ISSN No:-2456-2165

Preserving and Randomizing Data Responses in Web Application using Differential Privacy

Rubia Fathima
REVA Academy for Corporate
Excellence REVA University
Bengaluru, India

Abstract:- The Data reflects Humans as most of the data is online. Privacy for individual's data has become a prime concern. Securing and preserving that privacy has been a focus for long decades. In data analytics, machine learning and data science analysis over users' private data utilize to understand the individual user's responses to present data publicly. The issue with private data presented online that consisted of personal private data was sensitive and confidential was a significant issue, so a particular group of mathematicians and cryptographers came together to resolve this issue by introducing the concept of Differential Privacy.

Differential privacy is assurance over information privacy without damaging the chance of having privacy risk by including some amount of Random noise in the form of robust data to the original dataset. Differential privacy is also a tool with an algorithm that helps maintain Privacy by Preserving and Randomizing data responses—measuring the accuracy of statistical data by performing analysis. To Perform this process of differential privacy, IBM developed an open-sourced algorithm called Diffprivlib[1]**. With this library, the project has created a Front-End Web application that can perform data analysis that involves different mechanisms, models, and Tools.**

This project is an attempt to integrate all mechanisms, models, and tools involved in DiffPrivLib[1]**. The primary purpose of this paper is to showcase the work on differential privacy that consists in developing a user-friendly web application that can be open-sourced. This application is designed in a python programming package and will experiment with the dataset to perform the analysis to show the impact of differential privacy algorithms on different values on epsilon with accuracy and privacy.**

Keywords:- *Differential Privacy, Python Programming, Open-Source Library, Data Science, Machine Learning, Data Analytics.*

I. INTRODUCTION

As exposure of electronic data over the internet has become specific, detailed, and abundant, maintaining individual privacy has become the top priority. So, securing these personal private data in the datasets has been processed in mathematical computation involved in differential privacy algorithms.

Individual user's Data that involves any personal private information requires differential privacy to be applied. The most popular definition of privacy is differential privacy, coined as the new mathematical term came in 2004 in data privacy. It ensures publicly visible data does not make any changes for a single individual if there are changes in the dataset. It resolves issue by adding random noise to the mechanism at work. The need for an increase in adding robustness in the form of noise, maintain meaning full data pattern, the mathematically stringent definition of privacy, and computation that would be rich in a class of algorithms that satisfy the definition of differential privacy. [1]

Preserving Privacy and Security guarantees the data pattern has been spinning around the compliance process and to perform powerful collection and curation of data to apply appropriate policies for user's private data in any form

Differential Privacy Guarantees:

1. The raw data holding individual responses will not be unauthorized access and does not need to be modified.
2. Maintaining an individual's privacy will be valued over mining important details from data.
3. Manage resilience to post-processing; output generated from the secret differential algorithm will not affect the differential privacy of the algorithm. In other words, the data analyst that does not have additional knowledge about the dataset cannot increase privacy loss by looking and thinking at the output of the Differentially Private algorithm.[2]

The infield of research study on Machine Learning the Scope has been growing, and flooding like a tidal surge, the in-depth analysis in data privacy standard has emerged to differential privacy algorithms in the subjects of Cryptography and Security. IBM has presented great work by creating the Differential Privacy library, a general purposed and open-sourced library for investigating, experimenting, and developing differential privacy applications in a python programming language called "Diffprivlib." The library includes all host mechanisms, the building blocks of differential privacy[1], and several other applications to experiment in Machine learning and other data analytics tasks. This project demonstrates the idea of differential privacy in a web application that can be useful for any data analyst or accountant to perform analysis in the form of mathematical computations for the supervised dataset. This application solves

ISSN No:-2456-2165

the problem of having cost-effective results related to privacy as an open-source web application.

A. Scope of Work

The main scope of the work is to develop a front-end web application developed using digital technologies such as AI/ML, Cryptography[3], and Security. The study and experiment related to this project can help the enterprise in the following manner

1. The Web application developed for IBM's Differential Privacy library having covered mechanism, model, and tools.
2. Application developed can perform analysis that is free and cost-effective
3. "Diffprivlib" provides an extensive collection of mechanisms, the fundamental building blocks of differential privacy that handle the addition of noise. The Parameter used to set this value is denoted as epsilon (ε) to the dataset; ε controls so much noise or randomness to a raw dataset.
4. As the application is available opensource can be utilized to perform experiments by small-size companies to high-level companies
5. Helps non-technical person or provides ease of work for a data analyst who can perform different computations and understand the data
6. Experiments were performed only on the supervised dataset.
7. For Accountant, Data analysts can perform computations to understand if there is any Privacy Leakage.

Differential privacy is becoming a significant factor to business because:

1. It helps businesses analyze leakage and maintain data privacy that can comply with GDPR and CCPA without determining the ability to understand their customer behavior. While not complying with regulations can result in a severe and heavy number of fines. As per the 2021 recent report[4] from international law firm DLA Piper's cybersecurity and data protection team has been infringing with €273 million of fines for not complying with GDPR since May 2018. These fines are a drop in the bucket if they start considering the level of GDPR compliance and the European economy's size. The below-given figure shows the total value of GDPR fines imposed from 25th May 2018 to 2020 appears to have increased as per the country's check for GDPR compliance with more comprehensive and automated approaches.[4] The below-given Figure1 shows the Total values of GDPR fines.

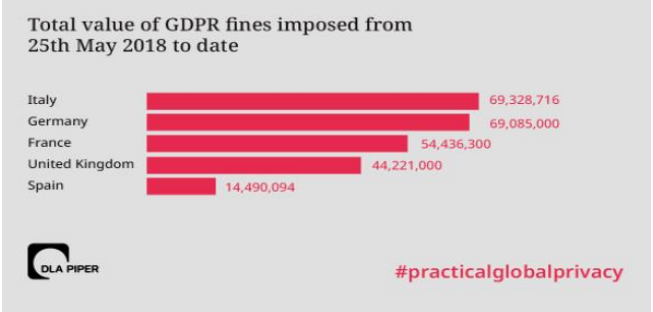


Fig 1: *Total value of GDPR Fines Imposed since May 2018*[4]

2. Data privacy violations such as breaches also damage the reputation of business "2020 Cost of a Data Breach Report" by IBM[4] states that lost business due to diminished reputation was the most significant cost factor of a data breach with a yearly average of $1.52 million. The report also states that customer's personally identifiable information was the costliest data type to compromise in the data breach they studied.
3. Differential privacy enables trade to share their information securely with other organizations to collaborate with them without gambling their customer's privacy
4. It is resistant to privacy attacks based on supplementary information to prevent possible linking attacks on de-identified data.

B. Abbreviations and Acronyms

Abbreviation	Long Form
GDPR	General Data Protection Regulation
CCPA	California Consumer Privacy Act
HIPAA	Health Insurance Portability & Accountability Act
GLBA	Gramm-Leach-Bliley Act
PPA	precise poverty alleviation
PATE	Private Aggregation of Teacher Ensembles
NIST	National Institute of Standards and Technology
DP	Differential Privacy
AI	Artificial Intelligence
ML	Machine Learning
SQL	Structured Query Language
PII	Personally, Identifiable Information

II. LITERATURE SURVEY

The project aims to have deep analysis in research for differential Privacy has become a de facto standard to maintain and guarantee privacy as it has become one of the hot research topics in the field of healthcare especially. From privacy engineering, there have been many new programs held to understand the performance of the differential privacy tool in Machine Learning[5] and Data Science[6]. The Review of background theory and existing literature, tools from renowned journals/technical reports/websites showcases the limitations and advantages of those tools and previous work by few researchers that must support the problem formulation on the deployment of differential privacy. NIST has a list of de-identification tools in the ranking challenge from created and during the Privacy Engineering program[7]; each application has research papers that explained limitations in the market. There are around 23 de-identification tools in the list of NIST. Most open-sourced tools are applied with limited edition and paid versions come with fully functional modules but are not cost-effective.

Hence, the problem stated above from the literature survey explains the limitations and drawbacks of few tools; therefore, the problem statement that this project is trying to explain is addressing the same and enable in the future with an open-source web application from IBM called Diffprivlib. This tool is by a non-IT person that performs administrative work or general staff to generate the output for getting privacy in data, perform automatic computations on non-Zero count,

Mean, Variance and Standard Deviation, Histogram and train model to get values that any generalist has understood in the form of a graph.

Based on the above literature study, **it is clear** that this project can address whitespace with a survey on different De-identification tools performed by NIST[7]. The available tools have fully functional differential Privacy computations with paid versions; otherwise, if the application is open source, the service is limited to have used only a limited mechanism involved. This project efficiently implements all available tools, models, and mechanisms in Diffprivlib with a lightweight python framework on the web application. The non-IT person with minimum data science knowledge can utilize the tool to perform differential privacy computations and mechanisms. While looking at other tools in the market, there may be a precedence but maintaining resilience and sustainability.

III. THE OBJECTIVE OF THE STUDY

The purpose of the capstone project is to have a well-integrated user-friendly web application that is high in resiliency and sustainability using IBM's Diffprivlib library, that can perform computations based on digital technologies such as AI/ML, Cryptography, and Security. The objectives of this work are below

1. The primary objective of the differentially private algorithm is to measure privacy and reduce risk.
2. The secondary objective is to investigate different datasets and experiment with developed web applications in differential privacy.
3. Understand the importance of mathematical noise to the data can help preserve an individual's privacy and confidentiality.
4. Enables organizations to customize the privacy policy to prevent attacks like Linkage and reconstruct the original data.
5. The user's learning dataset will have to improve with Observation and understanding of the parameters of IBM's open-sourced differential privacy library along with machine learning and other data analytics task.
6. With differential privacy, algorithm parameter Epsilon (ε) changes the accuracy and increases the dependency of Epsilon ε over privacy.

The novelty of this project is to provide the computational algorithm that performs sampling of the dataset so that original data can be preserved with post-processing and retained in final results. The tool that the originality of preserves data.

From the above Scope and objectives, an open-source library that develops will support many mechanisms, models, and tools related to Machine learning. The main point to be considered is applying the algorithm so that data privacy maintains the same original dataset without changing patterns. Making data robust by adding some noise can be controlled by Epsilon and data measures with accuracy. Calculates non-Zero count to understand the truthfulness of the data for mathematical computations like Mean, Standard Deviation, Variance of the dataset.

From the research and literature review on the available tool for Differential Privacy, the inference is that most of the DP application is just data masking and anonymization tools but to preserve and maintain data accuracy without making considerable changes epsilon. Randomizing responses[8] guarantees the privacy of the data and protects it from different attacks like Reconstruction attacks, Linkage attacks, etc. The main aim of data curator/creator is to develop front-end applications and perform the proper computation for precise datasets.

IV. PROJECT METHODOLOGY

The project was developed based on SDLC phases and followed by Data Science or Machine Learning lifecycle, and the methodology contains six different points that followed a Waterfall model approach. Below the given diagram, Figure 2 has shown project methodology.

