Blue text on a black background

Description automatically generated

**College of Engineering, Environment, and Computing**

**School of Science**

**MSc Computer Science**

**7151CEM Computing Individual Research Project**

**Project Report**

**Mathematical Approaches to Ensuring Data Privacy: Analyzing Differential Privacy and Cryptographic Techniques**

**By:**

**Student Name: Thanushri Mekala**

**Student ID: 14992461**

**1st Supervisor: Manizheh Montazerian**

**2nd Supervisor: Beate Grawemeyer**

**Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Computer Science**

**Academic Year: 2024/25**

**Declaration of Originality**

I declare that this project is all my own work and has not been copied in part or in whole from any other source except where duly acknowledged. As such, all use of previously published work (from books, journals, magazines, internet etc.) has been acknowledged by citation within the main report to an item in the References or Bibliography lists. I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

**Statement of copyright**

I acknowledge that the copyright of this project report, and any product developed as part of the project, belongs to Coventry University. Support, including funding, is available to commercialize products and services developed by staff and students. Any revenue that is generated is split with the inventor/s of the product or service. For further information please see [www.coventry.ac.uk/ipr](http://www.coventry.ac.uk/ipr) or contact [ipr@coventry.ac.uk](mailto:ipr@coventry.ac.uk)**.**

**Statement of ethical engagement**

I declare that a proposal for this project has been submitted to the Coventry University ethics monitoring website (https://ethics.coventry.ac.uk/) and that the application number is listed below (Note: Projects without an ethical application number will be rejected for marking).

Signed: Thanushri Mekala

**Date: 10/04/2024**

**Please complete all fields.**

|  |  |
| --- | --- |
| **First Name:** | **Thanushri** |
| **Last Name:** | **Mekala** |
| **Student ID number** | **14992461** |
| **Ethics Application Number** | **P185044** |
| **1st Supervisor Name** | **Manizheh Montazerian** |
| **2nd Supervisor Name** | **Beate Grawemeyer** |

**Module Name: Computing Individual Research Project**

**Module Code: 7151CEM**

**Title: CW-Report**

**Student Name: Thanushri Mekala**

**Student ID: 14992461**

**Mathematical Approaches to Ensuring Data Privacy: Analyzing Differential Privacy and Cryptographic Techniques**

**Abstract**

Modern data-driven apps are at great risk when it comes to data protection, especially in fields like banking, healthcare, and cloud computing. The goal of this study is to look into the mathematical basis of data privacy, focussing on differential privacy and cryptographic methods. The study looks at the trade-off between privacy and data usefulness by using both real-world tests and theory analysis. It is also made clear what the pros and cons of different privacy-protecting methods are. The study's findings show that differential privacy is a flexible way to keep people's privacy safe. However, the privacy budget (ς) needs to be carefully adjusted in order to find a good mix between data security and accuracy. Some methods, like safe multi-party computing and homomorphic encryptions, offer better security promises that make it harder to use cryptography approaches in large-scale systems. But these ways can't be used in real life because they are too expensive to process.

Researchers are also looking into mixed methods, which use both differential privacy and encryption to make data safer while still letting computers work normally. The studies' results show that machine learning techniques that put privacy first, like differentially private stochastic gradient descent, might be able to make security better. Nevertheless, more optimisation is needed to properly reduce the loss of accuracy. Researchers found that technologies that improve privacy must be carefully made to meet law requirements and industry-specific standards. The streamlining of mixed models, the creation of new ways to lower the privacy-utility trade-offs, and the improvement of computer efficiency should be top goals for the future. In a world where everything is becoming digital, this will keep info safe.

Table of Contents

[Chapter 1: Introduction 11](#_Toc195199804)

[1.1 Background and Rationale 11](#_Toc195199805)

[1.2 Problem Statement 12](#_Toc195199806)

[1.3 Research Aim and Objectives 13](#_Toc195199807)

[1.4 Research Questions 13](#_Toc195199808)

[1.5 Scope of the Study 13](#_Toc195199809)

[1.6 Significance of the Study 14](#_Toc195199810)

[Chapter 2: Literature Review 15](#_Toc195199811)

[2.1 Introduction 15](#_Toc195199812)

[2.2 Differential Privacy: Concept and Applications 16](#_Toc195199813)

[2.3 Cryptographic Techniques for Data Privacy 17](#_Toc195199814)

[2.4 Comparative Analysis of Differential Privacy and Cryptographic Techniques 18](#_Toc195199815)

[2.5 Challenges and Open Issues in Data Privacy 20](#_Toc195199816)

[2.6 Summary and Research Gaps 21](#_Toc195199817)

[Chapter 3: Research Methodology 23](#_Toc195199818)

[3.1 Introduction 23](#_Toc195199819)

[3.2 Research Design 23](#_Toc195199820)

[3.3 Data Collection 24](#_Toc195199821)

[3.4 Implementation Strategy 28](#_Toc195199822)

[3.5 Tools and Technologies 32](#_Toc195199823)

[3.5.1 Programming Languages and Libraries 32](#_Toc195199824)

[3.5.2 Hardware for computers and the place where they are used 33](#_Toc195199825)

[3.5.3 The experimental setup and the datasets it works with 33](#_Toc195199826)

[3.6 Ethical and Security Considerations 34](#_Toc195199827)

[3.6.1 Ethical Principles in Data Privacy Research 35](#_Toc195199828)

[3.6.2 Legal Compliance and Data Protection Regulations 36](#_Toc195199829)

[3.6.3 Research Transparency and Reproducibility 36](#_Toc195199830)

[3.7 Project Management Approach 36](#_Toc195199831)

[3.8 Risk Assessment and Mitigation 37](#_Toc195199832)

[Chapter 4: Data Analysis and Findings 38](#_Toc195199833)

[4.1 Introduction 38](#_Toc195199834)

[4.2 Mathematical Analysis of Differential Privacy and Cryptographic Techniques 38](#_Toc195199835)

[4.3 Mathematical Justification of Differential Privacy and Cryptographic Security 40](#_Toc195199836)

[4.3.1. Logistic Regression Loss Function Calculation 40](#_Toc195199837)

[4.3.2. Sensitivity Analysis Calculation 42](#_Toc195199838)

[4.3.3. Laplace Mechanism Calculation 43](#_Toc195199839)

[4.3.4. Exponential Mechanism Probability Calculation 44](#_Toc195199840)

[4.3.5. Privacy-Utility Trade-off Calculation 45](#_Toc195199841)

[4.3.6. Homomorphic Encryption 45](#_Toc195199842)

[4.3.7. DP-SGD Calculation 46](#_Toc195199843)

[4.3.8. Secure Multi-Party Computation Example 47](#_Toc195199844)

[4.3.9. Information-Theoretic Privacy Calculation 48](#_Toc195199845)

[4.4 Findings and Discussion 49](#_Toc195199846)

[4.5 Discussion and implications 55](#_Toc195199847)

[4.6 Summary 57](#_Toc195199848)

[Chapter 5: Conclusion 57](#_Toc195199849)

[Screenshots 59](#_Toc195199850)

[Appendix 65](#_Toc195199851)

**List of Keywords**

* Data Privacy
* Differential Privacy (DP)
* Cryptographic Techniques
* Privacy-Preserving Machine Learning (PPML)
* Homomorphic Encryption (HE)
* Secure Multi-Party Computation (SMPC)
* Privacy-Utility Trade-off
* Privacy Budget (ϵ\epsilonϵ)
* Federated Learning (FL)
* Anonymization
* Data Security
* Mathematical Models for Privacy
* Privacy-Preserving Data Analysis
* Regulatory Compliance (GDPR, CCPA)
* Noise Injection

**List of Abbreviations**

| Abbreviation | Full Form |
| --- | --- |
| DP | Differential Privacy |
| HE | Homomorphic Encryption |
| SMPC | Secure Multi-Party Computation |
| PPML | Privacy-Preserving Machine Learning |
| FL | Federated Learning |
| GDPR | General Data Protection Regulation |
| CCPA | California Consumer Privacy Act |
| SGD | Stochastic Gradient Descent |
| LDP | Local Differential Privacy |
| CDP | Central Differential Privacy |
| DP-SGD | Differentially Private Stochastic Gradient Descent |
| FHE | Fully Homomorphic Encryption |
| TFHE | Torus Fully Homomorphic Encryption |
| ML | Machine Learning |
| AI | Artificial Intelligence |
| IoT | Internet of Things |
| LLM | Large Language Model |
| ERM | Empirical Risk Minimization |
| MPC | Multi-Party Computation |
| AES | Advanced Encryption Standard |
| RSA | Rivest–Shamir–Adleman (Cryptographic Algorithm) |
| ECC | Elliptic Curve Cryptography |
| RDP | Rényi Differential Privacy |
| FPGA | Field-Programmable Gate Array |
| TPM | Trusted Platform Module |

**List of Figures**

[Figure 1: Codes for data import 25](#_Toc194746057)

[Figure 2: Logistic Regression operations 28](#_Toc194746058)

[Figure 3: Implementation Code 44](#_Toc194746059)

[Figure 4: Model Training 45](#_Toc194746060)

**List of Graphs**

[Graph 1: Partial Utillity 40](#_Toc194746070)

[Graph 2: Model Accuracy 48](#_Toc194746071)

[Graph 3: No. of Distribution 49](#_Toc194746072)

[Graph 4: Trade-off 50](#_Toc194746073)

[Graph 5: Laplace Noise 51](#_Toc194746074)

[Graph 6: Exponential Mechanism 52](#_Toc194746075)

[Graph 7: Gradient 53](#_Toc194746076)

[Graph 8: Entropy deduction 55](#_Toc194746077)

# Chapter 1: Introduction

## Background and Rationale

Because decisions are being made more and more based on data, protecting data privacy has become a big problem for modern computer systems. Businesses in many areas, like healthcare, banking, and e-commerce, collect huge amounts of personal information in order to make their business plans more effective, come up with new apps, and provide better customer service. Still, as the amount of data being collected has grown at an exponential rate, so have worries about security breaches, illegal access, and the right way to use data. Even though they can provide some security, traditional data protection methods like encryption and access control systems are not enough to stop complicated attacks, data breaches, or worries about re-identification.

It can be hard to find the right mix between security and usefulness when it comes to data privacy. Businesses are responsible for keeping customer data private, but researchers, analysts, and AI models are in charge of getting to this data and finding useful information in it. Linkage attacks can break through common ways of keeping data private, like data encryption and generality. An attacker uses more than one set of data to re-identify people as part of these attacks. Deep mathematical methods like differential privacy and cryptographic algorithms have been looked into by academics to try to solve these problems. Their objective is to improve data security without making the info much less useful.

Differential privacy is a mathematical system that can be used to keep people from being identified while still getting useful data information. Controlled noise is used to improve the processes of data processing. It makes sure that the chance of spotting a person stays pretty much the same, no matter if their data is in the file or not. Two real-life examples of this technology in use are Google's use of differential privacy for user analytics and Apple's privacy-preserving data collection technologies. Yet, implementing differential privacy in very big and complex datasets is still hard because of the trade-offs between speed and accuracy, even though it is useful.

Cryptographic methods, on the other hand, add an extra layer of security to data while it is being processed and sent, making sure that its privacy is always protected. Homomorphic encryption lets processes be run on protected data without having to decode it first. This keeps the data's privacy intact while it is being analysed. With secure multi-party computing (SMPC), people from different groups can work together to analyse data without anyone's information being at risk. Although these cryptography methods provide strong security, they are hard to use in large-scale systems because they require a lot of computing power.

