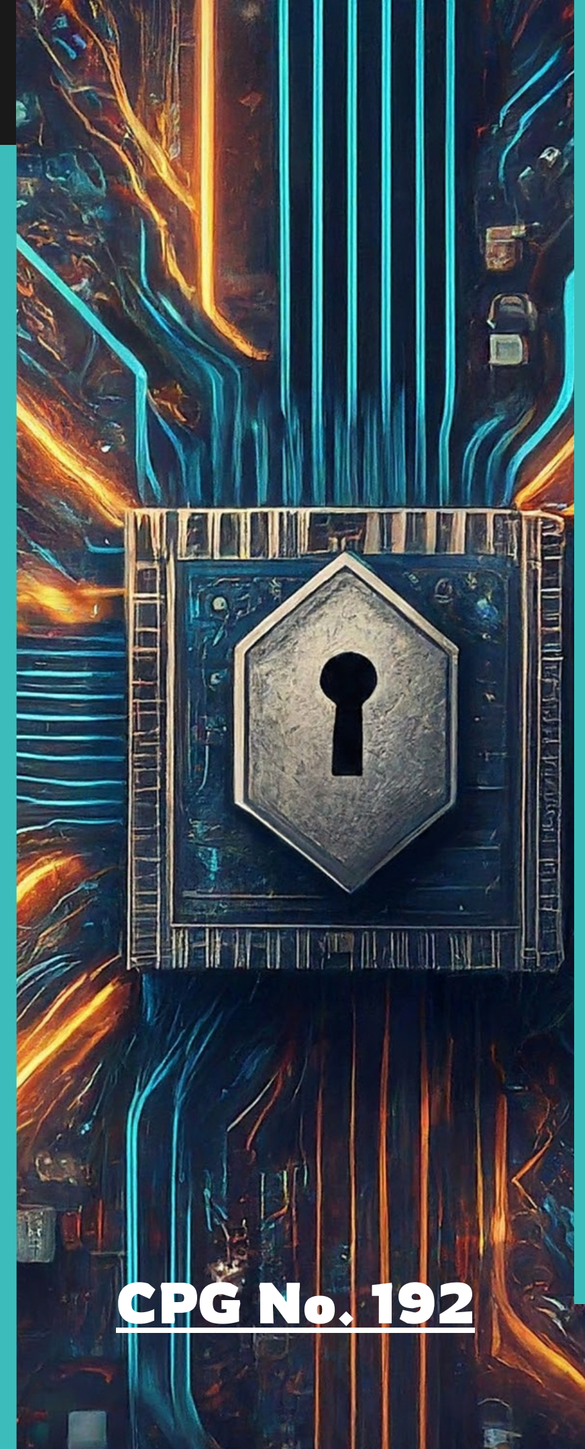
PRIVACY-PRESERVING SENSITIVE INFORMATION MASKING SCHEME



UNDER THE MENTORSHIP OF:

DR. ROHAN SHARMA
(ASSISTANT PROFESSOR)

DR. SHIVANI SHARMA
(ASSISTANT PROFESSOR)

