# INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

## Dharwad, Karnataka- 580009


**A Project Report On**

# "PRIVACY PRESERVING LEARNING SYSTEM"

Submitted by

**M Joshasree (19BCS069)**

**P Lalithaanjale (19BCS087)**

**S Pranay Sai Teja (19BCS102)**

**M Anupama (19BCS123)**

Under the Guidance of

**Dr. Malay Kumar**

Asst. Professor, Dept. of CSE

# CERTIFICATE

This is to certify that the Project Work entitled— **"PRIVACY PRESERVING LEARNING SYSTEM"** is a bonafide work carried out by M.Joshasree(19BCS069), P.Lalithaanjale(19BCS087), S.Pranay Sai Teja(19BCS102), M.Anupama(19BCS123) in fulfillment for the Mini Project of Bachelor of Technology in Computer Science & Engineering of the Indian Institute of Information Technology Dharwad during the year 2021-2022. The Project Report has been approved as it satisfies the academics prescribed for the Bachelor of Technology degree.

**Signature of Supervisor(s)**

**Name(s)**

**Department(s)**

**(Month, Year)**

# DECLARATION

We declare that this written submission represents my ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

M. Joshasree
19bcs069

P. Lalithaanjale
19bcs087

S. Pranay Sai Teja
19bcs102

M. Anupama
19bcs123

# ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of a large number of individuals who have been responsible for the successful completion of this project.

It is our privilege to express our sincerest regards to our project coordinator, **Dr. Malay Kumar**, Asst. Professor, Department of Computer Science & Engineering, Indian Institute of Information Technology Dharwad, for his valuable inputs, able guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of our project.

We take this opportunity to thank all our lecturers who have directly or indirectly helped our project. We pay our respects and love to our parents and all other family members and friends for their love and encouragement throughout our career. Last but not the least we express our thanks to our friends for their cooperation and support.

# ABSTRACT

The motive of this project is to apply Differential Privacy concept for the dataset (containing information about demographic, behavioral and medical risk factors of individuals with a predicting factor of future risk of coronary heart disease(CHD)) and applying Machine learning Classification Algorithm to the dataset to predict the future risk of CHD.

We can use Differential Privacy for many applications such as Machine Learning, Deep learning, Database Management system etc.The project basically starts with the analysis, preprocessing, normalization of the dataset and followed by development of a machine learning classification algorithm by two ways namely non-private machine learning algorithm and private machine learning algorithm(using pipeline) along with the assessment of advantages and disadvantages respectively. The ultimate goal is to implement a privacy preserving learning system. So we choose the Differential privacy concept in Machine learning in order to implement an application regarding privacy preserving learning system. The designed machine learning model helps to overcome the disadvantage of the linkage attacks on datasets,Membership Inference Attacks(MIA) etc.

## CONTENTS

# INTRODUCTION

## 1.1 Overview

**Privacy-Preserving Machine Learning** is a step-by-step approach to avoiding data breaches in machine learning projects. We can use techniques like compressive privacy, differential privacy and synthetic data generation to improve your privacy. ML engineers of all skill levels and superiors will benefit from implementing these privacy-preserving methods into their model construction process.

Face recognition, cloud data storage, and other complex privacy-enhancing technologies are demystified through real-world use scenarios. We face a slew of current and future machine learning privacy issues, but we also have a variety of methods to address them. We'll build a machine learning system that protects user privacy without sacrificing data quality or model performance in this project.

Anonymization process is usually done on the servers(and big datasets) of the companies that collect our data(including personal and confidential data).But, we exactly don’t know to which extent we can trust this anonymization process as there are some examples where this anonymization didn’t work.