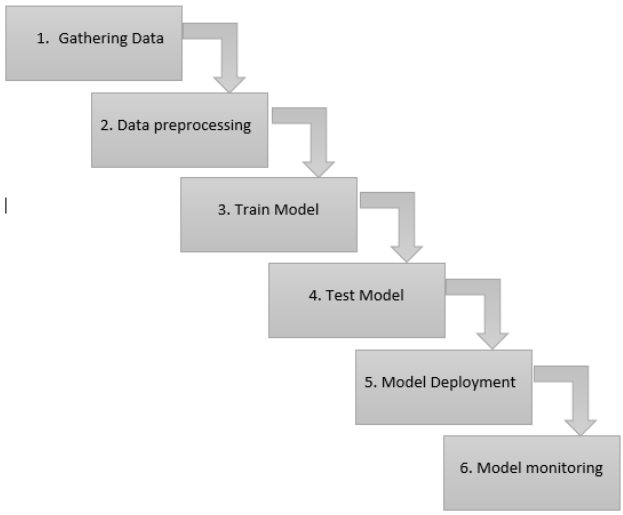


Fig 2: Project Methodology

A. *Process Steps*

1.

2. **Gathering Data**: During the first phase, the dataset is captured by uploading the dataset

3. **Data Preprocessing**: The data is analyzed to understand the use case; in the case of Differential privacy, the most looked into is identifying the critical column with personal private data of the user or data owner. They are here preprocessing the data for feature extraction, feature analysis, and feature visualization.

4. **Train Model**: For the Diffprivlib[2] Library models that classification into supervised and unsupervised learning; the below-given table has provided a sort between different types of models, for open-sourced Web application has been implemented with integrated differential privacy and currently with limited to supervised learning and only classification model GaussianNB, Linear regression, and Logistic Regression is applied. It is the responsibility of the user to choose the appropriate model, keeping the uploaded dataset in mind.

5. **Test Model**: The selected model has to follow a specific mechanism included in Diffprivlib[2] Library with Web application has captured the Analysis for supervised dataset having Classification models such as GaussianNB, Linear regression, and Logistic Regression. With each use case, test a model by choosing a model among GaussianNB, Linear regression, and Logistic Regression and then choose the column to that column has to be target and display.

6. **Model Deployment**: The application is deployed on a localhost python-based server using version 3.Code is also available on private GitHub account It is of a lightweight Python-based Framework called Flask. Diffprivlib[2] by default comes with functionality and familiarity of NumPy [17] and Scikit-learn [14] packages, meaning functions and models are instantly recognizable, with default parameters ensuring accessibility for all. The web application is with python libraries such as NumPy, Scikit-learn, and SciPy packages.

7. **Model Monitoring**: The Web application over differential privacy algorithm involves data visualization that performs and outputs a Mean, Variance, and standard deviation over the dataset. At the same time, it can also perform a non-zero count between Non-Private and Private data to measure the truthfulness of the algorithm.

Overall, the concept of the project is to manage and perform the different privacy mechanisms involved in Diffprivlib Library[9]. Still, while this project performs only for the supervised dataset for classification and regression model, in future analysis, this project will improve the deploying mechanism for unsupervised raw data to perform using the K-means algorithm.[10]

V. RESOURCE REQUIREMENT SPECIFICATION

The below-given Figure 3 shows diagram explains the process utilized for developing the Robust Differential privacy data post-processing

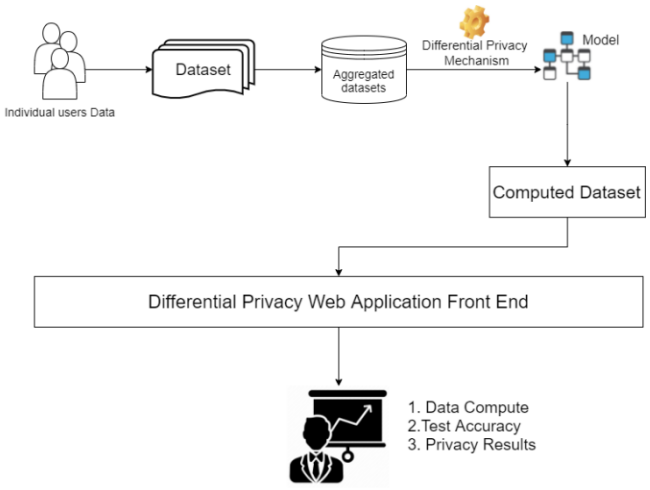


Fig 3: *Process flow and modules in Differential Privacy*

A. Data Components:Diffprivlib is of four major components

1. **Mechanisms**: These are the building block of differential privacy uses in all models that actualize differential privacy. Mechanisms have little or no default settings for use by experts implementing their models. They can be that as it may be utilized exterior models for partitioned examinations, etc.[11]
2. **Models**: This module includes machine learning models with differential privacy. Diffprivlib currently has models for clustering, classification, regression, dimensionality reduction, and preprocessing.
3. **Tools**: Diffprivlib comes with several nonexclusive devices for differentially private information analysis. Data incorporates differentially private histograms, taking after the same arrange as NumPy's histogram function.[12]
4. **Accountant**: The BudgetAccountant class can track security budget and calculate add up to security misfortune utilizing progressed composition techniques.

The function flow diagram in Figure 4 explains the process utilized for developing the robustness in data using a Differentially privacy algorithm in the below-given chart.