CPG No. 192

TEAM MEMBERS

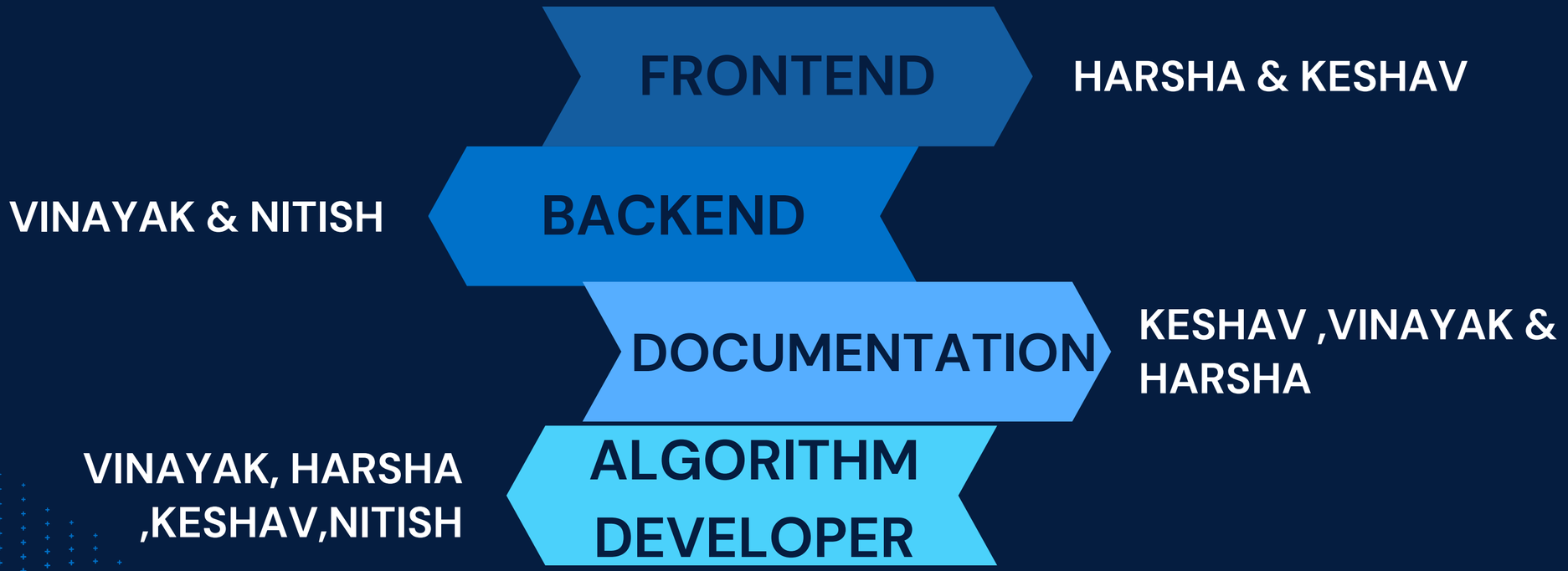
VINAYAK LAL
102116050

S NITISH KUMAR
102166003

KESHAV MITTAL
102116051

HARSHA VARDHAN
102116055

Contributions of Individual Team Members



PROJECT OVERVIEW

The rapid growth of information technology and e-commerce has led to vast data volumes, enabling insights in market trends and healthcare analytics. However, this availability raises privacy concerns, with sensitive data vulnerable to exposure through data mining. Privacy-preserving data mining (PPDM) addresses this, but traditional methods may compromise data utility. Our project aims to develop an algorithm for a privacy-preserving masking scheme that overcomes the drawbacks of already existing algorithms. By using our algorithm, we aim to safeguard sensitive information while maintaining data utility and confidentiality without deleting data and aims at providing near optimal solution.

NEED ANALYSIS

1

PROTECTION OF INDIVIDUAL PRIVACY

2

REGULATORY COMPLIANCE & LEGAL OBLIGATIONS

3

DATA SHARING AND COLLABORATION

4

UPHOLDING TRUST AND ETHICAL CONSIDERATIONS

5

BALANCING PRIVACY PRESERVATION AND DATA UTILITY

LITERATURE SURVEY

1

A Sanitization approach for hiding sensitive itemsets based on particle swarm optimization

This research paper proposes a novel sanitization approach that utilizes PSO for hiding sensitive itemsets in datasets. The approach aims to minimize the distortion introduced to non-sensitive itemsets while effectively concealing sensitive ones. By formulating the problem as an optimization task and leveraging the collective behavior of swarms, the proposed approach demonstrates promising results in preserving data utility while ensuring privacy.

2

Victim item deletion based PSO inspired sensitive pattern hiding algorithm for dense datasets

This paper presents a sensitive pattern hiding algorithm inspired by PSO, specifically designed for dense datasets. The algorithm, called VIDPSO, focuses on identifying and deleting victim items to conceal sensitive patterns effectively. By incorporating PSO's optimization capabilities, VIDPSO achieves significant improvements in terms of privacy preservation and data utility compared to existing techniques.