For example,In 2006, Netflix started a competition in which teams had to create an algorithm that could predict how an individual would rate a movie.
To help with this, Netflix provided a dataset of 100 M ratings, submitted by almost 480 thousand users. Netflix anonymized the dataset by removing the names of individuals and replacing some ratings with fake and random ones. However, this seems to be anonymous but it isn’t. Later in 2008, University of Texas published a paper that said they had successfully found the individuals and their ratings by combining the dataset with IMDb dataset. These are called linkage attacks.To avoid such types of vulnerabilities, the Differential Privacy concept comes into picture.

## 1.2 Classical definition of Differential Privacy

In the context of statistical and machine learning analysis, differential privacy (DP) is a strong, mathematical definition of privacy. DP is a privacy protection requirement, according to this mathematical concept, that various methods for evaluating sensitive personal information have been designed to meet. In short, Differential privacy mathematically guarantees that anyone viewing the result of a differentially private analysis will essentially make the same inference about any individual's private information, whether or not that individual's private information is included in the input to the analysis.

DP can be achieved by introducing a minimum distraction(noise) in the information given by the database. The introduced distraction is immense enough that it is capable of protecting privacy and at the same time limited enough so that providing information to analysts is still useful.The key idea or implementation of the concept relies on **"Blurring the data"**.