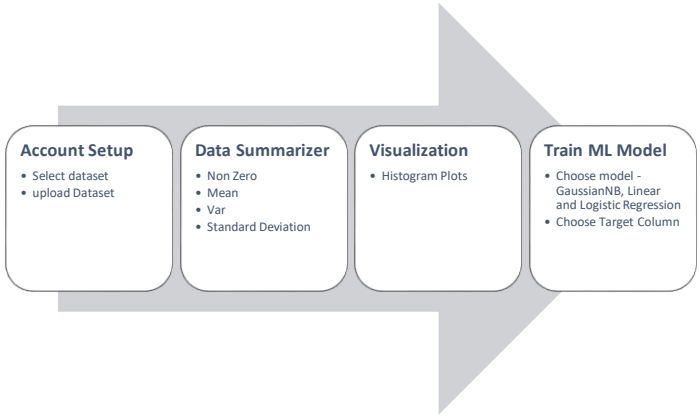


Fig 4: *Data Flow diagram for DP web Application*

B. Process Overview Steps

1.**Account Setup**: Select the account to set up a temp session for a data analyst or accountant. Setup budget account
1.Select the data set that would perform sampling
2.preview from the data selected
3.select the personal private data that is to mask
4.Upload-Epsilon value by Choosing epsilon Value

2.**Data summarizer**: Display the data of Private and Non-Private and perform comparative analysis method and display in the form of a graph having checked with accuracy

1. **Non-Zero Count**: This shows the number of non-zero elements for each column in the dataset. The screenshot in the next chapter clearly shows the change in the data distribution of select columns of the dataset involved in the differential private methods.
2. **Mean**: This shows the relative change in mean values of different columns between the non-private/standard method and differentially private counterpart.
3. **Variance**: This shows the relative change in variance values of different columns between the non-private/standard method and differentially private counterpart.

4. **Standard Deviation**: This shows the relative change in traditional deviation values of different columns between the non-private/usual method and differentially private counterpart.

3. Visualization - The histogram plots of both non-private/normal plotting functions and the Differential Private plotting methods.

4. Train ML Model - Apply machine learning principles on the dataset to train the model and obtain a comparative report of the performance of non-private/standard and the unique private ML models. Training a model to analyze the test dataset

a) **Choose Model:** This option is to choose one of the many models implemented with integrated differential privacy.

Currently, the below models:
1.Gaussian Naive Bayes Model
2.Logistic Regression
3.Linear Regression

The above given includes both classification models a well as regression models. It is the responsibility of the user to choose the appropriate model, keeping the uploaded dataset in mind.

b) **Choose the Target column**. This option uses to set the dependent or the Y-column involved in the process of training models.

The process is limited to have performed computations over a supervised dataset to generate the results in the form of a Graph that can present data to check accuracy over privacy loss.

VI. SOFTWARE DESIGN

This project follows Flask for Python-based architecture for software design. For high-level architectural diagram is given below in Figure 5

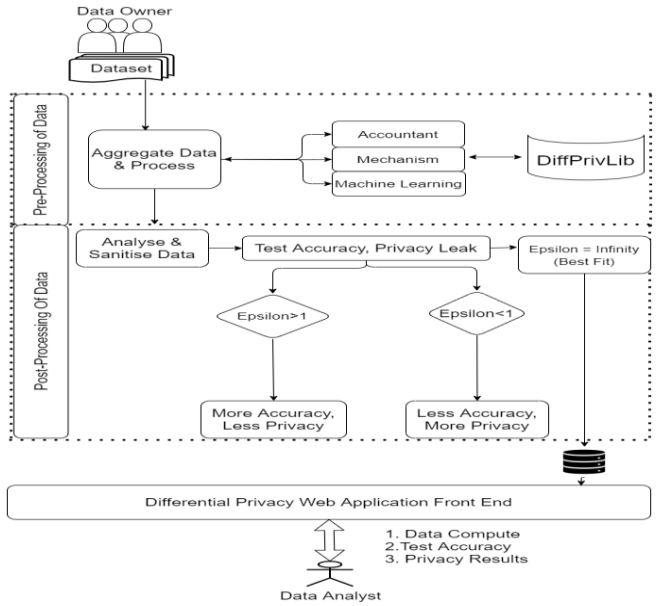


Fig 5: Architectural diagram for Differential privacy using an algorithm

The Software architecture of Python programming, above given architecture, performs the analysis divided into parts for ease of understanding.

On Back End

1. **Aggregate Data and Process**: Dataset used that can used Mechanism, Model and tools to perform analysis
2. **Analyze and Sanitize data:** post-Processing involves the analysis to perform a test on the training data for the different values of epsilon

On the Front End

1. Perform computations like non-Zero count, Mean, var and SD
2. Test accuracy by comparing with Private and Non-Private data
3. Display the data for different ranges of Epsilon Values

A. Low-Level diagram:

Flowchart to explain each module involved in the web application are Accountant, Mechanism, Model, and Tools

1. **Pre-Processing Data** given flowchart provides data flow for accountant module to show, so the preprocessing of the dataset in Figure 6 involves applying mechanism like GaussianNB to perform aggregation of the dataset. The next step would be part of post-processing analysis, use sampling of a dataset for sanitization.