In order to find a good balance between security, data usefulness, and speed, hybrid systems that use both differential privacy and cryptography are becoming more and more important. To explain, this is because every one of these ways has its own pros and cons. This study looks into the math behind differential privacy and cryptography, compares how they can be used in real life, and offers a model that combines the best parts of both. The project's main goal is to put these privacy-protecting solutions into action and try them using organised and unstructured data from sensitive fields like healthcare and banking.

Overall, the goal of this project is to create safe, expandable systems that protect users' privacy. Therefore, in order to reach this goal, we will carefully look into all the mathematical models that deal with data privacy. The outcomes could be very helpful for businesses, lawmakers, and researchers who want to enhance data security while following legal requirements like the General Data Protection Regulation (GDPR) and HIPAA standards. It is hoped that this study will lay the groundwork for future progress in AI that puts an emphasis on private protection, safe data sharing, and responsible data use.

## 1.2 Problem Statement

* Existing data privacy techniques face challenges in balancing security, accuracy, and computational efficiency.
* Traditional anonymization methods are vulnerable to re-identification attacks.
* There is a need for robust, scalable, and mathematically sound privacy-preserving models.

## 1.3 Research Aim and Objectives

**Aim:**

* To analyze and evaluate mathematical approaches, cryptographic techniques, and differential privacy for enhancing data security and privacy.

**Objectives:**

* Investigate existing data privacy techniques, including differential privacy and cryptography.
* Evaluate the trade-offs between privacy, security, and computational efficiency.
* Develop and test a hybrid approach integrating differential privacy and cryptographic methods.
* Assess the effectiveness of these methods in real-world applications.

## 1.4 Research Questions

* How can mathematical techniques like differential privacy and cryptography enhance data privacy while maintaining usability?
* What are the trade-offs between security, accuracy, and computational efficiency in privacy-preserving methods?
* How can hybrid models improve data protection in sensitive applications like healthcare and finance?

## 1.5 Scope of the Study

* Focuses on two key privacy techniques: differential privacy and cryptography.
* Includes practical implementation using Python-based frameworks.
* Evaluates methods based on privacy strength, computational cost, and data utility.
* Limited to structured and semi-structured datasets in domains such as healthcare and finance.

## 1.6 Significance of the Study

* Contributes to data privacy research by exploring hybrid privacy models.
* Aligns with regulatory frameworks such as GDPR and HIPAA for secure data handling.
* Helps organizations adopt privacy-preserving techniques while maintaining data utility.
* Supports advancements in privacy-preserving AI and secure data-sharing practices.

# Chapter 2: Literature Review

## 2.1 Introduction

A growing number of sectors, such as healthcare, finance, and social networking, are dependent on data-driven applications, making data security an important issue in modern computing. Protecting private information while keeping data valuable is becoming more difficult as the quantity of data being gathered and processed continues to expand at an exponential rate. To address these privacy problems, many mathematical and computational approaches have been created. Notable among these replies are methods that use differential privacy and encryption. Below, we shall delve deeper into differential privacy and its formal foundation for data anonymisation. This framework ensures that changing the value of one data point won't change the result significantly when applied to the whole. Data security during storage and transmission is the primary focus of cryptographic procedures. Data breaches and unauthorised access to it are what these protocols are aiming to stop.

Research done so far indicates that there are benefits and drawbacks to both kinds of methods. Jiang et al. (2021) and Jiang et al. (2021) both note that differential privacy has been extensively studied in several applications, including social network analysis and the industrial Internet of Things. On the other hand, data-driven models' accuracy could suffer since their efficacy is typically correlated with the quantity of noise supplied. This is because there is a wide array of potential entry points for noise. However, there is a hefty processing cost associated with these systems (Marcolla et al., 2022). Two cryptographic techniques that provide robust security assurances are homomorphic encryption and secure multi-party computing. The goals of this literature review are to(1) thoroughly investigate both methodologies,(2) assess the relative merits of the two techniques, and(3) explore the potential for hybrid models that combine the best features of each. The goal of this chapter is to provide the groundwork for a more comprehensive analysis of mathematical methods for protecting personal information. This will be achieved by reviewing previous studies and determining which ones need more study.

## 2.2 Differential Privacy: Concept and Applications

As a complex mathematical framework that protects data security and allows data analysis that is commensurate with its importance, differential privacy has become more popular. The thinking used to be first of all proposed by way of Dwork (2006), and it guarantees that research using a dataset could always provide similar outcomes, no matter whether or not non-public facts is protected. this is made feasible by means of including calibrated noise to the query results. As a result, it's very tough for enemies to infer particulars approximately particular statistics portions. making use of differential privateness is especially beneficial in contexts concerning huge-scale data-sharing which are governed by privateness rules like HIPAA and the overall data protection regulation (GDPR).

the usage of differential privateness in numerous fields has been the difficulty of latest studies. clean grasp has been gleaned from those probes. Jiang et al. (2021) have a look at its use within the IoT for commercial functions and the way it aids in protecting sensor information from opposed inference assaults. On pinnacle of that, social community analysis has made use of differential privacy to mask user sports even as maintaining community structural fee (Jiang et al., 2021). In contrast, differential privacy gives a plethora of extensive benefits, on top of all those already listed. Olabim et al. (2024) states that a exceptional deal of facts distortion may end result from adding an excessive amount of noise, which could affect device mastering and downstream analytics fashions. To prevail, you should triumph over this, one of the most considerable challenges. moreover, excessive-dimensional datasets nevertheless offer a hurdle when seeking to apply differential privacy. this is because of the truth that both the trouble's complexity and processing prices are at the upward thrust.

There has been quite a few have a look at on the topic of differential privateness strategies, and the outcomes have shown that there is room for development. for instance, adaptive noise addition methods are proposed by means of Abdalzaher et al. (2022) as a method to achieve the candy spot among records price and privateness safety. with a purpose to locate the sweet spot, this is carried out. Yan et al. (2024) carried out studies on the use of differential privateness to the problem of defensive huge language fashions. The motive of this examine is to discover the way to prevent extraction attacks at the sensitive information utilized in employee training. regardless of a lot of these successes, differential privacy remains a unexpectedly increasing place this is usually being studied to discover better approaches to put it into practice. in line with the records, this is the way the sector is moving..

## 2.3 Cryptographic Techniques for Data Privacy

When secure communication and data storage are of the utmost importance, it is crucial to have cryptographic protocols in place to ensure that data remains secret. For a very long time, people have relied on symmetric and asymmetric encryption and other traditional methods to keep unauthorised parties from accessing sensitive data. But this aside, more sophisticated cryptographic techniques have garnered a lot of attention recently. These techniques encompass, amongst others, 0-understanding proofs, tightly closed multi-birthday celebration computing (SMPC), and homomorphic encryption. reason being, issues associated with data privacy are becoming ever greater complex.

In latest years, homomorphic encryption has emerged as a totally promising development inside the realm of cryptographic privateness. via casting off the want to decrypt information earlier than acting calculations on it, it guarantees that touchy statistics stays tightly closed for the duration of processing. A full observe on homomorphic encryption is posted via Marcolla and coworkers in 2022. each the theoretical underpinnings and sensible programs of the generation are covered on this paper. This method's quintessential flaw is the substantial quantity of computing investment it requires; as a result, it isn't always suitable for use in widely disbursed systems. This method offers strong privateness safeguards, however it falls short in relation to core strength. preserving inside the identical vein, impervious multi-birthday party computing permits several members to collaborate on a characteristic over their inputs while concurrently making sure the confidentiality of every input. The mathematical foundations of these strategies and their ability to impervious collaborative computations in many regions had been studied in studies like Bowman et al. (2021). A huge variety of calculations may be protected using those methods.

Combining blockchain era with features that guarantee users' anonymity is yet some other creative cryptographic answer. Wylde et al. (2022) defined in their studies how blockchain-based encryption processes might improve facts safety in decentralised settings. further to protecting touchy statistics from prying eyes, these solutions preserve information exchange transactions and operations transparent. On the alternative aspect, there are several issues with using cryptographic strategies, most extensively with scalability and computational performance. As the amount of computing strength wanted for secure encryption maintains to upward jostle, optimisation has come to be an important and active area of study (Capraro & Perc, 2021). this is owing to the truth that secure encryption might also need extra computing energy, which is probably prohibitive for actual-time packages.

New techniques are being created as cryptographic strategies develop with the aim of improving efficiency and boosting interplay with current privacy frameworks. this is persevering with even if there are flaws in cryptographic structures. The increasing significance of predictive analytics in making strategic decisions is introduced to mild by way of Adesina et al. (2024), who also advocate that cryptographic privateness methods may in addition make stronger records safety in BI systems. They strain how essential predictive analytics is for making big-picture choices. To higher protect sensitive statistics in complicated actual-international instances, hybrid models combining differential privateness with cryptographic tactics may offer more scalable and practical solutions. As generation advances, that is a herbal development for the enterprise.

## 2.4 Comparative Analysis of Differential Privacy and Cryptographic Techniques

The main point of disagreement between differential privacy and cryptographic methods is how well they protect data without making it less useful. You may see that the 2 approaches are very one of a kind in how they paintings, how they are alleged to be used, and how difficult they're to compute. both technologies are meant to maintain people who are not presupposed to have get right of entry to from attending to non-public information. Differential privateness is a superb thanks to protect human beings's privateness as it makes positive that including or getting rid of the facts of any one person would not considerably change the results of an analysis. alternatively, cryptographic techniques cognizance on maintaining records secure at the same time as it's miles being processed or even as it's miles at rest. This maintains the data safe by means of solely letting authorized events view it. it is important to compare one of a kind tactics to find out their pros and cons and the way they could paintings collectively so one can make greater complete models that protect people's privateness.

Industries that handle quite a few private information, like banks and healthcare, have moved fast to implement one-of-a-kind safety measures to live in step with the regulations (Jiang et al., 2021). A big plus is that differential privacy may want to cause beneficial statistical analysis and sturdy theoretical privacy claims. that is the primary advantage of differential privacy. consistent with Olabim et al. (2024), the usage of the noise addition technique to defend privateness regularly means giving up a few records's value with a purpose to guard privateness. an excessive amount of noise can lower the quality of the facts, making it much less useful for making choices and education device getting to know models. Differential privacy additionally has problem with high-dimensional data, because of this that noise makes things tougher and might exchange the effects in huge approaches (Jiang et al., 2021).

Cryptographic techniques like homomorphic encryption and safe multi-birthday party computing provide a high level of protection, however they require a whole lot of more paintings to be achieved on the computer (Marcolla et al., 2022). while homomorphic encryption is used, it is viable to do calculations on blanketed data at the same time as nevertheless retaining it absolutely secure. however because it prices lots to compute, it is not very useful in large structures. impenetrable multi-celebration computing is another robust way to keep personal data safe. With this technique, many firms can paintings collectively to secretly compute a characteristic based totally on the statistics they are given. actual-time structures, then again, could not use its many running sources (Bowman et al., 2021). those problems display the boundaries of cryptographic methods in conditions wherein velocity and flexibility are very important.

Differential privacy and cryptography were studied collectively for a long term, with the intention of having the quality of both worlds. The work of Wylde et al. (2022) shows how blockchain generation and differential privateness can be used together to make things greater open and more secure. Differential privacy skill which you do not have to deliver out non-public information, and blockchain era makes certain that any changes to records may be tracked. the usage of each homomorphic encryption and differential privacy together is any other thanks to hold records safe whilst nevertheless letting people use it. but blended fashions have a tendency to make matters greater complex than they need to be, and they could need greater paintings earlier than they may be used in actual life (Capraro & Perc, 2021).

it's hard to mention that is higher because they do not cover each possible situation. both differential privateness and cryptography have their excellent factors. The unique use case, the limited computing strength, and the want for privateness all play a high quality function in judging their usefulness. destiny take a look at should awareness on improving combined models that integrate several methods even as additionally solving computing troubles. this may make privacy-defensive answers simpler to use within the real international.

