A black and grey logo

Description automatically generated

**Group**

* Dubai CW PG Thursday Group 6

**Group Members**

* Mohamed Aman
* Faizan Watare
* Fardeen Khan
* Pratibha Yadubanshi

**F21DL Coursework Part 1 – Data Analysis and Bayes Nets**

**Executive Summary**

In this report, we present the results and findings from Part 1 of our assignment, focusing on data analysis and the use of Naïve Bayes classifiers on the dataset. This part of the assignment allowed us to gain insights into the dataset, explore the performance of the Naïve Bayes classifier, and conduct experiments based on the lectures from week 1 – 5.

**Exploratory Data Analysis (EDA)**

During the data preprocessing phase, we found the following insights in our dataset:

* The dataset consists of a total of 9690 images, each represented by 2304 features (48x48 pixels), emphasizing the size and dimensionality of the dataset.
* The dataset doesn’t have any null values.
* The dataset consists of 10 distinct classes. However, this is an imbalanced dataset, with class distributions showing significant disproportions between classes.
* This was found out as a result data visualization by matplotlib and seaborn libraries.

**Normalization**

* We used MinMax Scaler to scale the pixel values of the dataset between 0 and 1. This ensures that all features (pixels) are on a common scale.

This meticulous data preprocessing phase provided the foundation upon which our subsequent analysis and modeling were constructed. It highlighted the dataset's scale, dimensionality, and class imbalances, all of which informed our approach and decision-making in the subsequent stages of our assignment.