3

Privacy-preserving data mining using differential privacy

This seminal paper introduces the concept of differential privacy as a framework for privacy-preserving data mining. Differential privacy aims to ensure that the presence or absence of an individual's data does not significantly affect the outcome of data analysis. By adding carefully calibrated noise to query responses, differential privacy offers strong privacy guarantees while allowing for meaningful data analysis.

LITERATURE SURVEY

4

Privacy-preserving data publishing: A survey

This survey paper provides a comprehensive overview of privacy-preserving data publishing techniques. It covers a wide range of approaches, including k-anonymity, l-diversity, t-closeness, and differential privacy. The survey discusses the strengths and limitations of each technique and provides insights into their real-world applicability across various domains.

5

Privacy-preserving collaborative filtering using randomized perturbation-based approach

This research paper proposes a randomized perturbation-based approach for privacy-preserving collaborative filtering. The approach introduces random noise to user-item ratings to protect individual privacy while maintaining the accuracy of recommendation systems. By incorporating perturbation techniques, the proposed approach offers robust privacy guarantees without compromising recommendation quality.

6

Privacy-preserving machine learning: Threats and solutions

This paper provides an overview of privacy threats in machine learning models and proposes solutions for mitigating these threats. It discusses various attacks on machine learning models, such as membership inference and model inversion attacks, and presents privacy-preserving techniques, including differential privacy, federated learning, and secure multi-party computation.

PROBLEM STATEMENT

Given the increasing concerns surrounding privacy and security in the era of big data, there is a critical need for innovative approaches to preserve sensitive information while maintaining data utility. Traditional deletion practices in privacy-preserving data mining (PPDM) may lead to data loss and reduced utility, necessitating the exploration of alternative techniques. Our project aims to address this challenge by developing a novel approach that integrates data aggregation techniques with evolutionary optimization methods such as Particle Swarm Optimization (PSO). By effectively concealing sensitive patterns while minimizing data loss and artificial pattern generation, our solution seeks to strike a delicate balance between privacy preservation and data utility, offering a robust framework for secure data transformation and collaboration with third parties.

OBJECTIVES

1. Developing a Novel Privacy-Preserving Masking Scheme
2. Overcoming the Drawbacks in already existing algorithms
3. Optimization of Masking Parameters
4. Comprehensive Evaluation and Validation
5. Cross-Domain Application and Impact

ASSUMPTIONS

Homogeneous Data Distribution

The project assumes that the sensitive information within the dataset is uniformly distributed and does not exhibit significant skewness or bias.

Access to Representative Data Samples:

It is assumed that the project has access to representative data samples that accurately capture the characteristics and distribution of sensitive information.

Availability of Computational Resources

The project assumes access to sufficient computational resources, including hardware infrastructure and computing resources, to support the development, optimization, and evaluation of the masking algorithms.

Compliance with Privacy Regulations

It is assumed that the developed privacy-preserving masking scheme complies with relevant privacy regulations and legal requirements

CONSTRAINTS

Algorithm Complexity and Scalability

The project faces constraints related to algorithm complexity and scalability Complex masking algorithms may introduce computational overhead and scalability challenges, requiring optimization techniques to enhance efficiency and scalability.

Data Sensitivity and Confidentiality

The project operates within the constraints of data sensitivity and confidentiality necessitating stringent measures to protect sensitive information from unauthorized access and disclosure.

Integration with Existing Systems and Workflows

The project must integrate seamlessly with existing data processing systems and workflows, imposing compatibility, interoperability, and implementation complexity constraints.

