Title: De-Identification of Facial Images and Tabular Data to Preserve Data Privacy using L-Diversity

Problem Statement: In recent times, much software MNCs have implemented facial recognition software for services like surveillance, photo-tagging, maintaining identity records etc. So, the question has arisen whether our identity is safe from leaks of sensitive data and are the companies following proper safety guidelines. Image recognition has become an integral part of many organization's operations such as Facebook implementing tagging people in uploaded pics, government installing surveillance cameras, CCTV cameras in banks and other institutions, etc. Thus, there is a need to protect this info from untrusted sources by privatizing the data. With this comes the need of preventing hackers/adversaries from accessing sensitive information. In this project, our aim is to implement one such method of anonymizing facial images using de-identification techniques.

Motivation: In an age of widespread data collecting and sharing, the safeguarding of people's sensitive information has become critical. Facial photos and tabular data frequently contain personal information that, if revealed, can lead to identity theft, discrimination, and other types of harm. This study is driven by the urgent need to create effective de-identification algorithms that might limit these hazards while enabling the responsible use of data for research, AI applications, and collaborative analysis. This research intends to adopt L-diversity not only to fulfill legislative requirements but also to sustain ethical norms, retain public confidence, and strike a balance between data value and privacy. Ultimately, the goal is to enable enterprises to reap the benefits of data-driven insights while preserving individuals' privacy rights in an increasingly linked world.

Project Outcome: There are many methods available to implement de-identification of facial images such as blurring out the image, hiding certain facial features or modifying the facial features by adding some noise, calculated by taking out average values of certain facial features such as skin- color, shape etc. SAP HANA Cloud, which provides a safe and scalable platform for storing and retrieving massive volumes of data, may be utilized to store and manage these anonymized photos, which we would be implementing in this project.