### 1.2.1 Illustration

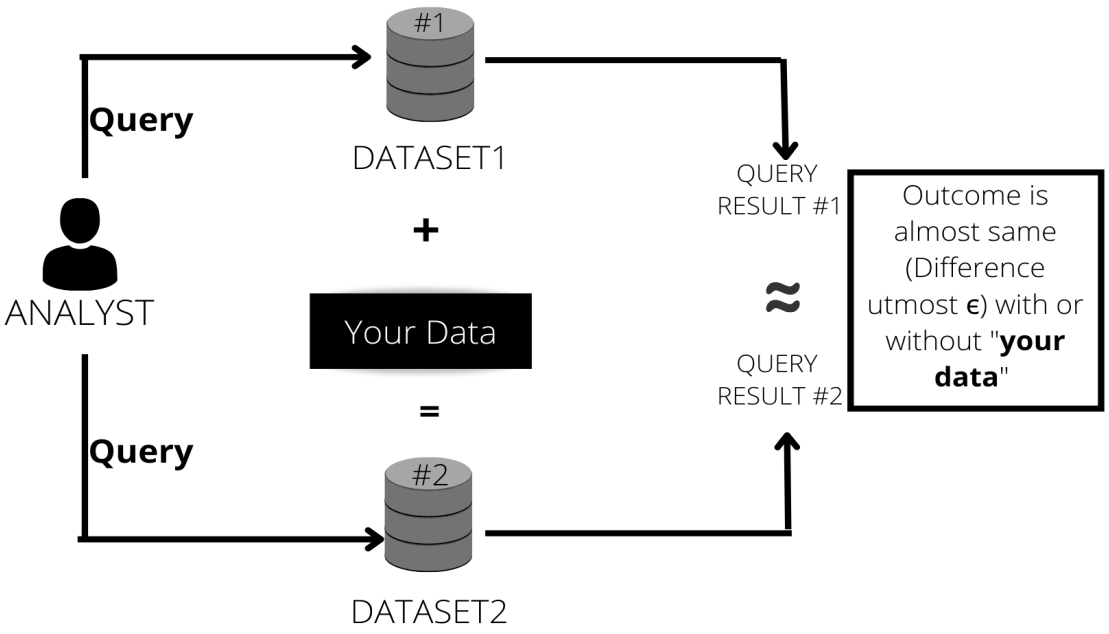



*Figure 1.1*

### 1.2.2 Use-case diagram

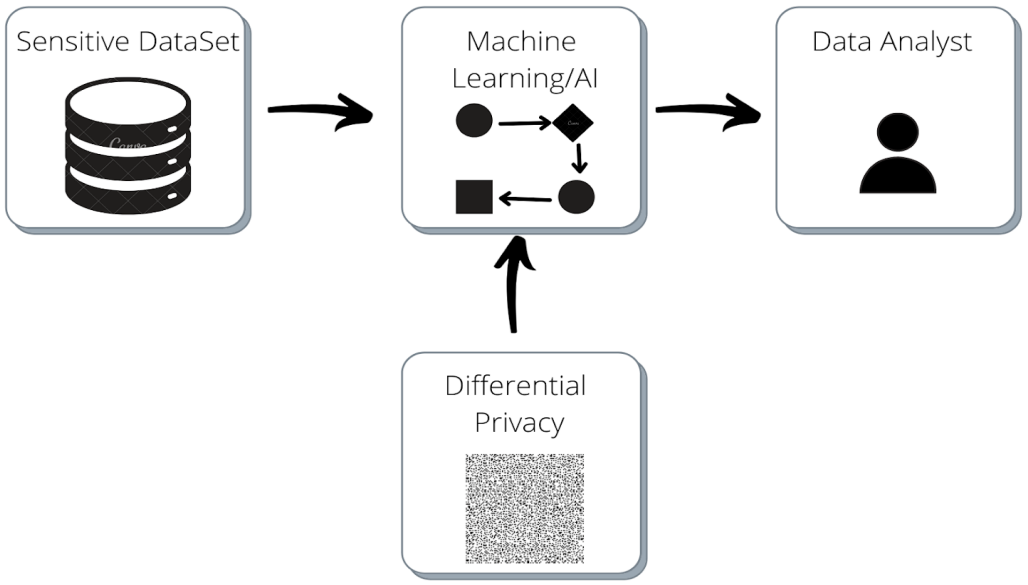



*Figure 1.2*

## 1.3 Background of Differential privacy

### 1.3.1 Mathematical definition

Let's say we have two datasets D and D′, which differ by utmost a single record/row. In addition, we consider a randomized mechanism/algorithm M[·] that operates on the datasets to gain a required output. This mechanism is differentially private if the results of M[D] and M[D′] are almost identical for every choice of D and D′.

More precisely, a mechanism M[·] is $\epsilon$-differentially private if for all subsets of output S $\subset$ Range[M] and datasets D and D′,

$$\mathbf{Pr(M[D] \in S) \leq exp[\epsilon] * Pr(M[D'] \in S)}$$

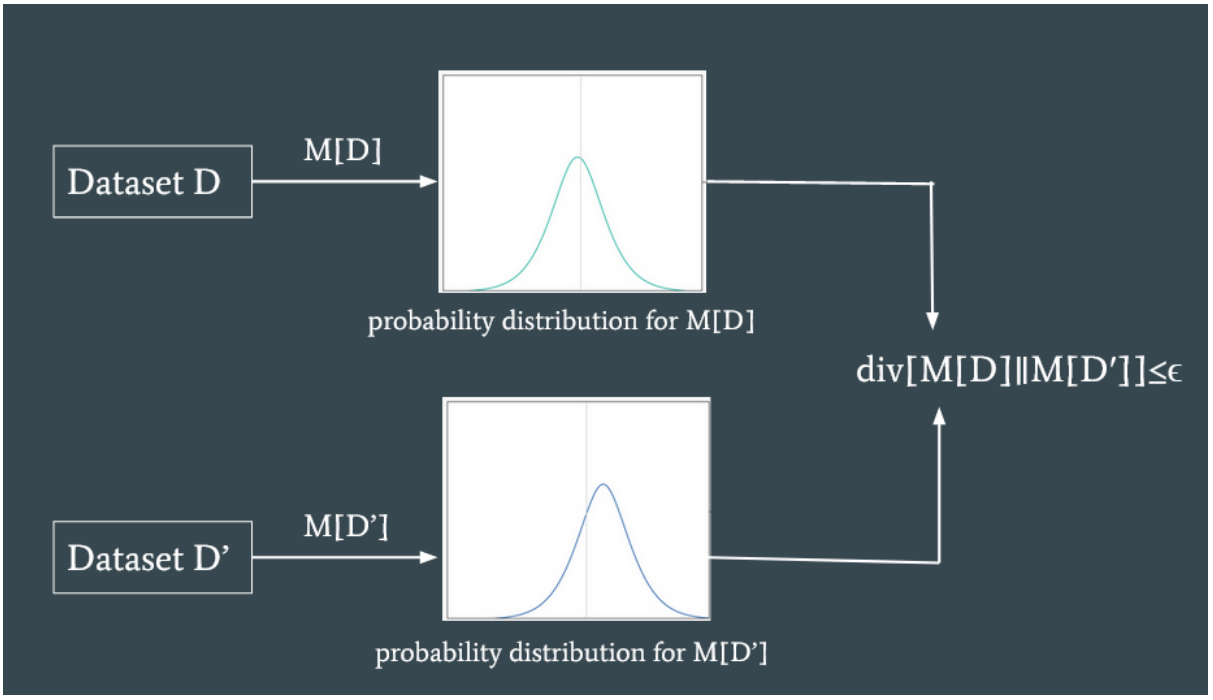
### 1.3.2 Definition in terms of Divergence