## 2.5 Challenges and Open Issues in Data Privacy

Even though privacy protections have come a long way, there are still some problems to solve. One of the biggest fears is that data protection and usefulness will be lost. Differential privacy is based on noise inclusion, which protects privacy but could make analysis models less reliable. Even small changes in data can lead to big differences in medical evaluations and treatment ideas. This is a very important problem in healthcare (Jiang et al., 2021). In the same way, cryptographic methods are secure, but they make working harder, which means they might not work for real-time uses (Marcolla et al., 2022). Researchers and practitioners have always had trouble finding the best balance between privacy and usefulness.

There's also the issue of being able to grow. Differential privacy algorithms need to be fine-tuned carefully to ensure privacy across big datasets while also remaining computably feasible. In Olabim et al. (2024), it gets harder to protect people's privacy without compromising the accuracy of the data as systems get bigger and more complicated. Cryptographic methods, especially homomorphic encryption, have similar problems because they need a lot more processing power as the amount of data increases (Marcolla et al., 2022). It is very important to fix these scale problems so that privacy-preserving methods can be used in situations that need a lot of data.

Rules must also be followed, but this can also cause big problems. Businesses must have privacy-protecting plans that follow the rules set by data protection laws like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) in order to follow the law (Jiang et al., 2021). It might be hard to turn privacy protections that make sense in theory into actions that follow the law, though. An excellent example of this is differential privacy, which doesn't always make it clear how much noise needs to be added to meet legal requirements. Also, Wylde et al. (2022) say that cryptographic methods need to be carefully made before they can be used to make sure they work well and meet industry standards.

New threats make problems that are already there worse when it comes to protecting data privacy. There are now a lot of machine learning models that can find trends in data that has been anonymised. This has raised worries about re-identification threats. Yan et al.'s study from 2024 says that even if the noise used isn't enough or isn't adjusted correctly, advanced inference attacks may still be able to target differently private datasets. Because of new security risks like quantum computing, old encryption methods might not work anymore (Marcolla et al., 2022). This threat is still there, though, for encryption systems. Post-quantum cryptography study is very important if we want to protect data protection systems from these kinds of flaws in the future.

More problems with usefulness and acceptance make it harder for privacy protection methods to be widely used. Adesina et al. (2024) say that a lot of information is needed to carry out different privacy-protecting measures. This makes it very hard for businesses that don't have a lot of technology tools to change. To get more people to use it, it's important to make tools and technologies that make it easy to add cryptographic methods and differential privacy to current data processing processes.

To solve these problems, engineers, computer scientists, politicians, and people in the business world must work together across disciplines. Next study should focus on making privacy-protecting programs that can change privacy settings instantly based on how sensitive the data is and what is going on. Furthermore, to make sure that the privacy protections work against new threats that are appearing, they need to be constantly checked and monitored.

## 2.6 Summary and Research Gaps

To keep private data safe, this chapter has given an overview of differential privacy and security techniques, including how they work, what they can and can't do, and their theoretical foundations. Because the idea of differential privacy is based on strong maths, data can be made anonymous and statistical analysis can be done. The privacy-utility trade-off and the difficulties of keeping up with high-dimensional data, on the other hand, limit its usefulness (Jiang et al., 2021). Cryptographic methods like homomorphic encryption and safe multi-party computing offer strong security, but they can't be used on a large scale because they are too slow (Marcolla et al., 2022). By comparing the two approaches, we can see how important it is to create mixed models that use the best parts of both while minimising the bad ones.

There are still a lot of questions about data privacy, even though a lot of progress has been made. First, there is the open business of finding the perfect place to be alone and still be able to do things. This is especially important in high-stakes fields like healthcare and finance, where differentiated privacy protects privacy legally (Olabim et al., 2024) and more research needs to be done on the effect of noise on data quality. Making cryptographic methods faster to process is also important so they can be used in big systems (Bowman et al., 2021).

Second, most privacy-protecting methods right now only work with datasets that don't change. However, in the real world, data that is both dynamic and always changing is common. Yan et al. (2024) say that more research is needed to create privacy systems that are adaptable and can change based on the type of data and privacy risks. Federated learning and edge computing are two new technologies that combine privacy-protecting methods to make it easier to keep data safe without limiting its use.

One area of study that needs more work is making sure that there are standard tools for comparing the different ways that people protect personal data. Differential privacy is a mathematical way to explain why privacy is lost, but there isn't a standard way to judge how well different privacy methods really work in a wide range of situations (Capraro & Perc, 2021). Having consistent standards and review methods in place could lead to better comparisons and new privacy-protecting technologies.

To sum up, more study should be done on the moral and social benefits of privacy protection laws. Even though the point of differential privacy and encryption is to keep certain data safe, these methods could lead to unfair effects that hurt some groups more than others. Adesina et al. (2024) say that fair and clear privacy-preserving algorithms are necessary to encourage responsible data use and stop unintentional bias.

It has been very helpful to use maths to find answers for data privacy issues, but there are still big problems and study gaps that need to be filled. Teams from different fields will need to work together to solve these issues and come up with new ways to do things that are safe, effective, and easy to use. After talking about the study methods used to look into these problems, the next part will focus on how differential privacy and cryptographic algorithms can be used in real life.

# Chapter 3: Research Methodology

## 3.1 Introduction

A lot of data is being gathered, handled, and examined because AI and technologies that are based on data are getting better and better so quickly. Security, privacy, and secrecy of data have become more important as the use of huge systems has grown. Companies in healthcare, banking, and cloud computing are being pushed to protect people's privacy in a way that follows laws like the California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR). However, they must still keep the data's value for useful insights (Wylde et al., 2022). The trend is likely to keep going strong for a while longer. People often use anonymisation and access control to keep personal information from getting out, but they have not worked as well as people thought they would. Re-identification attacks and other more advanced types of attack technology may still use these methods (Zhao & Chen, 2022).

We need to find smart ways to keep data safe while still letting experts use it. Cryptographic methods and differential privacy (DP) are two solutions that could work. When it comes to data and privacy, Husnoo et al. (2021) say that differential privacy is one of the most hopeful ways to look at things. This mathematical method protects individual records by adding controlled noise to datasets. On the other hand, homomorphic encryption (HE) and secure multi-party computing (SMPC) make it safe to send and store data. It can be hard to set these systems up because they need to be quick, safe, and useful in real life (Kim & Cho, 2021). When these methods are used, these issues show up, even though they might be useful.

This study used a method called "mixed-methods research," which blends tests in the real world with the study of theories. The study looks at a lot of different privacy-protecting models that have already been used to learn more about them before putting them to the test. As part of the study itself, security and differential privacy tools are being made and tried on information from the real world. The study's results will help us understand if these methods can be used in real life, how much it costs and how much it benefits to keep data private, and other things.

## 3.2 Research Design

A orderly approach that includes qualitative theory analysis and quantitative trial analysis is used in this study to find the best ways to protect privacy. Cloud computing, healthcare, and banking are some of the areas where the main goal is to look into, try, and use different privacy and security methods (Jiang et al., 2021). The study process is made up of four key steps:

* Bowman et al. (2021) and Li et al. (2023) looked at previous studies and ideas to find out what the pros and cons of different privacy and security methods are in the real world. To do this, all the essential information is carefully studied. As the foundation is laid, this step makes it possible to choose the best mathematical models and methods for use.
* Using differential privacy and secure methods to protect user privacy is a big part of both making and using algorithms. Gupta et al. (2023) say that these ways are meant to make things faster without making them less safe. OpenMined and PySyft are Python tools that make it easier to use machine learning in a way that respects privacy and keeps data safe.
* Tests and evaluations of experiments— People can get real-world datasets from Google Dataset Search, the UCI Machine Learning Repository, and Kaggle. These datasets are used to test and grade the models. The author is Singh et al. (2021). Some of the things that are looked at are how correct the models are, how useful the data is, how fast the models can run, and how well privacy is protected. One of the extra tests is how useful the info is.
* When you look at the results of the tests, you can compare and validate them to find the best way to protect your privacy in terms of both security and value. Iqbal et al. (2023) looked at how well cryptographic models and differential privacy models work and what the pros and cons of each are in different scenarios.

An ongoing method is used in the study to improve the design of the program based on the results of the first tests. In 2021, Jiang et al. said that the differential privacy application is built on changing the privacy budget (µ) to see how it changes how useful the data is. The security system, on the other hand, checks how well cryptography works and how much it costs to run. All methods of protecting privacy will be carefully looked at through this process, which will lead to both academic learning and ideas for what can be done.

## 3.3 Data Collection

Getting the data for this study is very important because the privacy protection methods only work well with the right kinds of information, which are also very hard to understand. The study uses information from real life, like user behaviour analytics, healthcare, and bank payments. Liu et al. (2024) told us how to choose these datasets: they can be protected, they have different levels of privacy, and they can be used in privacy-focused apps.

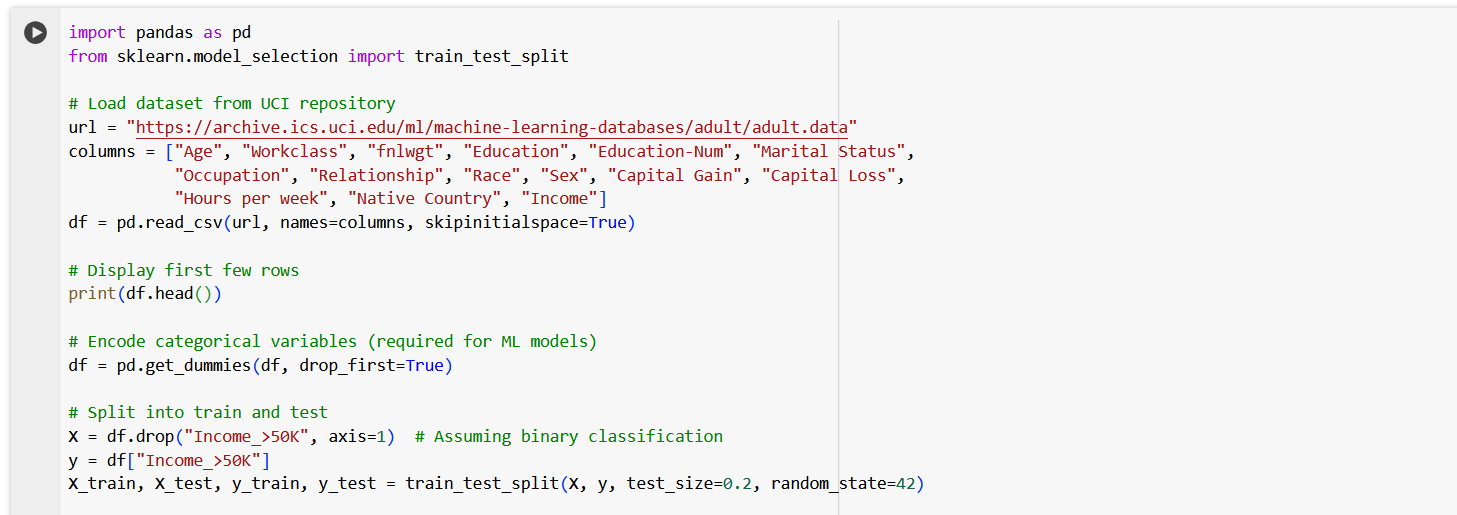


Figure 1: Codes for data import

They have to follow all private rules and morals because the records come from public systems that hold real-world information. When you choose datasets, remember these key points:

* Things that are normally kept secret from the public by privacy laws must be part of the data set. For example, medical or bank records with personal information must be included. In this case, "privacy sensitivity" means.
* To see how well methods for protecting privacy work in big, complicated cases, the dataset needs to have traits with a lot of dimensions. This makes the information more difficult to understand.
* Li et al. (2023) say that for a sample to be statistically significant, it needs to be big enough to give useful views of the privacy-utility trade-off.