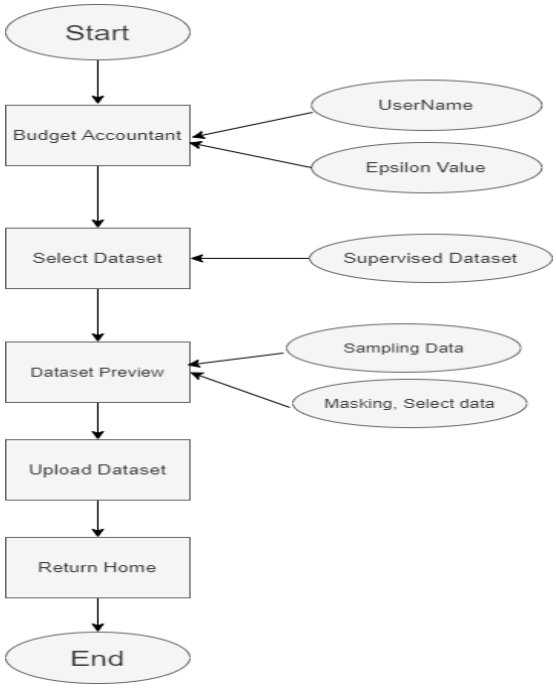


Fig 6: Pre-Processing Data Analysis

2. **Post- Processing Data** involves a complex process in the flowchart given in Figure 7, applies the algorithm for differential privacy using the Classification and regression model. The final step is to train a model using the GaussianNB mechanism to perform aggregation of the dataset

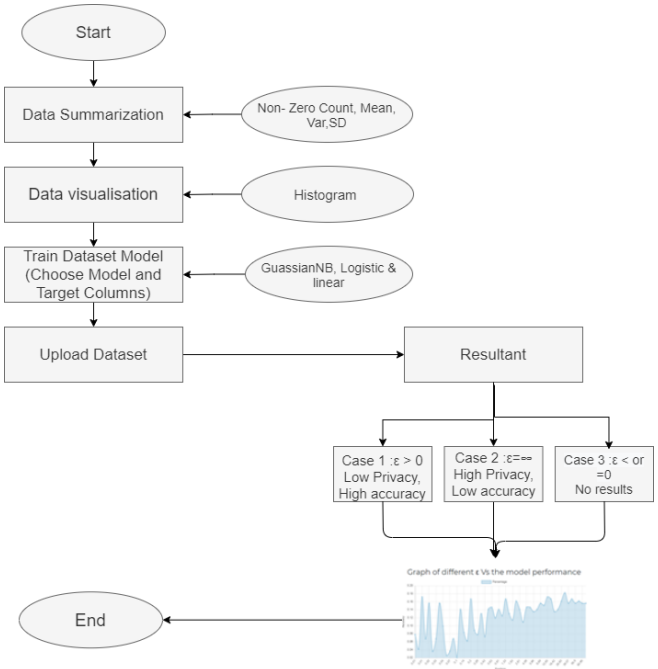


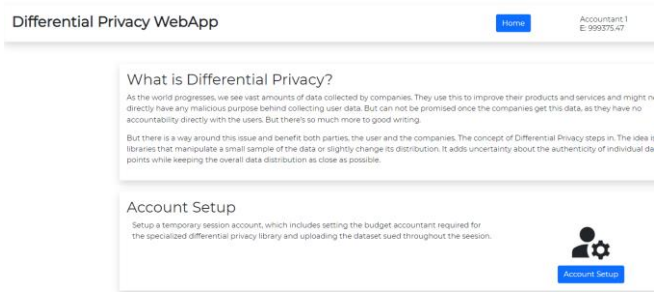
Fig 7: Post-Processing Data Analysis

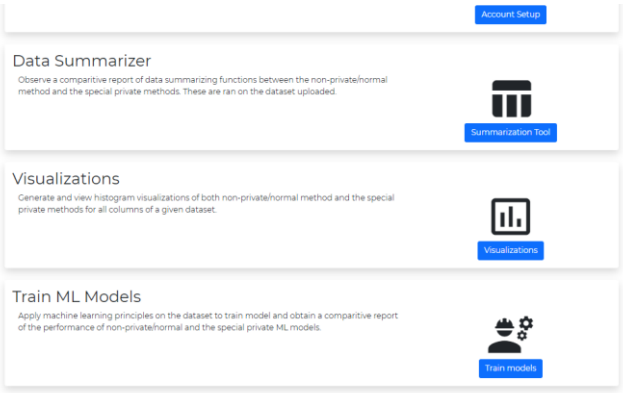
VII. IMPLEMENTATION

The Web application has to perform the computation on Differential privacy open-sourced algorithm using Python-based technologies such as Flask and other python libraries like Scikit[13], NumPy[12], SciPy[14] in the Diffprivlib[15]. The application is deployed on a localhost python-based server using version 3. Code is also available on private GitHub account It is of a lightweight Python-based Framework called Flask.

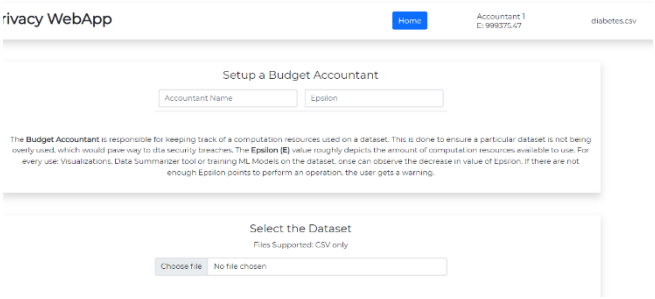
The below-given screenshot displays the front end and helps to understand the navigation and deployment of the Differential privacy library that has been integrated into open-sourced Web Application