Divergence is defined as the measure of the difference between the probability

distributions.Since the mechanism M[·] is randomized, there is a probability distribution over its output. The mechanism is ϵ-differentially private if and only if

$$\mathbf{div[M[D] \,/\!/\, M[D'\ ]]\leq\epsilon}$$

for datasets D and D′ differing by at most a single record. In other words, ϵ quantifies how large the divergence can be between the distributions of results when the mechanism is applied to two neighboring datasets.




### 1.3.3 Mechanisms

Differentially private algorithms are randomized algorithms with noise introduced at important moments. The Laplace mechanism is one of the simplest algorithms, and it may post-process aggregate query responses (e.g., counts, sums, and means) to make them differentially private.

#### 1.3.3.1 Laplace mechanism

Let f[·] be a deterministic function of a dataset D which returns a scalar value. For instance, it might count the number of rows that satisfy a condition. The Laplace mechanism works by adding noise to f[·]:

$$M[D]=f[D]+\acute{\epsilon}$$

where $\acute{\epsilon} \sim Lap_{\acute{\epsilon}}[b]$,
$b=\Delta f/\epsilon$,
And L1 sensitivity,

$$\Delta f=\max_{D,D'} \| f[D]-f[D'] \| 1$$

For classification purpose, we are using logistic regression algorithm in diffprivlib python library. In the pseudo code for logistic regression, It uses the mechanism named "Vector" which adds a Laplace-distributed random vector to the objective.

The naïve pseudocode implementation of the Laplace mechanism looks like this :

**Input:** Function *F*, Input *X*, Privacy level $\epsilon$

**Output:** A Noisy Answer

Compute *F (X)*

Compute sensitivity of of *F : S*

Draw noise *L* from a Laplace distribution with variance:

$$2 * \left(\frac{S}{\epsilon}\right)^2$$

**Return** *F(X) + L*

Below is example Java code for the Laplace mechanism specific to count queries :

```
import org.apache.commons.math3.distribution.LaplaceDistribution;
double laplaceMechanismCount(long realCountResult, double epsilon) {
        LaplaceDistribution ld = new LaplaceDistribution(0, 1 / epsilon);
        double noise = ld.sample();
        return realCountResult + noise;
}
```

## 1.4 Graphical representation with a range of epsilons

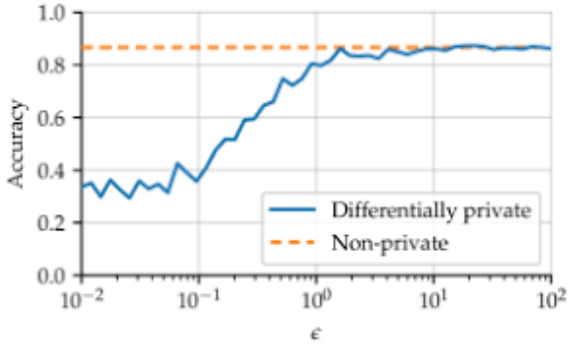



*Figure 1.3*

### 1.5 Problem statement and description

To preserve privacy of Individuals participated in the dataset, we are using the Differential Privacy concept for training the model followed by constructing a machine learning classification model (using Logistic regression) to predict the future risk of coronary heart disease.