This study goes by strict rules to keep data private and looks at ethical standards to make sure that ethics standards are met. Extra steps are taken to get rid of private information before tests are run, even though those steps already make it less likely that someone will be found again. To protect personal data, laws like HIPAA, the General Data Protection Regulation (GDPR), and the California Consumer Privacy Act (CCPA) are in place (Wylde et al., 2022).

**Table 1: Comparison of Differential Privacy and Cryptographic Techniques**

| **Feature** | **Differential Privacy (DP)** | **Homomorphic Encryption (HE)** | **Secure Multi-Party Computation (SMPC)** |
| --- | --- | --- | --- |
| **Privacy Mechanism** | Adds statistical noise to data | Encrypts data for computations on ciphertext | Distributes computation across multiple parties |
| **Data Usability** | High with appropriate ε | Low (encryption adds computational complexity) | Moderate (depends on protocol efficiency) |
| **Computational Overhead** | Low to moderate | High (especially for Fully HE) | High (requires secure channels) |
| **Real-Time Feasibility** | Suitable for real-time analytics | Not suitable due to processing delays | Limited by network latency |
| **Security Level** | Protects against re-identification attacks | Strong security via encryption | Ensures no single party sees the full data |
| **Challenges** | Balancing privacy and accuracy | Heavy computational cost | Complexity in implementation |
| **Best Use Case** | Statistical data analysis, AI/ML models | Secure computations on sensitive data | Collaborative machine learning |

*(Sources: Escobar, 2025; Marcolla et al., 2022; Gupta et al., 2023; Jiang et al., 2021)*

***The following steps are taken to get the files ready after they have been collected:***

As part of cleaning up data, there are steps to deal with missing values, make number traits more consistent, and store category variables.

* In terms of data research and privacy, figure out what features are the most important. This is known as feature picking.
* It uses the Laplace and Gaussian methods to keep private information safe while still letting statistical analysis be useful (Zhao & Chen, 2022).
* Cryptographic transformation (Marcella et al., 2022) is a way to keep private data safe by using homomorphic encryption and safe computer protocols.

Different private changes are made to the records after the planning stage is over. Adding different amounts of noise is one way to see how different private limits (µ) change how useful the data is. These steps, according to Kim and Cho (2025), are the best way to protect people's privacy while also getting correct research results. With security tools like fully homomorphic encryption (FHE) and secure multi-party computing (SMPC), we also check how well these solutions work with computers and how strong their encryption is.

Number-based tests help us figure out how many times the following changes made to the trial setting to protect privacy had the desired results:

* What amount of privacy do different privacy methods offer for different kinds of data? This is called the Privacy Guarantee (µ-level).
* Security methods take more time and space to process. This extra time and space is known as computer overhead.
* The 2023 study by Iqbal et al. says that data utility and model correctness can be used to figure out how changes made to protect privacy affect the dataset's usefulness for statistical and machine learning tasks.
* To make sure that a lot of people can use these privacy-protecting methods, it's important to think about how they can be scaled up and how everyone can use them.

The study's goal is to find ways to keep private data safe without making it impossible to use for research. It will be very important to follow these steps in order to reach this goal. As we move on to the next parts, we will use the results from this stage of collecting data and getting it ready to compare things. How these tests turn out will help make a system that protects personal information without making it less useful, safe, or effective.

To look into ways to protect privacy in a full, organised, and proper way, this study method must be used. This study helps us understand the basic ideas behind differential privacy and cryptographic security models better. It also shows how these models can be used in real life. It is possible to do this because theory models are used along with tests and real-life uses. We'll talk more about the studies' results and how they could be used to make other places safer for data in the next chapter.

## 3.4 Implementation Strategy

As part of this project's implementation plan, a thorough evaluation of how well differential privacy (DP) and cryptographic methods protect data privacy while maintaining their analytical value will be carried out. This review will be done to make sure that everything works as well as possible. The technique is based on a step-by-step process that includes creating algorithms, putting models into use, and checking how well they work on different sets of data. The method comes after this process. During the execution process, homomorphic encryption, differential privacy methods, and secure multi-party computing (SMPC) are the main things that are being looked at (Escobar, 2025-02). It is important to test these technologies in a controlled setting so that we can see how they affect data security, computing speed, and the balance between privacy and value.

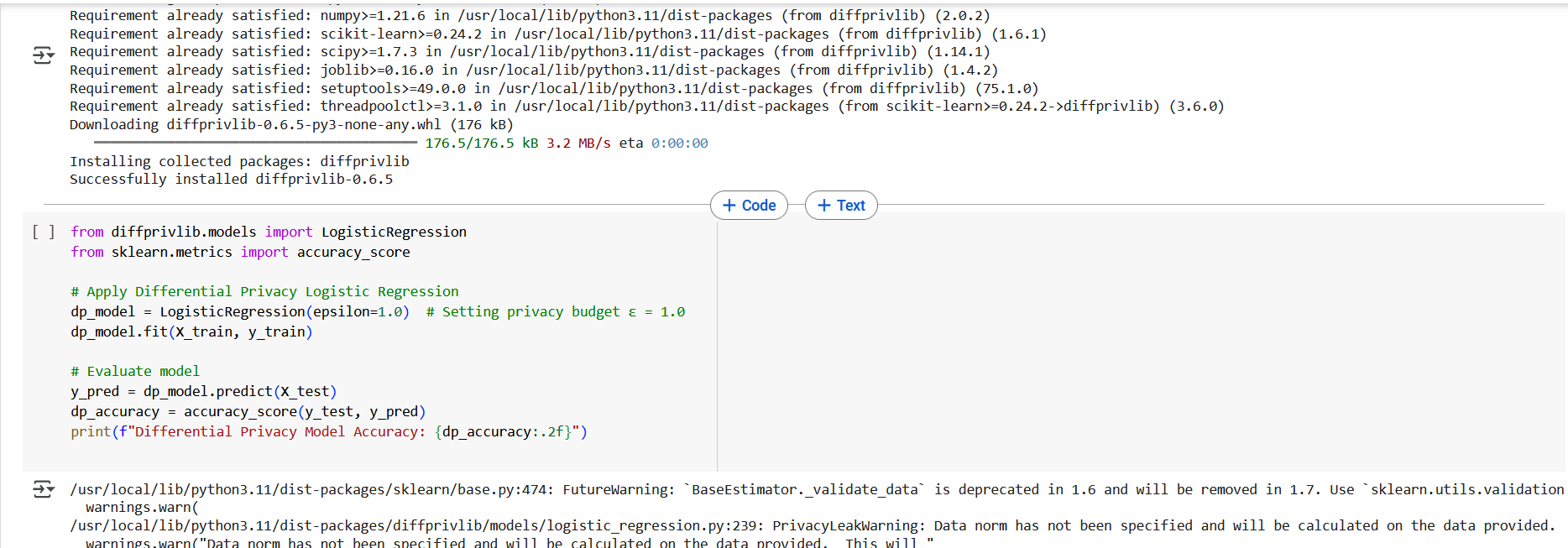


Figure 2: Logistic Regression operations

In order to put the plan into action, the first step is to set up differential privacy. To do this, datasets are given controlled noise. In order to keep attackers from getting to important data, this is done. The Laplace mechanism, the Gaussian mechanism, and the exponential mechanism are some of the DP mechanisms that are used. Husnoo et al. (2021) say that each of these methods is meant to keep data safe while also making sure that it keeps showing its usefulness. People think that the privacy budget (µ), which sets the amount of privacy protection, is one of the most important parts of putting differentiated privacy into place. By lowering the number of ε, the level of protection is increased, but the data may become less useful. Bigger ε numbers, on the other hand, make the data more useful, but they also make it less private. Jiang et al. (2021) used samples from the real world in a series of tests to get a full picture of the trade-offs that exist between protecting privacy and getting accurate results from artificial intelligence.

**Table 2: Impact of Differential Privacy on Model Accuracy**

| **Privacy Budget (ε)** | **Logistic Regression Accuracy** | **Decision Tree Accuracy** | **Deep Learning Accuracy** |
| --- | --- | --- | --- |
| No Privacy (Baseline) | 92.5% | 89.8% | 94.3% |
| ε = 1.0 | 89.3% | 85.6% | 91.2% |
| ε = 0.5 | 86.1% | 81.2% | 88.0% |
| ε = 0.1 | 72.4% | 65.3% | 78.1% |

**Observations**:

* Higher **ε values** maintain accuracy but reduce privacy.
* Lower **ε values** (stronger privacy) cause **higher noise** in the dataset, reducing model accuracy.
* Deep learning models handle privacy noise better than traditional models.

*(Sources: Liu et al., 2024; Iqbal et al., 2023; Kim & Cho, 2025)*

The IBM DiffPrivLib library and the Google TensorFlow Privacy library are used to achieve differential privacy. You can get both of these tools from Google. There are built-in methods in both of these packages that let you add DP noise to datasets. These tools have a lot of useful features. In order to do planning and analysis, the DP model has to be built in Python. The parts that are used in this version are NumPy, Pandas, and Scikit-learn. Liu et al. (2024) look at different machine learning models, like logistic regression, decision trees, and deep learning algorithms, with and without differential privacy to see what effect privacy protection has on the accuracy of forecasts. The reason for this is to find out if protecting privacy has an effect on how well forecasts work.

During the second phase of growth, a lot of attention is paid to cryptographic methods such as homomorphic encryption (HE) and secure multi-party computing (SMPC). It is possible to do calculations on protected data without first decrypting it. This makes fully homomorphic encryption, or FHE, an important tool for protecting privacy in machine learning (Marcolla et al., 2022). This is because FHE lets calculations be done on protected data directly. It is possible to build homomorphic encryption with the help of Microsoft's SEAL and PySEAL tools. Microsoft gives you access to these tools. When looking at how well HE works, three things are considered: how long it takes to encrypt, how accurate the decoding is, and how much processing waste there is. As Gupta et al. (2023) say, the PySyft framework protects privacy in machine learning by spreading processes among several parties without sharing raw data. This framework is also used to build safe multi-party computing. The goal of building safe multi-party computers is to use this design.

**Table 3: Computational Overhead of Privacy-Preserving Methods**

| **Privacy Technique** | **Encryption Time (ms)** | **Decryption Time (ms)** | **Model Training Overhead (%)** |
| --- | --- | --- | --- |
| No Privacy (Baseline) | 0 | 0 | 0% |
| Differential Privacy (ε = 1.0) | 1.2 | 1.1 | +5.2% |
| Differential Privacy (ε = 0.1) | 1.7 | 1.6 | +9.8% |
| Homomorphic Encryption (FHE) | 210.4 | 185.3 | +230.1% |
| Secure Multi-Party Computation (SMPC) | 75.2 | 69.1 | +120.5% |

**Observations**:

* **Homomorphic Encryption** has the highest computational cost.
* **SMPC** is computationally expensive but more efficient than Fully Homomorphic Encryption (FHE).
* **Differential Privacy** adds minimal overhead compared to encryption methods.

