

## **MediKUest: A Journey Towards Private Health Data**

Ahmet Emin Kaya - 75180

Cem Ozan Doğan - 72623

Oğuzhan Şanlı - 72126

Muzaffer Mert Akkan - 71456

### **SUMMARY**

In the 21st century, privacy has become a lot more important due to technological developments. The number of people who share their data increased exponentially in the last decade. That is why there needs to be differentially private methods to protect the privacy of the people.

The project we will do will consist of a medical DP system which has data about student health records. We will create an interactive user interface to query the database and get the results in a differentially private way.

### **MOTIVATION AND PROBLEM STATEMENT**

The problem we will tackle will be how to create a differentially private(DP) system to protect each student's identity. Sharing of medical data can be very important for medical research and decision-making as we have seen in the COVID-19 pandemic. However, sharing individuals' data directly may reveal important personal information which may fall into the hands of untrustworthy parties. One example of such disclosure of private information was seen in Australia in 2016 when the Australian Ministry of Health released medical billing records. Even though some measures were taken to ensure privacy, researchers were able to re-identify individuals from the given data [6,8]. To prevent such issues, we will design a differentially private user interface for medical data that will allow medical professionals and researchers to access the data they want while protecting individuals' privacy. The challenging aspect of our solution to this problem will be deciding on parameters like sensitivity since in our dataset we have very diverse data and wrong parameters might lead to too much utility loss. Moreover, once we process the queries, we need to find a way to present them in a way that is useful to medical researchers.

### **TECHNICAL APPROACH**

The solution of our problem consists of creating a database which will hold a large collection of ready to use medical records. On those medical records, differential privacy modifications will be performed in order to obtain a system that can be used to safely query modified medical records without any privacy leakage.

#### **Method Selection:**

We are going to use Python for development of Differential Privacy mechanisms which will provide differentially private queries from our medical dataset [1,2,7]. Reasons for that include factors such as all the team members being familiar with Python which will make the

development process easier. Python has many powerful libraries that we are planning to use during the development of our project and also Python has high readability which is a crucial factor when it comes to development and implementation of complex algorithms such as the DP algorithms [3,7].

When it comes to our frontend side of the project, we are going to use React.js as the backbone which will allow us to create an intuitive and dynamic user interface that will be displayed on our web page. Since React.js has a component based architecture it will make our development process easier and enable us to create a responsive web page as a result.

As for our backend side, we are planning to use the Python Flask framework which is a popular choice when it comes to selection of flexible and lightweight web frameworks. Python Flask will allow us to connect our frontend side and the DP mechanisms side of the project to our database since we will have to make many queries in order to succeed in our project. Furthermore, Python Flask will also handle the API requests and is completely compatible with Python which we are going to use for our DP mechanism which makes it an ideal choice for seamless integration and the general objectives of our project.

Here is our summarized action plan for this project:

1. Creating the database side systems.
2. Creating the query side Laplace Mechanism.
3. Creating the frontend.
4. Connecting the frontend and backend.
5. Testing the whole system.

#### **Database Construction and Differential Privacy Implementation:**

To begin, we will construct a secure database to store modified medical records. The database is planned to be designed with privacy as the highest priority which is going to be achieved with the implementation of DP mechanisms which is the main objective of our project.

The DP methods that we are going to implement will be implemented using Python as our choice of programming language. We will integrate the Laplace Mechanism and Exponential Mechanism which are popular and proven approaches when it comes to Differential Privacy[3,7]. The integration of the Laplace Mechanism and Exponential Mechanism will ensure that the privacy is maintained while providing appropriate responses for the queries made.

#### **Web Development and Backend Integration:**

Second phase of our project involves development and design of the frontend side of the project. For this part we are mainly going to use React.js as our framework that will enable us to construct an easy-to-use web application that is fast and will also allow the users to interact with our system efficiently.

Later when our web page is ready we will integrate the backend side into our frontend by using Flask. Python Flask framework will handle the communication with the DP system part of our project and will also enable the processing of queries to be sent into the frontend side where everything will be displayed. This integration is planned to ensure efficient request processing. A diagram of our system can be seen in Appendices-C.

### **System Testing:**

Testing of the project will be conducted at every stage of development to ensure that our Differential Mechanisms are working correctly and allow us to obtain private query results. Tests planned include unit testing and integration testing to ensure reliability and security of our project.

## **DELIVERABLES**

1. The project report
2. The source code for our project
3. Video recording for project demo

## **TIMELINE**

Team Members	Dec 18-24	Dec 25-31	Jan 1-7	Jan 8-15	Jan 16-23
Ahmet Emin Kaya	Proposal, Database, Frontend and Backend design	Frontend implementation	Connecting Backend and Frontend	Report, Bug Fixing, Testing	Report, Bug Fixing, Testing
Cem Ozan Doğan	Proposal, Database, Frontend and Backend design	Backend implementation	Connecting Backend and Frontend	Report, Bug Fixing, Testing	Report, Bug Fixing, Testing
Oğuzhan Şanlı	Proposal, Database, Frontend and Backend design	Backend implementation	Connecting Backend and Frontend	Report, Bug Fixing, Testing	Report, Bug Fixing, Testing
Muzaffer Mert Akkan	Proposal, Database, Frontend and Backend design	Frontend implementation	Connecting Backend and Frontend	Report, Bug Fixing, Testing	Report, Bug Fixing, Testing

## BIBLIOGRAPHY

- [1] "Medical Students Dataset." <https://www.kaggle.com/datasets/slmsshk/medical-students-dataset> (accessed: Dec. 21, 2023).
- [2] Liu, W., Zhang, Y., Yang, H., & Meng, Q. (2023). A survey on differential privacy for Medical Data Analysis. *Annals of Data Science.* <https://doi.org/10.1007/s40745-023-00475-3>
- [3] N. Fernandes, A. McIver, and C. Morgan, "The Laplace Mechanism has optimal utility for differential privacy over continuous queries." 2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), 2021, doi: 10.1109/lics52264.2021.9470718.
- [4] W. Huang, S. Zhou, T. Zhu, Y. Liao, C. Wu and S. Qiu, "Improving Laplace Mechanism of Differential Privacy by Personalized Sampling," 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Guangzhou, China, 2020, pp. 623-630, doi: 10.1109/TrustCom50675.2020.00088.
- [5] H. Jiang, J. Pei, D. Yu, J. Yu, B. Gong and X. Cheng, "Applications of Differential Privacy in Social Network Analysis: A Survey," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 108-127, 1 Jan. 2023, doi: 10.1109/TKDE.2021.3073062.
- [6] A. Dyda, "Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality." *Patterns*, vol. 2, no. 12, p. 100366, 2021, doi: 10.1016/j.patter.2021.100366.
- [7] L. Lu, Y. Li, Y. Zhou, F. Tian and H. Liu, "Adaptive Differential Privacy Interactive Publishing Model Based on Dynamic Feedback," 2018 International Conference on Networking and Network Applications (NaNA), Xi'an, China, 2018, pp. 218-222, doi: 10.1109/NANA.2018.8648706.
- [8] C. Culnane, B. Rubinstein, and V. Teague, 'Health Data in an Open World', 12 2017.doi: <https://doi.org/10.48550/arXiv.1712.05627>

## APPENDICES

### A. Drive link to our datasets

<https://drive.google.com/drive/folders/1sXfjgZSC5vhLFqxMxISoJEMkJ4MCKrK>

### B. The student medical dataset

<https://www.kaggle.com/datasets/slmsshk/medical-students-dataset>

### C. A diagram of how our system will work:

