University Of St. Thomas

# A purple shield with white letters Description automatically generated

# Artificial Intelligence

# Assignment – 6

## Project Title - Face Recognition

# Team Members:

Su, Duyen

Lim, Yeeyin

Verma, Pragya

Dixit, Mugdha

Patil, Priyanka Namdev

1. Project Title

* **Face Recognition**

1. Description of data source and web link(s)

* The "Face Recognition Dataset" was downloaded from UMASS website. Below given is the link to the dataset.
* https://vis-www.cs.umass.edu/lfw/index.html

1. # of records & # of attributes with description of each attribute

* Number of records: This dataset contains 49 images in each folder, and there are a total of 31 folders. The dataset consists of 1519(31×49) records or instances, which is 8MB of data.
* Also, as this is a collection of images there is no attribute for this dataset.

1. Some general statistics of the dataset with few sample records (or images or text)

* **General Statistics:**

Total Dataset Size: 8MB

Number of Folders: 31

Images per Folder: 49

Total Records: 1519

* **Following are all the folders available:**

A screenshot of a computer screen

Description automatically generated A screenshot of a computer

Description automatically generated

* **Sample of images in a folder:**

A collage of a person's face

Description automatically generated

1. Tools / methods / computing platform(s) your team (plan to) use in your study

* We are using Python as our programming language, and as there are lots of images, model will need high computing power hence, we will be using ‘Google Colab’ for the model training and prediction.

1. Detail description of what problems/questions your team plans to predict / study.

* Our primary objective in building the model is to prioritize precision and accuracy, ensuring the correct identification of individuals. Given that images of the same celebrity vary in angles, facial hair, accessories like goggles and glasses, and age progression, our model must account for these diverse scenarios. As the dataset includes a mix of images accuracy becomes paramount in our focus.
* Even the dataset was unbalanced, however the model we built leveraging the balance since the model predicts whether 2 images are identifying the same person by using different parameters.
* A well-trained model holds the potential to expand our scope to the development of an attendance system applicable in educational institutions and workplaces. Such an application could be used in tracking attendance processes, offering efficiency and convenience to administrators and employees alike.