### 1.6 Objectives

- Finding an appropriate dataset to do the prediction.
- The data is pre-processed, with redundant and extraneous information removed.
- A non-private mechanism is used to construct an optimized model that uses Logistic Regression as a classifier. Then, with the help of the diffprivlib inbuilt library, a private machine learning model is constructed.
- Demonstrating and comparing the results.

### 1.7 Motivation

Machine learning algorithms use their training data to extract knowledge. When dealing with personal or sensitive data, the model may expose information that is specific to a single data point. In the realm of natural language processing, a notable example of data leakage was recently reported . A neural network could memorize a random secret word placed into its training dataset and recreate the word using only the trained model. Similarly, an auto completion model may learn to guess a user's password based on their username each time it is entered. This problem has ramifications in other sectors, including health, census records, and finance, because it prevents companies from sharing trained models. So, to avoid such threats we use differential privacy for training the model and to preserve privacy.

# LITERATURE SURVEY

In this section we will study about various references which illustrate the basics, advantages, challenges, mathematics, approaches, mechanisms of Differential privacy.

1. **The algorithmic foundations of differential privacy. C Dwork, A Roth - Found. Trends Theor. Comput. Sci., 2014 - tau.ac.il**. The above mentioned paper helped us a lot in knowing the facts about Differential privacy, its advantages and challenges, its mathematical intuition and different types of mechanisms of Differential privacy.

2. **Privacy-preserving logistic regression. K Chaudhuri, C Monteleoni - Advances in neural networks, 2008 - proceedings.neurips.c.** From the above article we studied about the important tradeoff between privacy and learnability,when designing algorithms for learning from private databases.

3. **Mengmeng Yang, Lingjuan Lyu, Jun Zhao, Tianqing Zhu, Kwok-Yan Lam** in their paper **Local Differential Privacy and Its Applications: A Comprehensive Survey** explained about the concept of local differential privacy. They also distinguished between local and centralized differential privacy.

4. **Naoise Holohan,Stefano Braghin,P´ol Mac Aonghusa,Killian Levacher** in their paper titled **Diffprivlib: The IBM Differential Privacy Library - A general purpose, open source Python library for differential privacy,** they demonstrated the IBM Differential Privacy Library, an open source, general-purpose library for studying, experimenting, and implementing differential privacy applications in Python. The library includes a variety of techniques, which are the foundations of differential privacy, as well as a number of machine learning and other data analytics applications.
   The library's development prioritized simplicity and accessibility, making it ideal for a wide range of users, from those doing their first inquiries into data privacy to privacy professionals wishing to contribute their own models and processes for others to utilize.

5. The book **Programming with Differential Privacy by Joseph P. Near and Chiké Abuah** is helpful in understanding the python programming techniques of various mechanisms of Differential privacy and various types of sensitivities.