1. Home Page is the main page for all functional modules that are given in the below screenshot

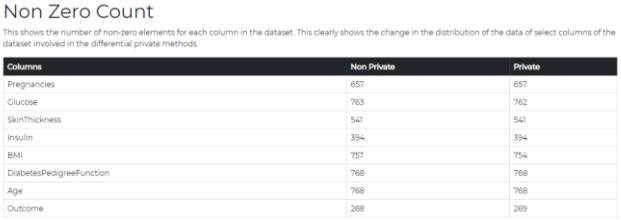




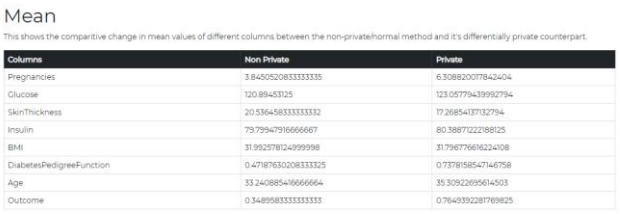
2. The accountant module has the functionality of account Setup to perform the data upload and preview, and the dataset performs the masking and clustering by shuffling columns.



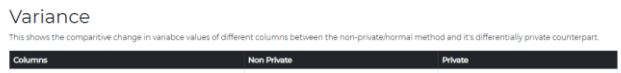
3. Data Summarizer that displays the comparative Analysis of Non-Zero Count, Mean, Variance, and Standard Deviation, Calculates the accuracy of the data.



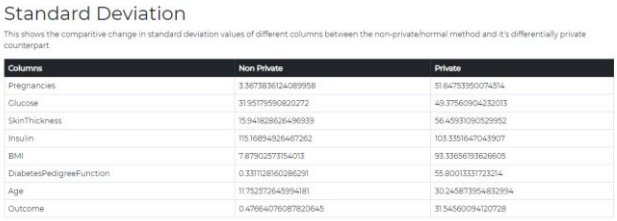
Columns	Non Private	Private
Pregnancies	657	657
Glucose	763	762
SkinThickness	541	541
Insulin	394	394
BMI	757	754
DiabetesPedigreeFunction	768	768
Age	768	768
Outcome	268	269



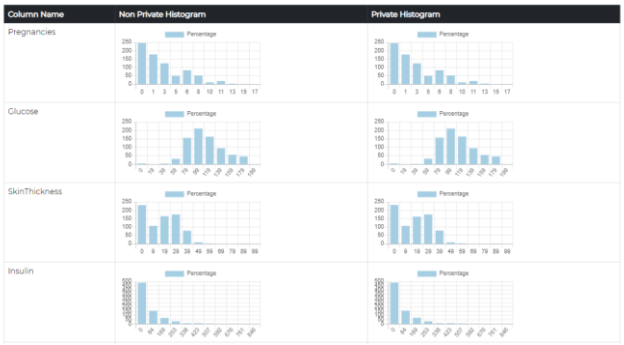
Columns	Non Private	Private
Pregnancies	3.8450520833333335	6.308820017842404
Glucose	120.89453125	123.05779439992794
SkinThickness	20.536458333333332	17.26854137132794
Insulin	79.79947916666667	80.38871222188125
BMI	31.992578124999998	31.796776616224108
DiabetesPedigreeFunction	0.4718763020833325	0.7378158547146758
Age	33.240885416666664	35.30922695614503
Outcome	0.3489583333333333	0.7649392281769825



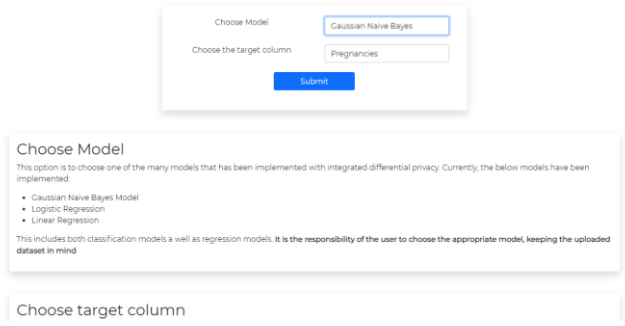
Columns	Non Private	Private
Pregnancies	11.339272393120657	3607.181431402947 5
Glucose	1020.9172617594401	1128.1764230416895
SkinThickness	254.14189999659726	3296.525740509182
Insulin	13263.886874728734	12788.735949955071
BMI	62.07904647827148 6	1675.8665306464165
DiabetesPedigreeFunction	0.10963569693840876	1760.6713210634136
Age	138.12296379937067	326.63190465055663
Outcome	0.22718644493055558	520.8563476137723



4. Data Visualization that displays data plotted using Histogram dataset. A Graph plots with X-axis with column value and Y-axis with epsilon value to compare data pre-and post-processing with differential privacy on Private and Non-Private in the below screenshot.



5. Train Model with three different computational mechanisms on classification and regression dataset for GaussianNB model, Linear regression, and Logistic Regression.



This chapter has implementation details that include a list of tools, commands for installation, and framework design with a screenshot of a web application; Next chapter provides complete in-depth details of testing business usecase.