METHODOLOGY

1

Data Input and Separation

2

Victim Identification and Selection

3

Identification of Portion of Data to be Modified

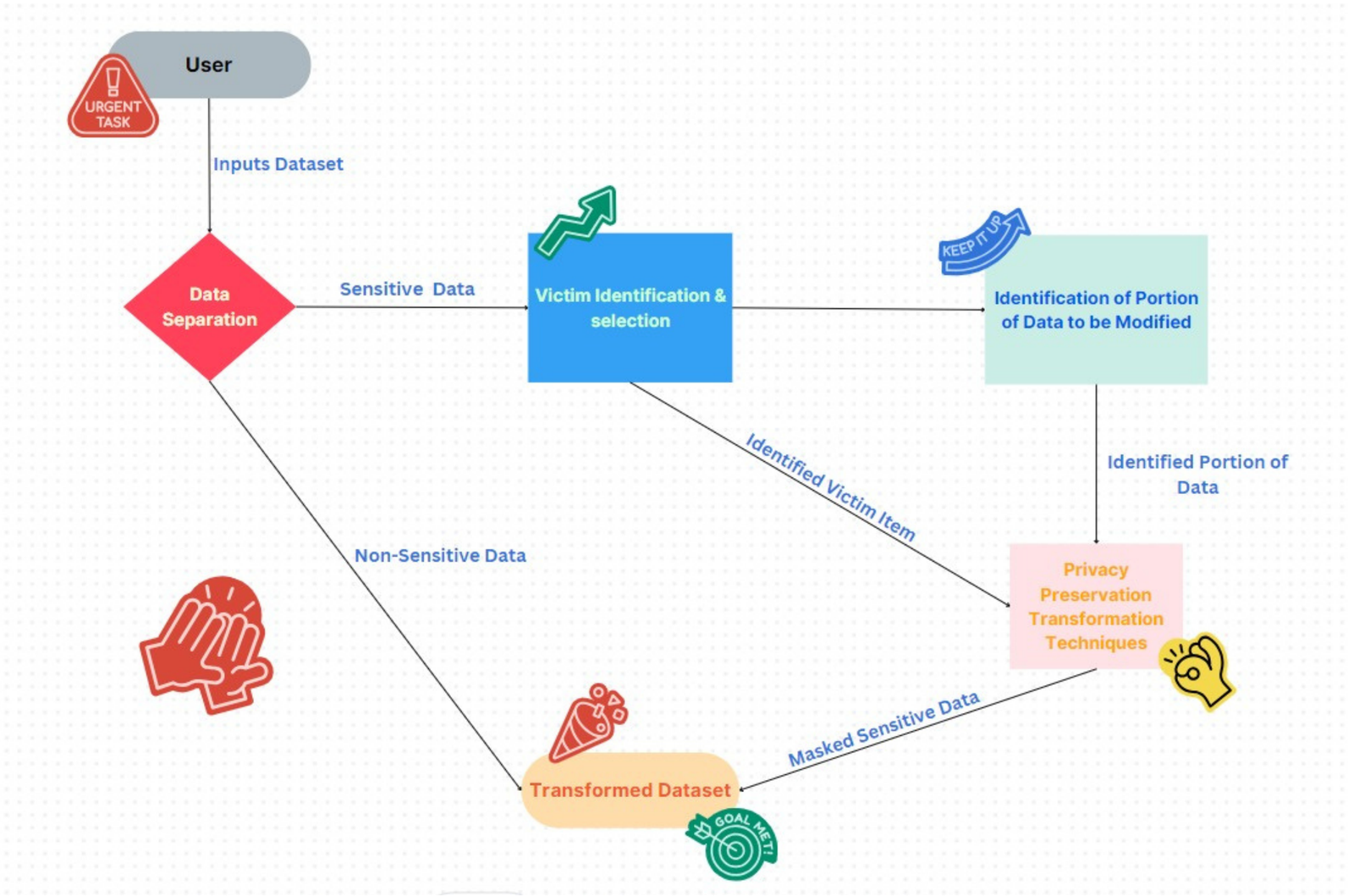
4

Privacy Preservation Transformation Techniques

5

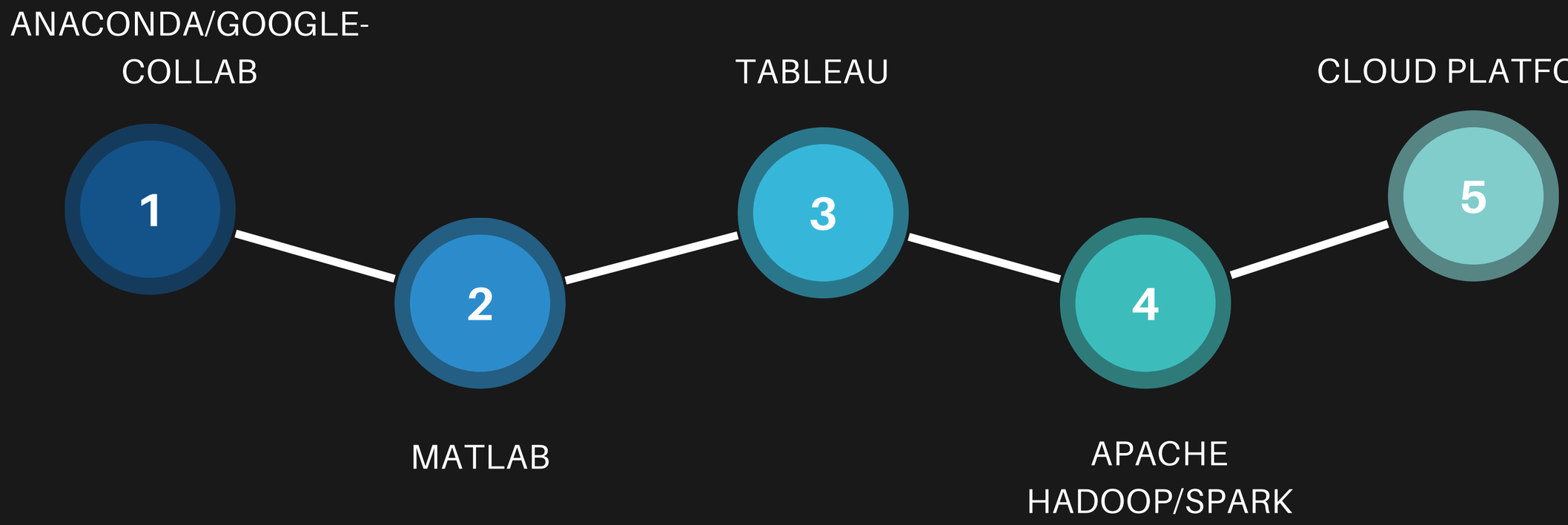
Minimization of Data Loss and Constraints

METHODOLOGY



PROJECT REQUIREMENTS

We’ll be using following tools/platforms in our project



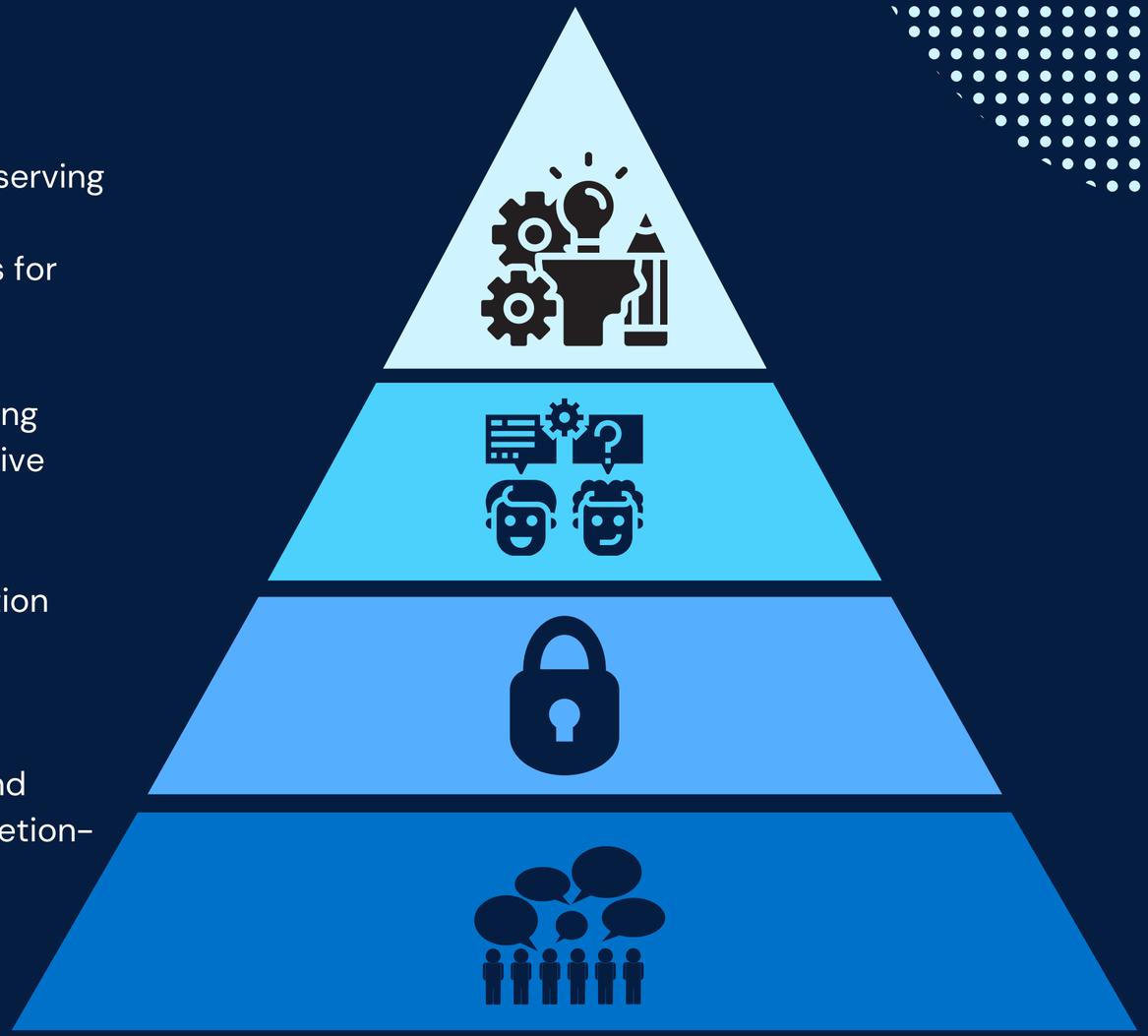
PROJECT OUTCOMES

01 Development of a novel approach to privacy-preserving sensitive information masking, integrating data aggregation techniques with innovative strategies for privacy preservation and utility enhancement.

02 Design and implementation of algorithms for adding artificial data to decrease the frequency of sensitive information while maintaining data utility.