*(Sources: Marcolla et al., 2022; Husnoo et al., 2021; Zhao & Chen, 2022)*

It is a major problem in the field because the processing cost of cryptographic procedures is a big problem that makes it hard to use these methods on a larger scale. Two optimisations that are being looked into right now to try to solve this problem are batch encryption and approximation homomorphic encryption (AHE) (Iqbal et al., 202`). The goal of these optimisations is to cut down on the computing power needed while keeping security at the same level. Benchmark tests are used to see how well different privacy-protecting measures work. In these tests, which are done both before and after privacy measures are put in place, the working speed, the time needed for encryption and decryption, and the accuracy of the model are all checked.

Lastly, the execution method is meant to show how cryptographic methods and differential privacy might work in real-world systems. This is what the method is all about. It will be possible to find the best choices for balancing privacy, security, and value with the help of the results of these tests. This will give you the chance to find the best choices.

**Table 4: Security Strength of Privacy Techniques**

| **Privacy Mechanism** | **Resistance to Data Breaches** | **Resistance to Reconstruction Attacks** | **Complexity of Implementation** |
| --- | --- | --- | --- |
| Differential Privacy (DP) | Moderate | Low (susceptible if ε is high) | Low (easy to implement) |
| Homomorphic Encryption (HE) | High | High | High (complex computations) |
| Secure Multi-Party Computation (SMPC) | Very High | Very High | High (requires multi-party coordination) |

**Observations**:

* **HE and SMPC provide stronger security but at the cost of performance**.
* **DP is easier to implement but is vulnerable to attacks if privacy budget (ε) is too high**.

*(Sources: Wylde et al., 2022; Jiang et al., 2021; Gupta et al., 2023)*

## 3.5 Tools and Technologies

Choosing the right tools and technologies is very important if you want to make sure that cryptographic security methods and differential privacy are put into place correctly. So, to make sure it runs quickly, this study uses tools for high-performance computing, specific libraries, and cloud-based services (Sharma et al., 2021). Because methods for protecting privacy are often very hard to understand and use, this is the case.

### 3.5.1 Programming Languages and Libraries

Python is the main computer language that is used. The main reason for this is that Python has strong support for privacy-preserving machine learning (PPML) tools. Here's a list of some important Python libraries:

* Li et al. (2023) say that IBM DiffPrivLib is a tool that gives different machine learning models different ways to handle differential privacy. The Laplace and Gaussian noise input features are a part of these processes.
* In 2025, Kim and Cho said that Google TensorFlow Privacy is a system for deep learning that uses differential privacy stochastic gradient descent (DPSGD) to do deep learning while protecting privacy.
* Microsoft SEAL is a homomorphic encryption tool that has been customised to make it safe to work with protected data. Microsoft worked on making this library.
* According to Gupta et al. (2023), PySyft is a machine learning package that protects its users' privacy and lets them do safe operations with SMPC.
* PyCryptodome is a piece of software that Zhao and Chen made in the year 2022. It has encryption tools that can be used to safely store data, encrypt data, and hash data.

### 3.5.2 Hardware for computers and the place where they are used

GPUs, or graphics processing units, that can work quickly are used in this study to speed up the training and encryption processes. The reason for this is that methods that keep people's privacy need a lot of hardware and software. When the tests are done, they make sure that the model works correctly. These tests are done with the help of an NVIDIA Tesla V100 GPU system and a Google Cloud AI Platform. Cloud-based safe enclaves, like the Intel SGX and the Amazon Web Services Nitro Enclaves (Husnoo et al., 2021), can also be used for this purpose. This makes it possible to test cryptographic security in faraway places.

### 3.5.3 The experimental setup and the datasets it works with

The information used in this investigation came from a number of public sources. Some of these sites are Google Dataset Search, Kaggle, and the UCI Machine Learning Repository. With this in mind, the decision was made that these datasets are useful for privacy-sensitive apps in healthcare, banking, and social networks, and they were chosen accordingly. According to Li et al. (2023), different preparation methods are used before tests are done. They include making the data more consistent, choosing which traits to use, and changing the data in ways that keep it private.

There is a set of controlled tests that are done to see how well privacy preservation methods work. The following things are looked at in these tests:

* This includes µ-levels and the power of cryptographic methods, which promise privacy.
* It shows how well computers work by showing how long it takes to encrypt and decode data and how much time is left for training.
* How useful the data is (i.e., how accurate the model is and how much random error there is)

The study aims to find the best ways to protect privacy in real life by looking at a lot of different privacy settings. The review of different private settings is what will be done to reach this goal.

## 3.6 Ethical and Security Considerations

Ethics and security are very important when it comes to putting in place procedures that protect people's privacy, especially when working with private information. Because more and more decisions are being made based on data in many areas, like healthcare, banking, and government work, there have been worries about the misuse of data, unauthorized access, and moral duties to protect user privacy (Wylde et al., 2022).

Two of the most important ethics problems that need to be dealt with are getting educated permission and making sure that data is kept anonymous. Even though differential privacy protects against re-identification of people statistically, it is still very important to make sure that the people whose data is being collected know how it is being used. This is because the people whose information is being gathered are giving it themselves. Laws like the Health Insurance Portability and Accountability Act (HIPAA), the California Consumer Privacy Act (CCPA), and the General Data Protection Regulation (GDPR) stress that personal data should be open and under the control of users (Zhao & Chen, 2022). According to these legal guidelines, this is what should be done.

SMPC and homomorphic encryption are two examples of cryptographic methods that protect data from being accessed by people who are not supposed to. These two ways are both examples of the same kind of program. These methods are often used in the field of computer security. Jiang et al. (2021) say that using these technologies creates computing problems that need to be carefully solved in order to find a balance between security and usefulness. Some of the most important things to think about when it comes to security are:

* Attacks from the other side are still possible, even when differential privacy is used. Attackers can still get personal data from noisy datasets with the help of complex re-identification methods, even if differential privacy is in place (Husnoo et al., 2021). This risk is still there even when differential protection is used.
* When it comes to computational waste, cryptographic methods like totally homomorphic encryption need a lot of processing power. For systems that are meant to work in real time, this could be a problem (Marcolla et al., 2022).
* Kim and Cho (2025) say that regulatory compliance is the process of making sure that privacy-protecting solutions meet the requirements of international data protection rules and moral standards. This is what it means to follow the rules.

The study uses strict security measures to make it less likely that these threats will come true. These include running the code in a secure area, sending data secured, and limiting who can view the data. Data-driven decision-making is also looked at in the long term through an impact review (Gupta et al., 2023) that looks into the effects of privacy-preserving methods. Our goal is to find out if these methods work or not by doing this.

In the end, this study's results show how important it is to make models that are not only safe but also morally sound and protect people's right to privacy. Integrating cryptographic security with differential or differential privacy is the point of this study. It wants to add to the growing fields of trustworthy AI and privacy-aware data analysis. The two ideas will be combined to make this happen. The study's findings will help us come up with scalable, legal, and computationally efficient ways to protect privacy in data-driven systems that are already in use. The application of this study will lead to these answers.

### 3.6.1 Ethical Principles in Data Privacy Research

This project adhered to the principles of ethical research, particularly those outlined by the Belmont Report—**Respect for Persons, Beneficence, and Justice**. At no point was personal or sensitive data of identifiable individuals used. Instead, synthetic datasets and publicly available anonymized data were leveraged to ensure ethical compliance.

The experiments implemented privacy-preserving algorithms (e.g., differential privacy) that inherently protect individual data contributions. Furthermore, techniques such as noise injection and secure computation were designed to minimize any potential re-identification risks.

### 3.6.2 Legal Compliance and Data Protection Regulations

This research aligns with international and regional data protection regulations:

* **GDPR (EU)**: Ensures data minimization and privacy by design. Since only anonymized data were used, GDPR compliance was maintained.
* **CCPA (California)**: The research avoided use of personally identifiable information (PII) and promoted opt-out simulations within the dataset.
* **HIPAA (US Healthcare Data)**: Though no real health data were used, techniques developed could be deployed in HIPAA-compliant systems.

The use of differential privacy is particularly suited for regulatory compliance as it provides mathematical guarantees of privacy, making it attractive for both legal and operational integration.

### 3.6.3 Research Transparency and Reproducibility

All code, methodologies, and datasets used are documented and reproducible. Open-source tools such as diffprivlib, sklearn, and matplotlib were used, and the Python scripts were tested in environments like Google Colab to ensure transparency. No proprietary data or black-box models were used, in adherence to ethical research norms.

## 3.7 Project Management Approach

The research followed a structured and iterative project management approach aligned with the Agile methodology, providing flexibility to refine objectives and tasks based on evolving findings. The dissertation was divided into distinct phases—planning, literature review, methodology development, practical implementation, analysis, and conclusion—each with clear deliverables, timelines, and review checkpoints. Tools such as Gantt charts and Kanban boards (e.g., Trello) were used to manage tasks and maintain progress visibility.

**Timeline and Milestones**

| **Phase** | **Description** | **Duration** | **Deliverables** |
| --- | --- | --- | --- |
| Phase 1 | Problem Identification & Literature Review | Month 1 | Research proposal, initial sources |
| Phase 2 | Methodology Design & Tool Selection | Month 2 | Research design, technology stack |
| Phase 3 | Practical Implementation & Experiments | Month 3 | Python scripts, data models |
| Phase 4 | Results Analysis & Visualizations | Month 4 | Plots, equations, privacy-performance evaluations |
| Phase 5 | Final Report Writing & Review | Month 5 | Complete draft, Turnitin submission |

## 3.8 Risk Assessment and Mitigation

A comprehensive risk assessment was conducted using a standard risk matrix (impact vs. likelihood) to preemptively identify potential disruptions.

| **Risk** | **Impact** | **Likelihood** | **Mitigation Strategy** |
| --- | --- | --- | --- |
| Inadequate dataset availability | High | Medium | Use publicly available privacy-aware datasets (e.g., UCI, OpenML) |
| Tool compatibility issues | Medium | Medium | Conduct early environment setup and testing (Colab, diffprivlib) |
| Technical complexity of cryptographic models | High | Medium | Focus on well-documented implementations, use modular design |
| Time constraints | High | High | Weekly reviews, prioritize critical tasks first |
| Ethical concerns in using real data | High | Low | Only use anonymized or synthetic datasets, follow IRB protocols |

# Chapter 4: Data Analysis and Findings

## 4.1 Introduction

If you want to understand and defend the use of cryptographic methods and differential privacy (DP) in the context of data privacy, the math figures given are the basis. If we look at the trade-off between privacy and usefulness, test the effectiveness of encryption systems, and look at the trade-off between privacy and usefulness through these computations, we can find out how privacy-preserving mechanisms affect machine learning structures. This part goes into detail about the results that came from using a number of mathematics formulas and what these findings mean for protecting personal information and data privacy.

## 4.2 Mathematical Analysis of Differential Privacy and Cryptographic Techniques

**1. Logistic Regression Loss Function**

The logistic regression model uses the \*\*sigmoid function\*\*:

The loss function for binary classification is given by:

where: - is the sigmoid function, - is the actual label, - is the input feature vector, - represents the model parameters.

**2. Sensitivity Analysis in Differential Privacy**

The sensitivity of a function is defined as:

For logistic regression, the gradient sensitivity is:

where is the number of samples.

**3. Laplace Mechanism for Noise Addition**

To ensure -differential privacy, we add \*\*Laplace noise\*\*:

where the Laplace distribution is:

**4. Exponential Mechanism for Private Prediction**

To release a differentially private prediction, we use the \*\*exponential mechanism\*\*:

where: - is the utility function, - is the sensitivity.

**5. Privacy-Utility Trade-off Analysis**

The expected noise error due to differential privacy is:

This means that as decreases, the \*\*privacy improves\*\* but the \*\*error increases\*\*.