VIII. ANALYSIS AND RESULTS

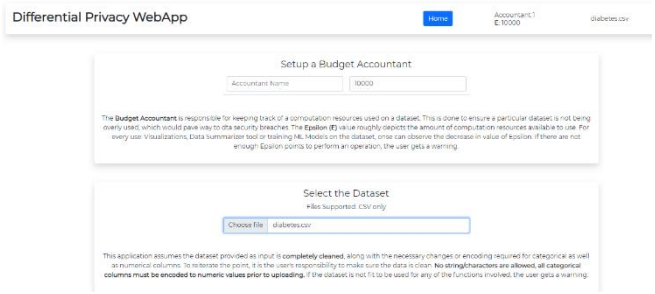
Project and analysis show that the web application built is open source using IBM's Differential Privacy Library called Diffprivlib[15]. The Final Outcome of the Private differential data is in the form of having values given in the form of a graph, and simple computational values that have compared to present the data in between private and non-Private data results that provide graph plots to display if the Differential privacy is applied has some amount of data leakage in this section. For the technical analysis performed, the computational results below

A. Business use-case

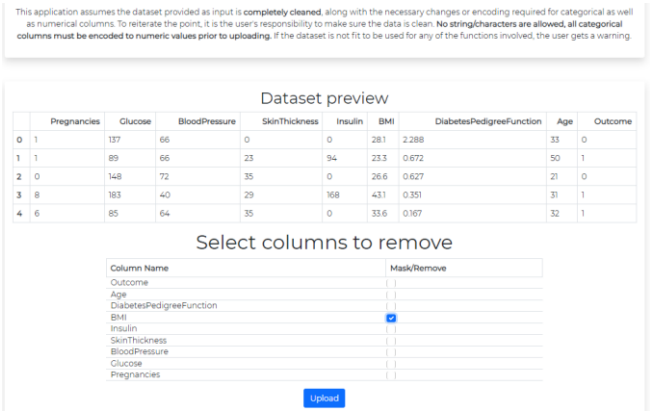
Use case:

Consider use case using healthcare dataset for diabetes as an example[16] to perform classification and regression to generate outputs for different Epsilon content results given in the next chapter.

1. Using the diabetes dataset, load the data with Accountant Name and CSV File. The dataset contains columns: Pregnancies, Glucose, blood pressure, skin thickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome



2. Perform Sampling and display data by masking column that contains the private information related to a user. In this example, select BMI as confidential to mask.



3. In data summarizer displays the data accuracy pre- and post-application of differential privacy algorithm; those values show how the accuracy is changing with each column on different epsilon values. The step helps in easy comparison and understanding of change inaccuracy of data.

Similarly for Mean Value change in comparison to the mean of each column denotes how much data is a change in accuracy post-processing, and further similarly for Variance

and standard deviation values to understand the differences in accuracy over different values of Epsilon

Mean

This shows the comparitive change in mean values of different columns between the non-private/normal method and it's differentially private counterpart.

Columns	Non Private	Private
Pregnancies	3.845052083333335	2.2152951617954
Glucose	120.89453125	120.67065933250805
BloodPressure	69.10546875	70.59581551915994
SkinThickness	20.536458333333332	20.568733697893208
Insulin	79.79947916666667	80.3636180804127
BMI	31.992578124999998	32.77090717107628275
DiabetesPedigreeFunction	0.4718763020833325	0.4969189533874655
Age	33.240885416666664	32.23105779614771
Outcome	0.3489583333333333	0.0

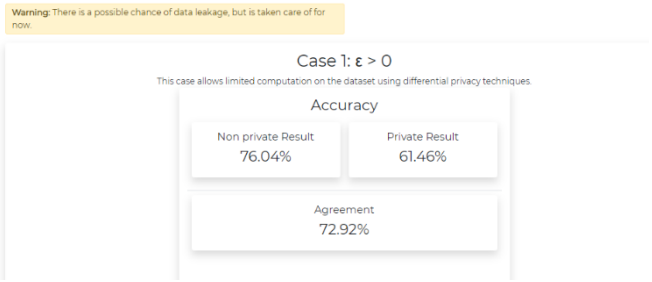
Variance

This shows the comparitive change in variabce values of different columns between the non-private/normal method and it's differentially private counterpart.

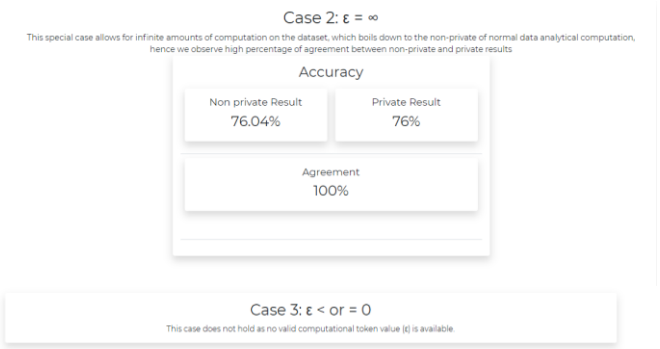
Columns	Non Private	Private

From the example of the diabetes dataset that performs GaussianNB, the below results are for the epsilon values. While Epsilon (epsilon= float("inf")) is applied with different values, check iteration

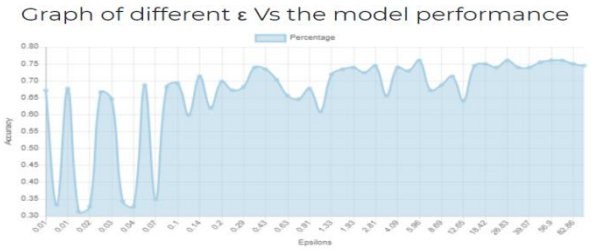
1. High Epsilon (i.e., greater than 1) gives better and more consistent accuracy but less privacy. Here the agreement gave 26.56%, which depicts that data confidentiality is less if epsilon is high.



2. Small Epsilon (i.e., less than 1) gives better privacy but worse and less consistent accuracy. Privacy is 89.58% percent while ε = ∞ provides accuracy and is the perfect fit dataset.



The Resultant Graph for all the epsilon values versus the accuracy results in the below-given screenshot



The above results in the graph show the values change on an overall dataset for the different epsilon values; in every case, if ε = ∞ chooses values randomly until it gets the perfect fit for privacy, it will give the least data accuracy.