## FLOW CHART

A flowchart is a diagrammatic representation of an algorithm, workflow or process. Flow chart depicts the steps in approaching a solution for the problem stated. The flowchart for the solution we are proposing to the stated problem is as follows:

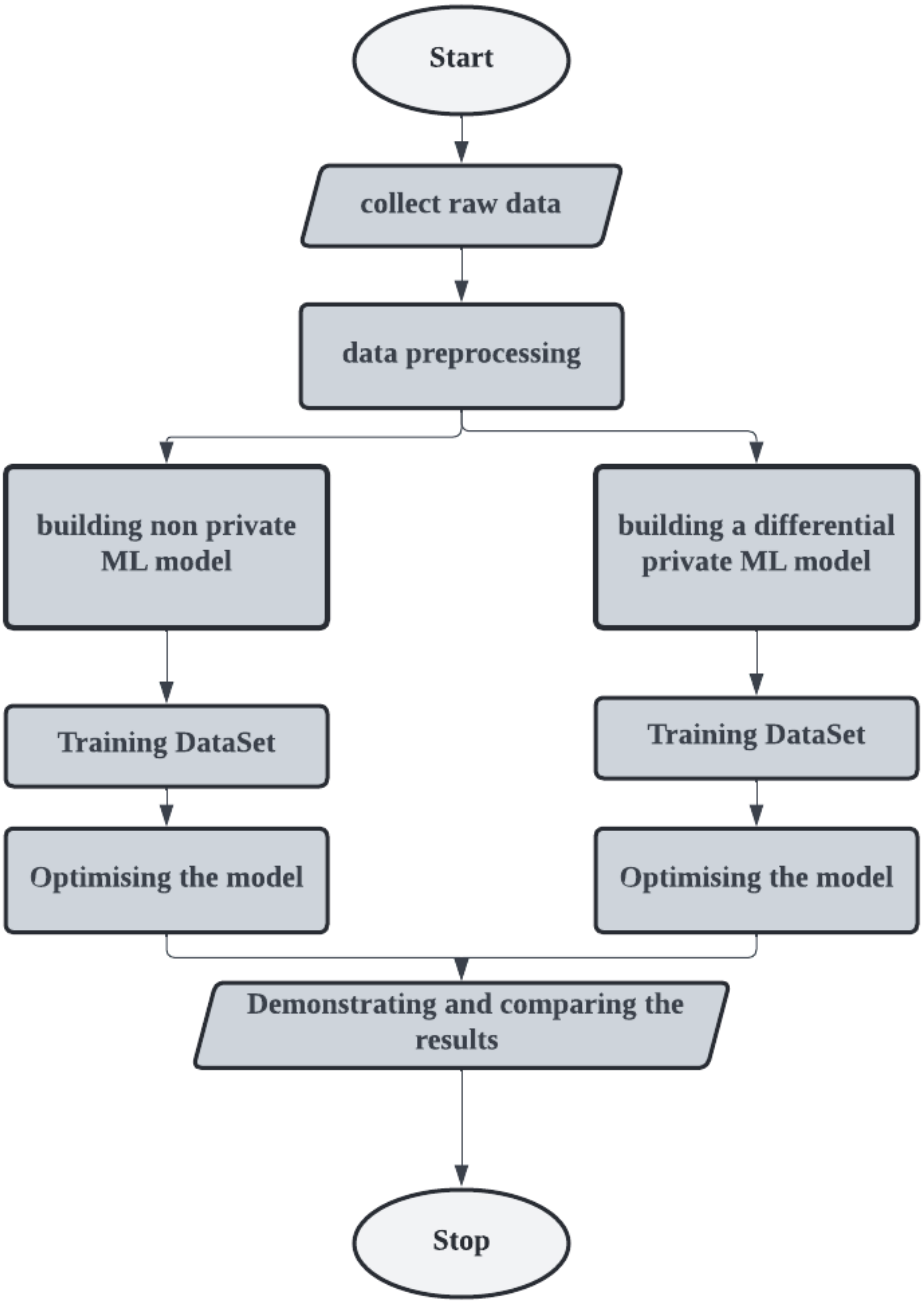



*Figure 3.1*

# IMPLEMENTATION

## 4.1 Introduction

Implementation is the manifestation of an application or the execution of a plan, idea, model, design, specification, standard, method, or policy. In other words, an implementation is the programming and deployment of a technical specification or technique as a programme, software component, or other computer system. For a particular specification or standard, there may be numerous implementations.