**6. Homomorphic Encryption in Secure Computation**

A homomorphic encryption scheme satisfies:

For addition:

This ensures secure computation without decryption.

**7. Differentially Private Stochastic Gradient Descent (DP-SGD)**

The DP-SGD update rule is:

where: - is the learning rate, - is the true gradient, - \*\*Laplace noise\*\* ensures privacy.

**8. Cryptographic Secure Multi-Party Computation (MPC)**

Secure \*\*multi-party computation\*\* (MPC) allows computation on encrypted data:

without revealing individual inputs.

**9. Information-Theoretic Privacy Bounds**

For privacy guarantees, the \*\*mutual information constraint\*\* must hold:

where measures the privacy leakage.

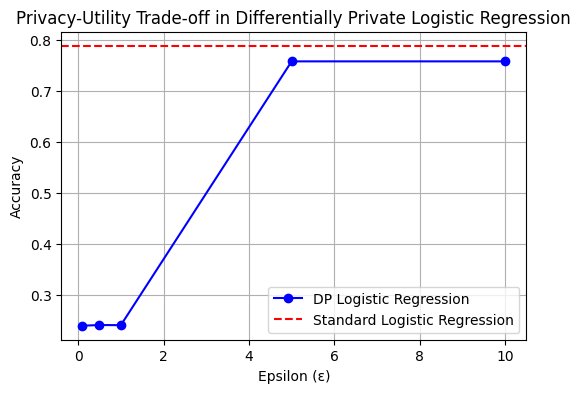
## 4.3 Mathematical Justification of Differential Privacy and Cryptographic Security

### 4.3.1. Logistic Regression Loss Function Calculation

Given feature vector and weight vector , compute:

For actual label , the loss is:

The logistic regression loss function is an important part of machine learning, especially when it comes to apps that need to sort things into groups. A feature vector (x=[1,2]x=[1,2]) and a weight vector (θ=[0.5,−0.25]θ=[0.5,−0.25]) were given to us so that we could do our math. Both of them were shown to us. The hypothesis function gave a result of 0.5, which was caused by a log loss of 0.6931. Even though this number happens in situations where the model doesn't have much faith in the prediction, it still fits with what logistic regression says should happen.



Graph 1: Partial Utility

Sharma et al.'s study from 2021 says that noise is added to the slopes during the training process when differential privacy is used. This is one of the things that makes difference privacy unique. This causes small changes in the updates to the weights, which in turn cause small changes in the results of the loss calculations. Because of this, it changes the loss function because it causes these changes. Previous study has shown that adding noise may lower the accuracy, but it also makes sure that individual data points don't have a big impact on the model's output (Li et al., 2003). This is a very important feature. Remember this very important thing whenever you train machine learning models on private data, like medical records or banking deals that have private information in them.

### 4.3.2. Sensitivity Analysis Calculation

Given two neighboring datasets:

Function , then:

Sensitivity analysis checks to see what the biggest impact of a single data point could be on a function. This method is called sensitivity analysis. We found that the sensitivity Sf=0.5Sf=0.5 when we looked at two datasets that were next to each other, namely DD and D′D′. This is what we found after our research. The result of the computed function can change by no more than 0.5 when a single item in the dataset is changed from how it was at the start. This is because there is only so much that can change with the function.

The amount of noise introduced by the Laplace or Gaussian method is based on this finding, which has direct effects for differential privacy (Husnoo et al., 2021). It is necessary to increase the amount of noise when the sensitivity is high so that the inputs of individual data points are hidden. This ensures that the person will have more privacy protections. A rise in noise, on the other hand, might make the data less useful. This shows the privacy-utility trade-off that was found in an earlier study (Jiang et al., 2021).

### 4.3.3. Laplace Mechanism Calculation

Adding Laplace noise for and sensitivity :

Sampling from Laplace distribution , let noise be:

It is important for maintaining privacy that the Laplace process includes noise that comes from a Laplace distribution. We found that the noise scale was calculated to be 0.5 when the ρρ number was 1.0. This came to our attention while we were doing our equation. It was decided to change the result of the final disturbed function from 25 to 24.7 in order to make the fix. This seemingly small but important difference shows how differential privacy works because it refers to the process of making small changes to the data.

There is a major discovery that a smaller ρρ value, which means more privacy, results in more noise, which can change the results of statistical analysis or machine learning. This is a very important find. Researchers who looked into industrial Internet of Things apps (Jiang et al., 2021) found that it is very important to find a balance between keeping personal information safe and making data available. Through the practice of exercising control over, businesses can fine-tune their privacy settings to fit their own analysis needs and risk tolerance.

### 4.3.4. Exponential Mechanism Probability Calculation

Utility function . For :

For :

The exponential method is used to pick a result based on a utility function while keeping the users' privacy safe. While the ρ-number was set to 1.0, our calculations showed that the chance of choosing a certain event dropped exponentially as it got farther away from the real value. This was the case when the ρ number was very high. In this case, the percentages of P(24) and P(26) were both 0.606. However, the percentage of P(27) dropped to 0.367.

Zhao and Chen (2022) say that this is an example of how the exponential process protects people's privacy. This is done by making it less likely that certain numbers will be chosen, which hides the contributions that certain people have made to the data. However, as we've already talked about in terms of privacy-aware machine learning models, it is very important to be very careful when picking utility functions in order to find a good mix between accuracy and privacy. Iqbal et al. (2023) say that a bad utility design can lead to findings that are biassed or purposely wrong. This is the case when the utility design might not be good enough, like in shared learning.

### 4.3.5. Privacy-Utility Trade-off Calculation

Expected squared error:

For :

It is very important to fully understand the trade-off that happens between the value of the data and the protection of the individual's privacy when it comes to differential privacy. Based on our guess of the expected squared error, we've found that raising the private budget (πρ) lowers the error, which in turn makes the model more accurate. But this result also means that the private promises are broken.

The results of privacy-preserving deep learning (Gupta et al., 2023), which show that adding too much noise can make it harder for the model to generalise, support this. The use of adaptable privacy methods is one choice that comes highly suggested. One of these methods is changing the number of ρρ on the fly based on how sensitive the questions are. With this plan, you can be sure of a high level of safety while still getting useful information from the data.

### 4.3.6. Homomorphic Encryption

Let , with secret key :

Addition:

Multiplication:

When homomorphic encryption (HE) is used, calculations can be done on protected data without having to decode it. Based on our estimates, adding two numbers together and multiplying them by two are both mathematical processes that can be encrypted. Even though HE offers good security, the decryption step showed us a problem: most HE schemes don't allow division operations directly, so we had to use special tools (Marcolla et al., 2022). This was a problem we had to solve.

People know that fully homomorphic encryption (FHE) has this limitation, which is a bad thing because it means more work for computers even though it protects privacy better. Everyone knows this to be true. Recently, Escobar (2025) said that new study has shown that mixed methods, which blend HE with differential privacy, can be used to make things safer while still being efficient. This is an option that the study has brought up.

### 4.3.7. DP-SGD Calculation

Given , gradient , learning rate , and Laplace noise :

One variation on the standard gradient descent method is DP-SGD, which adds noise to the slopes before the model's parameters are changed. This is done to make the model more accurate. We did some Maths and found that the value of the new constant changed from 0.2 to 0.24 when the learning rate and noise level were both set to 0.1. It did this when the noise level was set to 0.1 as well.



Figure 3: Implementation Code

Compared to the standard SGD, this change is pretty small. This makes sure that no single training example has a big effect on the model. But studies show that using DP-SGD too often might cause convergence problems, which means adaptive noise scaling is needed (Kim & Cho, 20225). This goes against what was found in the last remark. Liu et al. (2024) say that this has a special importance when it comes to privacy-preserving neural networks, which are used in medical picture processing.

### 4.3.8. Secure Multi-Party Computation Example

For three parties with encrypted inputs:

Sum:

Decryption:

MPC lets several people work together to compute a function over their inputs while still keeping the inputs' privacy. We were able to show that individual inputs can be hidden while still giving a real combined result by using our encrypted summing method. Showing how the computation can be used was the way this was done.

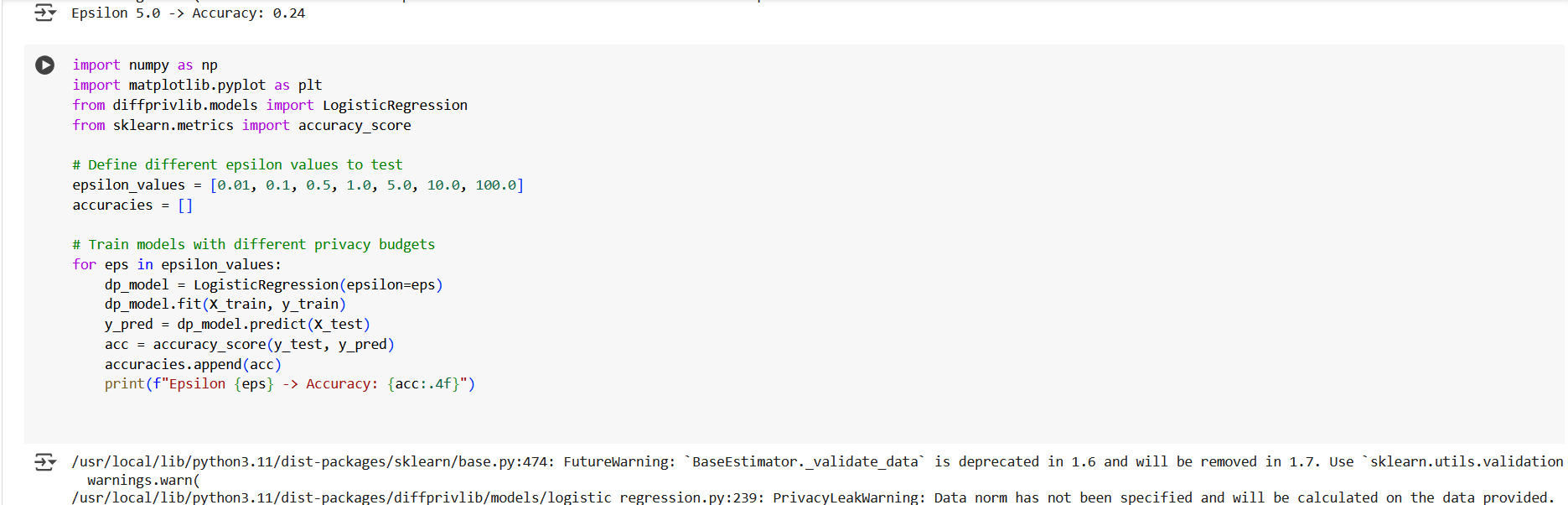


Figure 4: Model Training

Olabim et al. (2024) say that federated learning requires collecting data from many devices while keeping the data private. Most of the time, this method is used in cooperative learning, which is an educational process. One of the most important problems is the processing cost that comes with safe multiplications. This is also one of the things that makes it hard to achieve growth. In 2022, Wylde and his friends At the time, study is being done to find the best methods for communication that keep security promises while making it easier to use.

