***Group 3***:- *Report 6*

**CSE523 Machine Learning**

**Heart Attack Prediction**

| **Group Members** | **Roll No** |
| --- | --- |
| **Shivam Thakker** | **AU1940193** |
| **Devarsh Sheth** | **AU1940189** |
| **Pranav Gandhi** | **AU1940313** |
| **Meet Jhaveri** | **AU1940284** |

* 1st we found the common features in all 3 datasets and then looked out for the missing values and understood the other terms that were different in the other datasets.
* The feature values were quite differently written in the .csv file, so we had to make the values same to combine all the 3 files.
* We performed data cleaning and modeled the data, by sorting, filtering and replacing some values.
* After combining the data we had to check whether any rows were duplicated or not, because if rows are repeated then the model may overfit when applied.
* So in excel we ran a COUNTIFS function to check out the repeated rows.

**=COUNTIFS(criteria\_range1, criteria1, [criteria\_range2, criteria2], ...)**For reference we checked out a YouTube video:

<https://www.youtube.com/watch?v=Ihkgs7T3Do0>

* It turned out that out of 1571 rows 217 rows repeated, so we removed the repeated rows and now we will work further on the data.

# THANK YOU!!