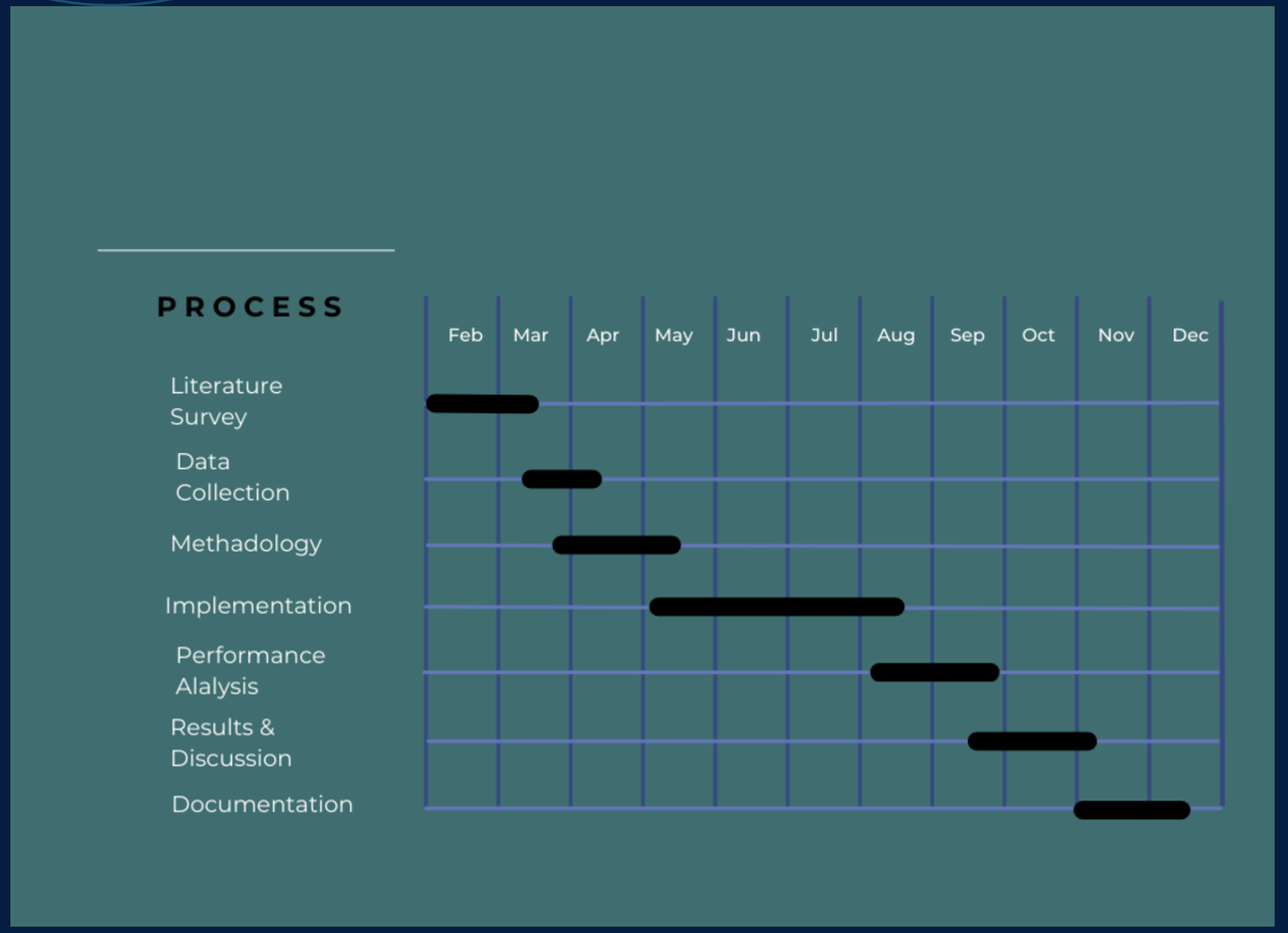
03 Creation of a secure data transformation application capable of real-time data transformation for collaboration with third parties.

04 Demonstration of improved privacy protection and enhanced data utility compared to traditional deletion-based methods.



WORK PLAN

The following plan will be followed to ensure the successful completion of the project:



REFERENCES

[1] Al-Rubaie, Mohammad, and J. Morris Chang. "Privacy-preserving machine learning: Threats and solutions." IEEE Security & Privacy 17.2 (2019): 49-58.

[2] Amiri, Ali. "Dare to share: Protecting sensitive knowledge with data sanitization." Decision Support Systems 43.1 (2007): 181-191.

[3] Dwork, Cynthia. "Differential privacy." International colloquium on automata, languages, and programming. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[4] Fung, Benjamin CM, et al. "Privacy-preserving data publishing: A survey of recent developments." ACM Computing Surveys (Csur) 42.4 (2010): 1-53.

[5] Jangra, Shalini, and Durga Toshniwal. "VIDPSO: Victim item deletion based PSO inspired sensitive pattern hiding algorithm for dense datasets." Information Processing & Management 57.5 (2020): 102255.

[6] Lin, Jerry Chun-Wei, et al. "A sanitization approach for hiding sensitive itemsets based on particle swarm optimization." Engineering Applications of Artificial Intelligence 53 (2016): 1-18

THANK YOU