## 4.2 Implementation of Non-private Machine learning model

Implementation of normal non-private Machine learning model includes data collection, data analysis, data preprocessing, training and testing the model.We are using a dataset that contains information about demographic, behavioral and medical risk factors of individuals with a predicting factor of future risk of coronary heart disease(CHD).To predict/classify the result we are using logistic regression.

## 4.3 Implementation of Differentially private Machine learning model

To implement the Differential privacy, we are using machine learning pipeline and Python inbuilt library named “diffprivlib”. Data collection, data analysis, data preprocessing, training and testing the model will be included as same as in the non private machine learning model. We should then define the parameters like bounds, datanorm, epsilon to gain differential privacy. And for classification purpose we are using  Logistic regression algorithm.

## 4.4 Programming language - python 3.9.5

The Python programming language was utilized to complete this project. It's a high-level programming language that lets you write interactive, interpreted object-oriented scripts. It uses several English terms rather than programming keywords because it was designed to be human accessible. Code readability is emphasized in the design philosophy. As a result, it's ideal for creating huge programmes. Python has an automatic memory management system and a dynamic tape system. It offers a wide number of extensive structured libraries that support numerous programming paradigms, including object-oriented programming, functional programming, and procedural programming.

## 4.5 Libraries used

Several pre-trained libraries were used during the project's development to make developing the models easier. The following are some of the most well-known libraries.

### 4.5.1 Sklearn

In Python, Scikit-learn (Sklearn) is the most usable and robust machine learning library. It uses a Python consistency interface to give a set of efficient tools for machine learning and statistical modeling, such as classification, regression, clustering, and dimensionality reduction. NumPy, SciPy, and Matplotlib are the foundations of this package, which is mostly written in Python.

### 4.5.2 Diffprivlib

Diffprivlib is a general-purpose python library for experimenting, researching, and implementing differential privacy applications.The IBM Differential Privacy Library (also known as diffprivlib) is written in Python 3, a prominent machine learning and data analysis computer language. Diffprivlib takes advantage of the NumPy (WCV11,) and Scikit-learn (PVG11,) packages' capabilities and familiarity, making functions and models readily recognisable, with default parameters providing accessibility for everybody. diffprivlib is free to use and modify, and users are encouraged to contribute to the library's functionality and features by using the MIT Open Source license.Diffprivlib contains a wide number of methods that handle the addition of noise. These mechanisms are the core building blocks of differential privacy. To meet differential privacy, these strategies are used behind the scenes in machine learning models and other tools, allowing the complicated process of obtaining differential privacy to be hidden from view.

### 4.5.3 NumPy

Numpy is a software library mainly used for Array Vectorization. It is implemented using the Python programming language. It supports large multidimensional arrays and matrices. It also includes many Mathematical functions to operate on single and multidimensional arrays. Numpy was originally written as Numeric and was created by Jim Huguin with support from several other Developers. It is known to be an ancestor of Numpy. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with

extensive modifications. NumPy is open-source software and has many contributors.

#### 4.5.4 Pandas

Pandas is majorly used in applications where Data is extracted in the form of frames. Our applications take input in the form of video feed and individual images need to be extracted from the video. Hence Pandas was the best choice for this operation. It processes data in various formats that includes csv, excel and several other file formats. The Pandas library is implemented using Python programming language and it allows several data manipulation operations such as Group By, join, merge, melt, concatenation as well as data cleaning features such as filling, replacing or imputing null values

## RESULTS

We will be using the same dataset for both non-private and differentially private machine learning models. First the non-private ML model will be designed classically and the accuracy is calculated. Then the dataset then undergoes differential privacy while training and then the accuracy is calculated after fitting the model. Finally demonstration and comparison of two models will be carried out.

