of the model and the reports can be snipped and saved as image file to be included in the zip folder for LMS submission. (LMS and GitHub)
6) Include your GitHub URL directing to your assessment 1 in a text file then submit to LMS. (LMS)
7) **Don't forget to credit/cite the source of the data on your GitHub page** https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset

*Please zip all the required files into one folder then submit to LMS.

**Please save the dataset and model in 2 different folders to GitHub.

Complete the assessment and submit the files to LMS and GitHub by 5pm. Good Luck!!!

	100%	50%	0%
Task Completion (30%)	Scripts can be executed without any error on trainer's local machine.	-	Scripts fail to be executed on trainer's local machine.
Project requirements (30%)	Able to achieve the objectives of the project using relevant and appropriate approach.	Able to achieve the objectives of the project but using inappropriate approach such as brute forcing the solution.	Fail to achieve the objectives of the project.
Exploratory data analysis (30%)	Demonstrates strong understanding on the objectives of the project and performs relevant approach to process the data. Necessary data processing techniques such as, data loading, data cleaning, features selection and data preprocessing are performed and well justified.	Shows comprehensive understanding of the objectives of the project but uses incorrect or irrelevant approach to process the data. For example, removing NaN data when there are limited number of samples in the dataset.	Shows limited understanding of the objectives of the project. Absence of data processing section in the code.
Code readability (5%)	Involves the usage of functions or methods for repeated tasks. Codes are easily readable and justified by including comments and description texts.	Minimal usage of functions or methods for repeated tasks. Available comments and descriptions but lack of details.	No usage of functions or methods for repeated tasks. Codes are difficult to read and understand. Missing descriptions and comments.
GitHub repo (4%)	Detailed and clear instructions of the project on README.md. Results such as graphs are also included in README.md as part of the project description.	Project successfully uploaded to GitHub repo but with incomplete README.md, Missing descriptions, instructions, and results.	Fails to upload project to GitHub repo and missing README.md
PEP8 compliance (1%)	Fully complies with PEP 8 Standard	Partially complies with PEP 8 Standard	Fails to comply with PEP 8 Standard
Total (100%)			