IX. CONCLUSIONS AND FUTURE SCOPE

A. *Conclusion*

The primary scope of work was to develop computational results for a dataset using a differential privacy algorithm from IBM's general purpose and an open-source python-based library called Diffprivlib[9] for simulation, experimentation, and development of this web application. Library also includes standard tools, mechanisms, and models with the functionality of version 0.4.1 Having implemented for classification with GaussianNB and regression- Linear and Logistic type of model using a supervised dataset that could display the visualization and histogram results for the different values of epsilon. A web application developed on packages like Numpy[12], Scikit[13], and SciPy[14]. Improving the efficiency of the model and provide more user-friendly data considering personal private data.

B. *Future scope of work*

Roadmap for the future looks at complex data

- The web application has the potential of a broader scope of development to include many other mechanisms and models.
- To build an ML application using diffprivlib can be used dynamically, not just for a private analysis.
- The library can upgrade for the latest packages and other tools to improve the efficiency and performance of Various datasets.
- The tool can use an unsupervised dataset that contains text with raw format using the k-means algorithm.
- From the security standpoint, recommendations are if the dataset is vulnerable to any attack; from IBM and Poneman institutes data breach report estimates "Mega Breach" up to 1 to 50 million costs.[4]

Organizations adopting the emerging technologies and securing them remotely during the pandemic will require some Machine learning of datasets to have privacy in compliance with GDPR standards. The future would rely on having an automated feature that could depend only on Artificial intelligence and Machine Learning, and this project can have enormous potential to develop and be made available for the larger community and educational purposes.

ACKNOWLEDGMENT

I would like to acknowledge the support provided by Chancellor Dr. P Shyama Raju, Dr. S.Y Kulkarni, Ex-Vice Chancellor, Dr. K. Mallikharjuna Babu, Vice-Chancellor, and Dr. M. Dhanamjaya, Registrar.

I would like to sincerely acknowledge the great support and encouragement by Director, Corporate Training for RACE, and all the eminent RACE Faculty who mentored with efficacy throughout the course.

I would like to acknowledge Special thanks to all my mentors, for helping and mentoring me throughout the Course.

And this whole course would not have been possible without the support of my family, my office colleagues and my classmates at REVA, who made learning interesting, fun and collaborative.

Last but not least, I would like to extend my heartfelt thanks to the Admin, Infrastructure team, and extended support staff at RACE, and the diligent interns, for all the help and support.

REFERENCES

[1]. F. K. Naoise Holohan, Steve Martinelli, "IBM's open sourced Differential Privacy library - Diffprivlib," 2019. https://github.com/IBM/differential-privacy-library/.
[2]. A. Alnemari, R. K. Raj, C. J. Romanowski, and S. Mishra, "Protecting Personally Identifiable Information (PII) in Critical Infrastructure Data Using Differential Privacy," *2019 IEEE Int. Symp. Technol. Homel. Secur. HST 2019*, pp. 6–11, 2019, doi: 10.1109/HST47167.2019.9032942.
[3]. M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, and S. Jana, "Certified robustness to adversarial examples with differential privacy," *Proc. - IEEE Symp. Secur. Priv.*, vol. 2019-May, pp. 656–672, 2019, doi: 10.1109/SP.2019.00044.
[4]. IBM Security, "Cost of Data Breach Report 2020," p. 82, 2020, [Online]. Available: https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/.
[5]. M. Abadi *et al.*, "Deep learning with differential privacy," *Proc. ACM Conf. Comput. Commun. Secur.*, vol. 24-28-Octo, no. Ccs, pp. 308–318, 2016, doi: 10.1145/2976749.2978318.
[6]. A. Watson, "Differential Privacy towards Data Science," [Online]. Available: https://towardsdatascience.com/tagged/differential-privacy.
[7]. "De-identification Tools," 2018, [Online]. Available: https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/focus-areas/de-id/tools.
[8]. N. Holohan, D. J. Leith, and O. Mason, "Optimal Differentially Private Mechanisms for Randomised Response," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 11, pp. 2726–2735, 2017, doi: 10.1109/TIFS.2017.2718487.
[9]. N. Holohan, "Diffprivlib: Differential Privacy Library," 2020, [Online]. Available: https://diffprivlib.readthedocs.io/en/latest/index.html.
[10]. D. Su, J. Cao, N. Li, E. Bertino, and H. Jin, "Differentially private K-Means clustering," *CODASPY 2016 - Proc. 6th ACM Conf. Data Appl. Secur. Priv.*, pp. 26–37, 2016, doi: 10.1145/2857705.2857708.
[11]. K. S. S. Kumar and M. P. Deisenroth, "Differentially Private Empirical Risk Minimization with Sparsity-Inducing Norms," vol. 12, pp. 1069–1109, 2019, [Online]. Available: http://arxiv.org/abs/1905.04873.
[12]. "Numpy's histogram function," [Online]. Available: https://numpy.org/doc/stable/reference/generated/numpy.histogram.html.
[13]. "Scikit," [Online]. Available: https://scikit-learn.org/stable/.
[14]. "Scipy," [Online]. Available: https://www.scipy.org/.
[15]. N. Holohan, S. Braghin, P. Mac Aonghusa, and K. Levacher, "Diffprivlib: The IBM Differential Privacy Library," pp. 1–5, 2019, [Online]. Available: http://arxiv.org/abs/1907.02444.
[16]. "Diabetes Dataset," [Online]. Available: https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html.