## CHALLENGES

DP was originally proposed for interactive statistical queries to a database. A randomized query function M (that returns the query answer plus some noise) satisfies $\epsilon$-DP if for all datasets D1 and D2 that differ in one record and all $S \subset Range[M]$, it holds that $Pr(M[D1] \in S) \leq \exp[\epsilon] * Pr(M[D2] \in S)$. In other words, the presence or absence of any single record must not be noticeable from the query answers, up to an exponential factor of $\epsilon$. The smaller $\epsilon$, the higher the protection. The noise to be added to the answer to enforce a certain $\epsilon$ depends on the global sensitivity of the query to the presence or absence of any single record. Mild noise may suffice for statistical queries such as the mean, while a very large noise is needed for identity queries returning the contents of a specific record.

The crucial part in developing differential privacy is to specify the value of $\epsilon$.Using High values doesn’t assure best privacy and hence there is much privacy loss compared to the cases taking small epsilon values which are near to zero.

In addition to distorting model updates, utilizing DP poses the following issues:

- Sequential composition applies because model modifications are protected in each epoch and are computed on the same (or, at least, not totally disconnected) client data in subsequent epochs. This indicates that as the number of epochs increases, the effective epsilon reduces exponentially, and the effective protection decreases exponentially. As a result, the only models that are substantially usable are for epsilon values that are meaninglessly enormous] (such as 50-100).
- A data set with each record containing the response of a different respondent is assumed in the original definition of DP. Then DP makes sure that no single respondent's record is visible in the DP-protected output that has been released. This safeguards the privacy of each individual respondent. When DP is used to safeguard a client's model update, however, all entries in the client's data set are owned by the client. When all records in a client's private data set are about the client, such as when the client's private data contains her health-related or fitness measurements, making any single record undetectable is insufficient to safeguard the client's privacy. As a result, in this scenario, the DP guarantee is of no use.

## CONCLUSION AND FUTURE SCOPE

Machine learning's goal is to extract relevant information from data while maintaining privacy by concealing information. As a result, it appears that reconciling these competing interests will be difficult. When mining sensitive data, however, they must frequently be balanced. Medical research, for example, is a critical application that requires both the extraction of meaningful data and the protection of patient privacy. Extracting generic features of entire populations without compromising individuals' private information is one technique to overcome the dilemma.

Differential privacy, one of the most prominent and powerful definitions of privacy, is the focus of this project. We look at how machine learning and differential privacy interact, specifically privacy-preserving machine learning techniques.

Our proposed project idea will be helpful in case of visualizing the comparison between private and non-private machine learning models.

Applying cryptographic approaches to DP, building efficient strategies for DP implementation in real life, and changing privacy budgets to test how effective it is are some of the future work possibilities. To summarize, differential privacy is still a topic that needs further investigation. Many obstacles remain, but with more effort, they may be overcome.

# REFERENCES

## 8.1 Reference papers

[1] The Algorithmic Foundations of Differential Privacy by Cynthia Dwork, Aaron Roth,2014.

[2] A Case Study on Differential Privacy by Bihil SELESHI, Samrawit ASSEFFA, June 2017.

[3] the IBM Differential Privacy Library by IBM.

[4] Preserving User Privacy for Machine Learning: Local Differential Privacy or Federated Machine Learning? by Huadi Zheng, Haibo Hu, Ziyang Han, Jul 2020.

[5] Differential Privacy in the Wild:A Tutorial on Current Practices & Open Challenges by Ashwin Machanavajjhala, Xi He, Michael Hay, 2016.

[6] Achieving Differential Privacy and Fairness in Logistic Regression by Depeng Xu, shuhan Yuan, Xintao Wu, 2019.

[7] Privacy-Preserving Learning Analytics: Challenges and Techniques by Mehmet Emre Gursoy, Ali Inan, Mehmet Ercan Nergiz and Yucel Saygin.

[8] Differential privacy in practice: Expose your epsilons! by Cynthia Dwork, Nitin Kohli, and deirdre mulliga, October 2019.