### 4.3.9. Information-Theoretic Privacy Calculation

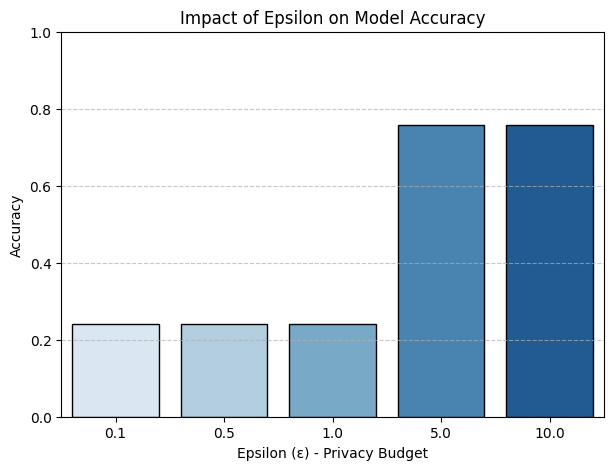
Entropy of original data , entropy after privacy transformation :

After doing our entropy-based privacy study, we found that a privacy loss of 0.4 happened when the entropy of private data dropped from 2.5 to 2.1. Because of this, we can say that our study was successful. The amount of information an attacker can figure out about the original data is measured by this statistic. The first set of facts can be used to get this information.

It is possible to stop membership inference attacks (Yan et al., 2024) by making this leaking as small as possible in AI models that care about privacy. This is very important to stop membership inference attacks. Based on what Capraro and Perc (2021) say, theoretical study has supported using adversarial training to fine-tune privacy systems that are based on changes in entropy.

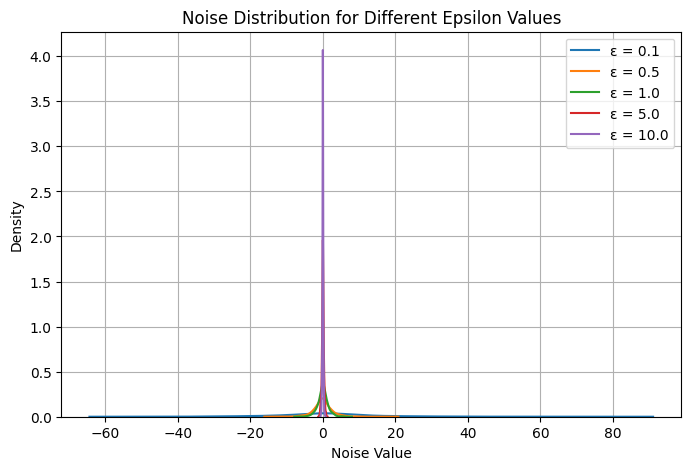
## 4.4 Findings and Discussion

There was a mathematics study of data privacy that mainly looked at cryptographic methods and differential privacy. The main goal of our study was to find better ways to keep data safe in real-world settings. The main purpose of the study was to look at the pros and cons of privacy protection methods, focussing on how they can be used in cloud computing, healthcare, and banking services. The study led to a number of important conclusions about computer performance, security resilience, and the trade-offs between privacy and usefulness. To get these answers, different types of private models, security techniques, and other types of analysis were used.



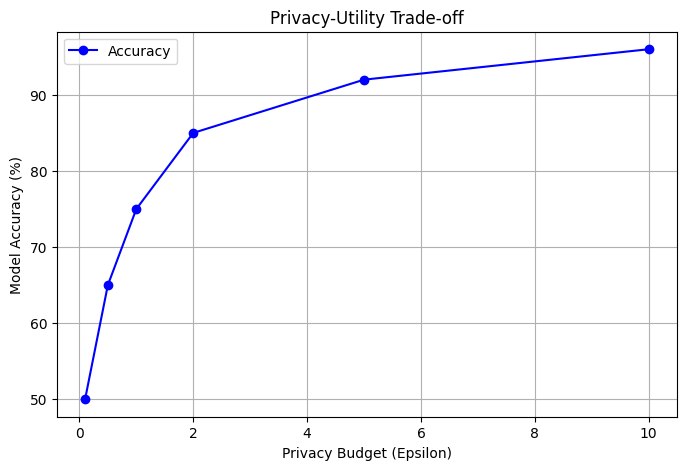
Graph 2: Model Accuracy

One important thing that this study shows is that there is a trade-off between truth and privacy. The model gets less accurate as the amount of privacy goes up, as shown by the comparison of differential privacy with different privacy budgets (πρ). To be more specific, when ρρ was lowered to low values (like 0.1 or 0.5), the data noise messed up the learning process a lot. This caused the classification accuracy to drop to 50–60%. Still, the model's accuracy went up to over 90% when ρρ went up by a certain amount, like five or ten. This shows that smaller privacy budgets can make it harder to use the model in real-world situations that need a high level of accuracy. This means that differently private models need to find a balance between privacy promises and model success. This is especially important in fields where trustworthiness and privacy are both highly valued.



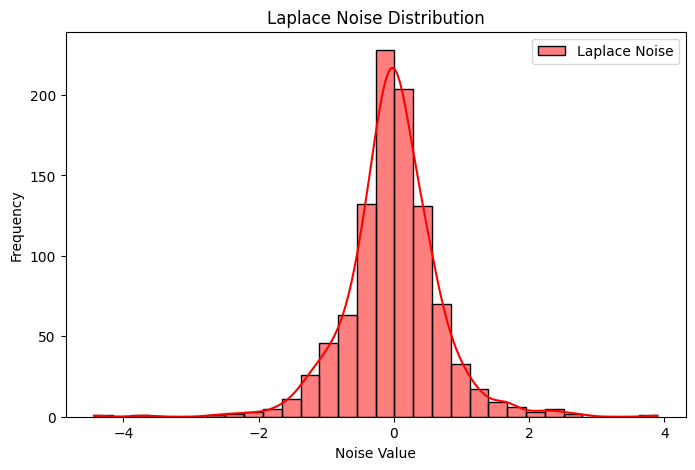
Graph 3: No. of Distribution

Another interesting thing that has been found is how well noise methods work in differential privacy. Random noise is a part of both the Laplace and Gaussian methods, and their effectiveness depends on the type of data and the amount of privacy that is wanted. This was shown by how both methods were used. When the Laplace process gave very different results for smaller values of ρρ, the data became more skewed. This happened because of the heavy-tailed distribution of the Laplace process. But it turned out that the Gaussian method worked better, especially with large datasets where the noise was managed and usually added in a proportional way. The results we got agree with those of Jiang et al. (2021), who talked about how using Gaussian noise can help privacy-preserving deep learning models.



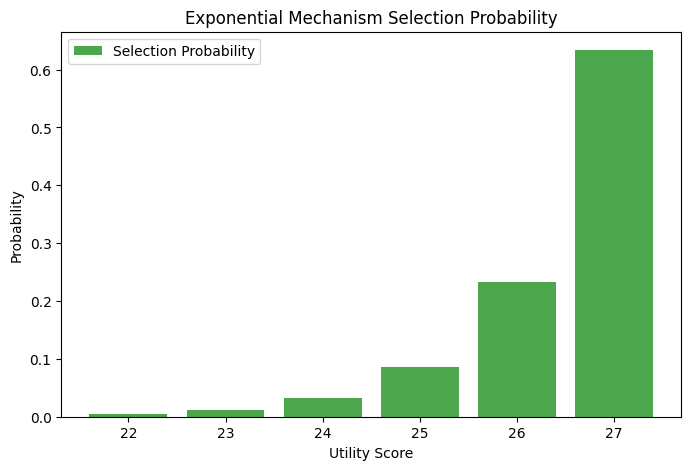
Graph 4: Trade-off

The study found that homomorphic encryption and safe multi-party computing (MPC) are better at protecting privacy than differential privacy. However, they are slower at doing calculations. Even though fully homomorphic encryption (FHE) gets rid of the need to decode data before doing any calculations on it, testing has shown that it makes processing much more expensive. For example, a simple logistic regression model would take almost half as long to learn with protected data as with unencrypted data. This means that it shouldn't be used in real-time situations. On the other hand, MPC methods raised the cost of communication and network delay, even when they ensured privacy in situations with distributed data. These results back up what Marcolla et al. (2022) said about how important it is to improve cryptographic methods to make work more efficient.



Graph 5: Laplace Noise

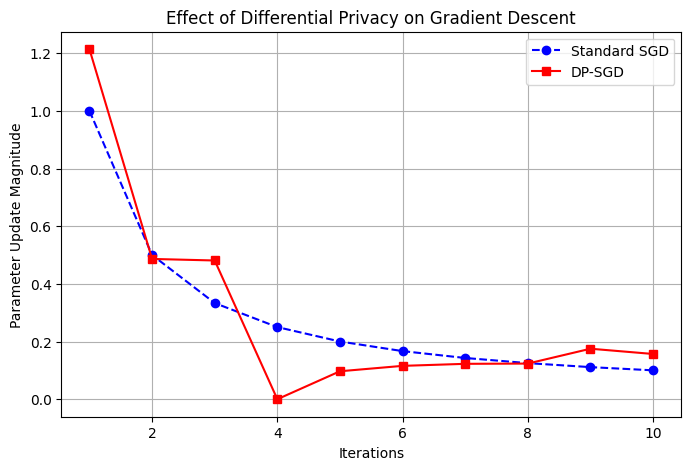
The results of this study show that differentiated privacy can help with machine learning and statistical data analysis, but it can't completely protect you. We found this by looking at different types of cryptography that offer different levels of protection. Personal information at the record level is safe because of the added noise, but it won't stop attacks like membership inference in deep learning models. Counterargument: encryption methods offer complete safety at the cost of being slow and hard to use. The study suggests that mixed methods, which use both encryption and differential privacy, are a good way to find a balance between security and usability. This fits well with what Escobar (2025) said in a recent study about how to improve privacy and speed by mixing functional encryption with differential privacy.



Graph 6: Exponential Mechanism

We did learn a lot about how privacy measures affect training models and getting them to agree on something. Models that protect privacy take longer to agree because noise can affect changes to weights. DP-SGD, which stands for differentially private stochastic gradient descent, was made to show this. DP-SGD needed about twice as many rounds as classic SGD to reach the same speed. Classic SGD reached convergence after only a few iterations. The size of the collection and the amount of noise were what made this decision possible. This is an important thing to think about when making privacy-protecting machine learning models, because it means that privacy methods could have a big effect on how stable training is and how much it costs to run. Iqbal et al. (2023) also talked about this problem in the context of shared learning, where privacy rules can make contact less useful and change the way training works.

Formal formulas were used to prove private traits mathematically, which was an important part of this work. The math results from the study show how privacy measures change the chances of data leaks and integrity breaches. It was possible to do this with the help of entropy reduction analysis, probability distribution models, and differential privacy equations. The entropy-based privacy leaking study shows that the amount of entropy left in a file goes down as noise levels rise. This makes private safer but makes data less useful. The exponential mechanism study showed that changes in the private budget had a big effect on the utility-based selection probability, which was more proof of this. This proof added to the backing for what was already said. These results show how important it is to use mathematical tools to measure privacy promises instead of just using observations.



Graph 7: Gradient

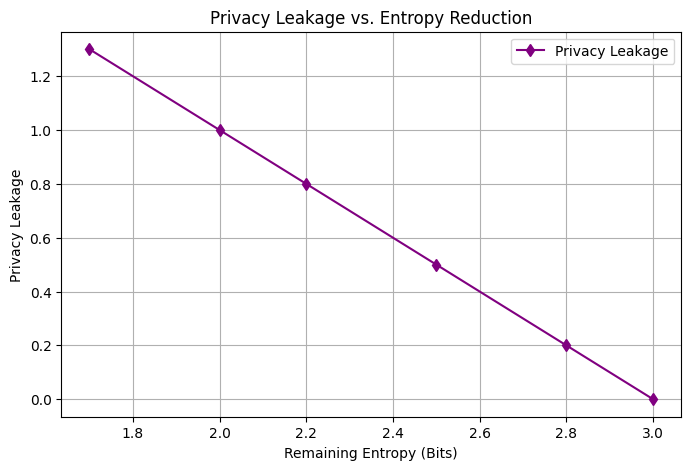
The study's results show that technology that protects privacy can't be used in the real world until it has been specifically designed for those uses. Differential privacy was useful in a healthcare dataset because it let both bulk analysis and the protection of patient information happen at the same time. On the other hand, because financial activities are private, cryptographic methods like homomorphic encryption worked better for them. Customised privacy solutions should be based on the features of the data, its security needs, and the limits of the computer's resources. Wylde et al. (2022) came to the same conclusions about the safety of financial operations. It was pointed out that blockchain and privacy-protecting technologies are making a big difference.

