DFS Project

Data Anonymization

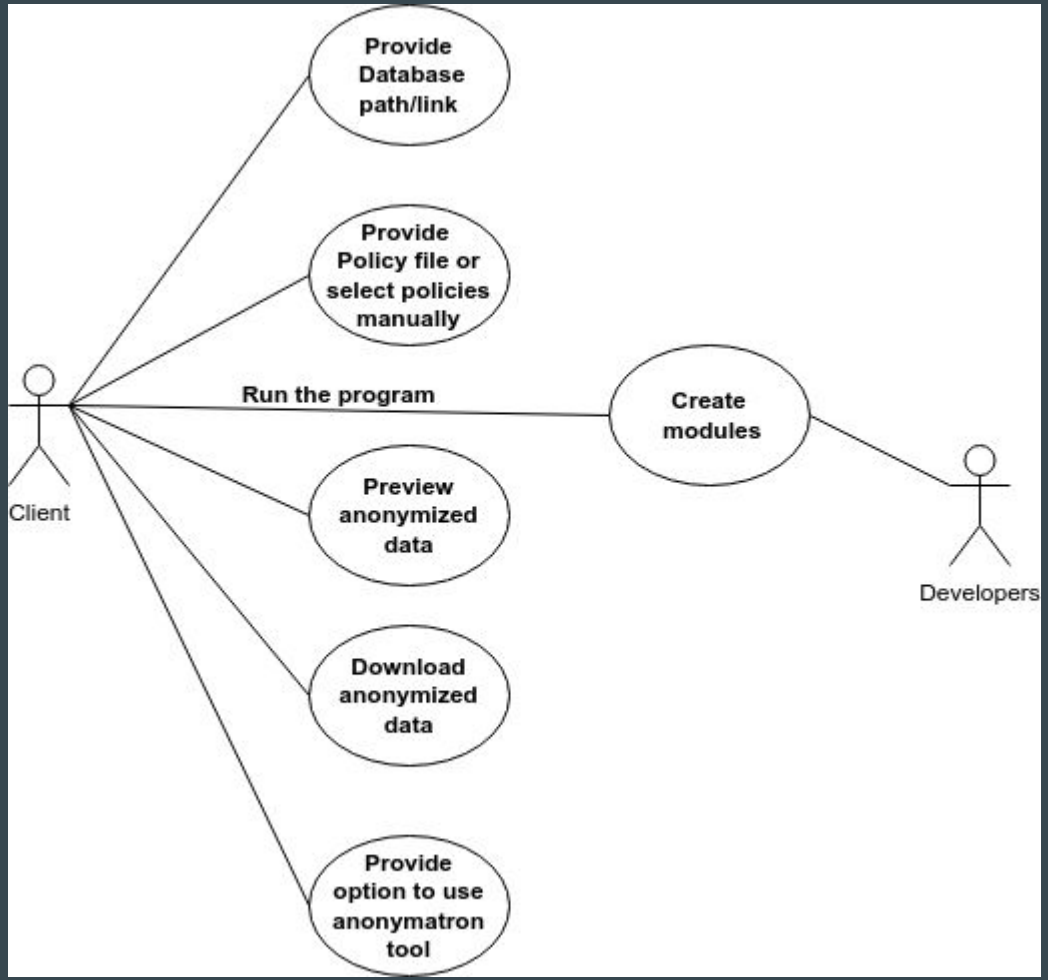
Overview

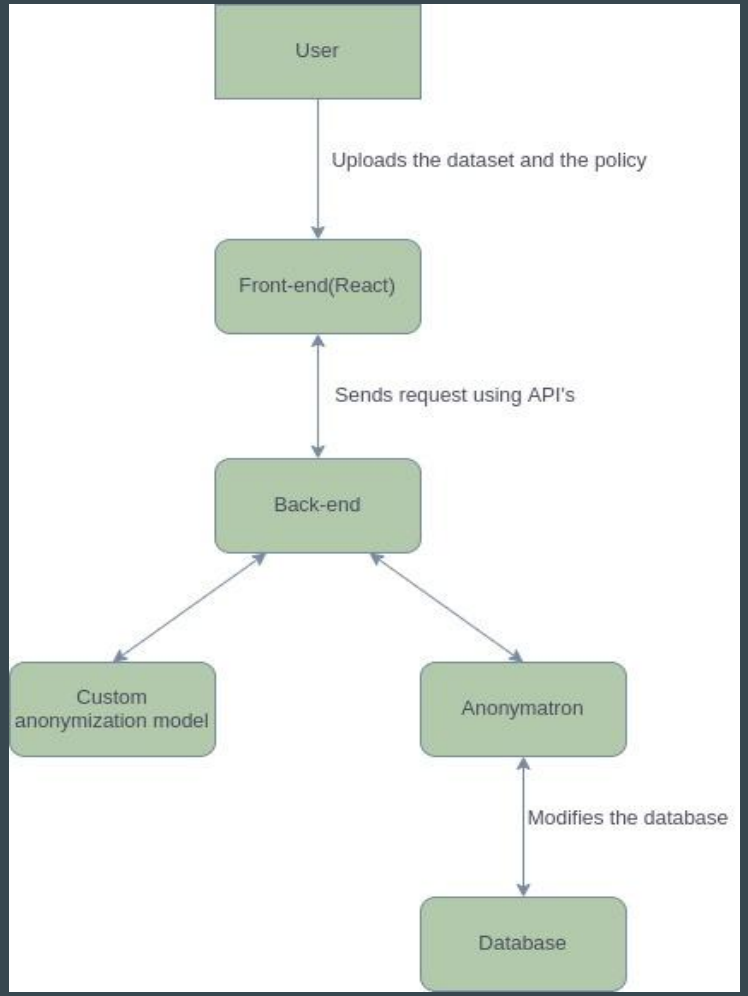
- The main aim of the project was to create a data anonymization tool which can be used in the data-foundation to make sure that the privacy of users is taken into consideration.
- We have implemented anonymization features like data masking, data perturbation, data pseudonymization, generalization, data swapping, using which the user can anonymize any dataset.
- We have also integrated Anonymatron tool which can anonymize the data which is stored in a database as well.

Key Requirements:

- A tool should be created using which the data-foundation's user can anonymize the datasets which he creates or wants to use.
- The user should be able to upload the dataset as well as the policy he wants to apply on each column.
- Different Data anonymization methods like Data perturbation, Data masking, Data perturbation, Data pseudonymization, Generalization, Data swapping, Dropping columns should be supported.
- The user should be able to download the anonymized dataset.

Use Case Diagram





Design diagram

- The design mainly consists of two parts, the first one is the front-end, which is created using react-js.
- The second part is backend which is implemented using Flask.
- The design of the system can be visualized using the following diagram.

Implementation

- Along with our implementation of data anonymization process, we have also integrated an existing tool available for data anonymization.
- In our implementation, the user has been provided with the following data anonymization policy options:
  - Data masking
  - Data perturbation
  - Data pseudonymization
  - Generalization
  - Data swapping
  - Dropping columns

Workflow of our Implementation

- The user is provided with two options to upload the dataset to be anonymized:
  - Uploading through file.
  - Providing link to the dataset.
- Once done, the user needs to provide the anonymization policies he would like to apply on the dataset. The user can do so by two methods:
  - Uploading a JSON file.
  - Selecting the policies manually.
- This would initiate the anonymization and the user would be previewed with the anonymized data.
- The app also supports functionality for the user to download the anonymized data.

Workflow of Integrated tool

- We have also integrated an open source anonymization tool “Anonymatron”.
- The user needs to upload an XML file containing metadata in a particular format required by the tool.
- This would initiate the anonymization and the user would be previewed with the anonymized data.
- The download functionality is also supported for this part.

Visualization

NAME	BRANCH	YEAR	CGPA	AADHAR
Nikhil	COE	2	9.0	2345678998765432
Sanchit	COE	2	9.1	3128758345235432
Aditya	IT	2	9.3	6314828279129927
Sagar	IT	1	9.5	7658675678567856
Prateek	MCE	3	7.8	3456456563456347
Sahil	MCE	2	9.1	1534634677534533

NAME	BRANCH	YEAR	CGPA
Ni****	COE	2	9.0
San****	COE	2	9.1
Ad****	IT	2	9.3
S****	IT	1	9.5
Pra****	MCE	3	7.8
S****	MCE	2	9.1

Individual Contribution

- Abhishek:-
  - Implemented front-end pages: **Upload Dataset page, Preview page, Anonymatron page.**
  - **Integrated** the **Anonymatron Tool** in the backend.
- Divyank:-
  - Implemented front-end pages: **Home page, Upload Policy page and Selecting Policy Manually page.**
  - Handling of **API's** with the backend and **MySQL** connection in python.
- Tirth:-
  - Implemented Backend features: **Anonymization Policies, Downloading Datasets, API's for front-end.**
  - Implemented the **K-anonymization** feature.

THANK YOU!

Made by:-

1) Abhishek Shah (2018101052)
2) Divyank Shah (2018101037)
3) Tirth Upadhyaya (2018101069)