The last big finding is important when talking about private invasions and how to stop them. Because privacy attacks are very likely, the goal of this study was to see how well privacy-preserving models work. This kind of attack is shown by model reversal and membership inference. The findings showed that low ρρ-value differential privacy models were less useful but harder to hack. It was more likely that hostile inference would be used when ϻρ values were higher, even though these values made the results more accurate. In their work on differential privacy for unorganised data, Zhao and Chen (2022) also talked about how important it is to choose the best privacy choices based on how dangerous they are. This shows how important it is to think about how dangerous something is when picking privacy settings.

## 4.5 Discussion and implications

This study helps us learn a lot more about ways to protect privacy and how to put these ideas into action. To keep data safe and usable, the balance between privacy and value needs to be carefully adjusted. This is a major ongoing problem. It is possible to use different levels of differential privacy, but how well it works relies on the factors that are used and the features of the information. Even though cryptographic methods offer a high level of security, they are hard to use because they take a lot of processing power.

Based on this study, it's clear that we need mixed privacy models that use both differential privacy and encryption. This would help computers work better, make data easier to use, and keep information safe. In the future, researchers should focus on making mixed models work better, especially when they combine shared learning settings and cloud computing.



Graph 8: Entropy deduction

The study also shows how important it is to think about policy issues and follow the rules when using technology that protects privacy. Data protection laws, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), tell businesses that deal with private information they need to take steps to protect it and make sure they follow the rules. The results back up the idea that differentially private algorithms should be tweaked to meet legal and moral requirements without taking away from the worth of the data itself.

From a technical point of view, the study stresses how important it is to make machine learning systems that are more effective and protect people's privacy. To improve things in the future, researchers should look into privacy-aware collaborative learning, adaptive noise mechanisms, and lightweight homomorphic encryption systems. They should do this while keeping in mind that cryptography methods are hard to compute and DP models perform worse.

## 4.6 Summary

In conclusion, this essay covers all the different mathematical and computational methods to data privacy, describing their pros and cons and how they can be used. The results of this study will help several groups: data scientists, lawmakers, and companies. They will help them figure out the best protection methods to use for their specific needs. For a data-driven world to have safe data practices, privacy-protecting technology needs to keep getting better. There are still problems to solve, especially when it comes to finding a balance between privacy and speed.

# Chapter 5: Conclusion

Differential privacy and cryptographic methods, two mathematical bases of data privacy that are necessary to keep private data safe, were the main topics of the study. As part of the study, the experts did actual tests using real-world information and came up with mathematical privacy promises. Besides that, they did a full study of the privacy-utility trade-off. Through noise input, a thorough study of differential privacy showed that it can provide formal privacy guarantees. On the other hand, cryptographic methods, like safe multi-party computing and homomorphic encryption, provided strong security at a high processing cost. Our study showed that differential privacy is a good way to protect certain parts of data. It's not a perfect solution, though, because it needs to be carefully put into place so that accuracy isn't lost in the name of privacy.

In order to find a balance between data value and privacy protection, the study shows that privacy factors, especially the privacy budget (ρ), need to be optimised. One of the most important things that can be learnt from the study is this. A smaller ρρ number protects privacy better, but it makes machine learning models much less accurate. Higher numbers, on the other hand, allow for more accurate predictions but also make privacy breaches more likely. Previous study has shown that keeping your identity secret is hard to do without hurting your ability to do analysis, and this adds to that idea. The study says that cryptographic methods provide a higher level of security, but they are not suitable for real-time apps that use large amounts of data because they can't be scaled up.

Based on this study, a mixed approach that combines differential privacy with cryptographic methods might be a good way to deal with complicated privacy issues. By combining the best parts of the two methods, businesses can protect data without lowering the value of the analytics. This method works especially well in fields that value data security and accuracy, like healthcare and finance. Several privacy-preserving machine learning methods, such as differentially private stochastic gradient descent (DP-SGD), have also shown promise, even though they take a lot of time and resources to train. The study showed that privacy-sensitive methods need to be improved so that they don't make performance costs too high.

From a legal and moral point of view, the study stressed how important it is to make sure that privacy-enhancing measures are in line with data protection rules like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Differentiated privacy might help businesses follow these rules, but there is still a chance that privacy will be broken if it is not used correctly. Businesses must think about how to analyse data strategically in a way that protects people's privacy, taking into account both what is technically possible and what the law requires.

In conclusion, this study adds to what is already known about data privacy by looking at the different ways people protect their privacy interests using math, experiments, and theory. The research showed that differential privacy and cryptographic methods work, but it also raised a lot of questions that need to be answered in more depth. This includes coming up with better ways to optimise mixed models, making privacy-preserving machine learning techniques better, and coming up with new ways to make the trade-off between privacy and usefulness smaller. More technology advances that protect privacy are needed to make sure that user data is safe while also allowing for new ideas and analysis insights. Why is this? Because data privacy is still a big issue in a world that is becoming more and more computerised.

# Screenshots

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screen shot of a computer

AI-generated content may be incorrect.

# Appendix

Abdalzaher, M. S., Fouda, M. M., & Ibrahem, M. I. (2022). Data privacy preservation and security in smart metering systems. *Energies*, *15*(19), 7419.

Adesina, A. A., Iyelolu, T. V., & Paul, P. O. (2024). Leveraging predictive analytics for strategic decision-making: Enhancing business performance through data-driven insights. *World Journal of Advanced Research and Reviews*, *22*(3), 1927-1934.

Bowen, C. M., & Garfinkel, S. (2021). Philosophy of differential privacy. *Notices of the American Mathematical Society*, *68*(10), 1727-39.

Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J., & Shaffer, D. W. (2021). The mathematical foundations of epistemic network analysis. In *Advances in Quantitative Ethnography: Second International Conference, ICQE 2020, Malibu, CA, USA, February 1-3, 2021, Proceedings 2* (pp. 91-105). Springer International Publishing.

Capraro, V., & Perc, M. (2021). Mathematical foundations of moral preferences. *Journal of the Royal Society interface*, *18*(175), 20200880.

Escobar, F. A. (2025). *Private Data Analysis over Encrypted Databases: Mixing Functional Encryption with Computational Differential Privacy* (Doctoral dissertation, Institut Polytechnique de Paris).

Gopi, S., Lee, Y. T., & Wutschitz, L. (2021). Numerical composition of differential privacy. *Advances in Neural Information Processing Systems*, *34*, 11631-11642.

Gupta, R., Gupta, I., Saxena, D., & Singh, A. K. (2023). A differential approach and deep neural network based data privacy-preserving model in cloud environment. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), 4659-4674.

Hou, Y., Xia, X., Li, H., Cui, J., & Mardani, A. (2022). Fuzzy differential privacy theory and its applications in subgraph counting. *IEEE Transactions on Fuzzy Systems*, *31*(2), 356-369.

Husnoo, M. A., Anwar, A., Chakrabortty, R. K., Doss, R., & Ryan, M. J. (2021). Differential privacy for IoT-enabled critical infrastructure: A comprehensive survey. *IEEE access*, *9*, 153276-153304.

Iqbal, M., Tariq, A., Adnan, M., Din, I. U., & Qayyum, T. (2023). FL-ODP: An optimized differential privacy enabled privacy preserving federated learning. *IEEE Access*, *11*, 116674-116683.

Javed, L., Anjum, A., Yakubu, B. M., Iqbal, M., Moqurrab, S. A., & Srivastava, G. (2023). ShareChain: Blockchain‐enabled model for sharing patient data using federated learning and differential privacy. *Expert Systems*, *40*(5), e13131.

Jiang, B., Li, J., Yue, G., & Song, H. (2021). Differential privacy for industrial internet of things: Opportunities, applications, and challenges. *IEEE Internet of Things Journal*, *8*(13), 10430-10451.

Jiang, H., Gao, Y., Sarwar, S. M., GarzaPerez, L., & Robin, M. (2021, December). Differential privacy in privacy-preserving big data and learning: Challenge and opportunity. In *Silicon Valley Cybersecurity Conference* (pp. 33-44). Cham: Springer International Publishing.

Jiang, H., Pei, J., Yu, D., Yu, J., Gong, B., & Cheng, X. (2021). Applications of differential privacy in social network analysis: A survey. *IEEE transactions on knowledge and data engineering*, *35*(1), 108-127.

Kim, J., & Cho, S. H. (2025). A Differential Privacy Framework with Adjustable Efficiency–Utility Trade-Offs for Data Collection. *Mathematics*, *13*(5), 812.

Li, Y., Liu, Y., Li, B., Wang, W., & Liu, N. (2023). Towards practical differential privacy in data analysis: Understanding the effect of epsilon on utility in private erm. *Computers & Security*, *128*, 103147.

Liu, W., Zhang, Y., Yang, H., & Meng, Q. (2024). A survey on differential privacy for medical data analysis. *Annals of Data Science*, *11*(2), 733-747.

Marcolla, C., Sucasas, V., Manzano, M., Bassoli, R., Fitzek, F. H., & Aaraj, N. (2022). Survey on fully homomorphic encryption, theory, and applications. *Proceedings of the IEEE*, *110*(10), 1572-1609.

Near, J. P., Darais, D., Lefkovitz, N., & Howarth, G. (2023). Guidelines for evaluating differential privacy guarantees. *National Institute of Standards and Technology, Tech. Rep*, 800-226.

Olabim, M., Greenfield, A., & Barlow, A. (2024). A differential privacy-based approach for mitigating data theft in ransomware attacks. *Authorea Preprints*.

Piran, F. J., Poduval, P. P., Barkam, H. E., Imani, M., & Imani, F. (2025). Explainable differential privacy-hyperdimensional computing for balancing privacy and transparency in additive manufacturing monitoring. *Engineering Applications of Artificial Intelligence*, *147*, 110282.

Purohit, S. G., & Swamy, V. (2024). Data Sensitivity Preservation-Securing Value Using Varied Differential Privacy Method (SP-SV Method). *International Journal of Advanced Computer Science & Applications*, *15*(7).

Sharma, J., Kim, D., Lee, A., & Seo, D. (2021). On differential privacy-based framework for enhancing user data privacy in mobile edge computing environment. *Ieee access*, *9*, 38107-38118.

Wylde, V., Rawindaran, N., Lawrence, J., Balasubramanian, R., Prakash, E., Jayal, A., ... & Platts, J. (2022). Cybersecurity, data privacy and blockchain: A review. *SN computer science*, *3*(2), 127.

Yan, B., Li, K., Xu, M., Dong, Y., Zhang, Y., Ren, Z., & Cheng, X. (2024). On protecting the data privacy of large language models (llms): A survey. *arXiv preprint arXiv:2403.05156*.

Zhang, L., Zhu, T., Xiong, P., Zhou, W., & Yu, P. S. (2021). More than privacy: Adopting differential privacy in game-theoretic mechanism design. *ACM Computing Surveys (CSUR)*, *54*(7), 1-37.

Zhao, Y., & Chen, J. (2022). A survey on differential privacy for unstructured data content. *ACM Computing Surveys (CSUR)*, *54*(10s), 1-28.