Patient Health Information in the Cloud

EN.601.419 Cloud Computing

Elizabeth Mathew, Varun Venkatesh

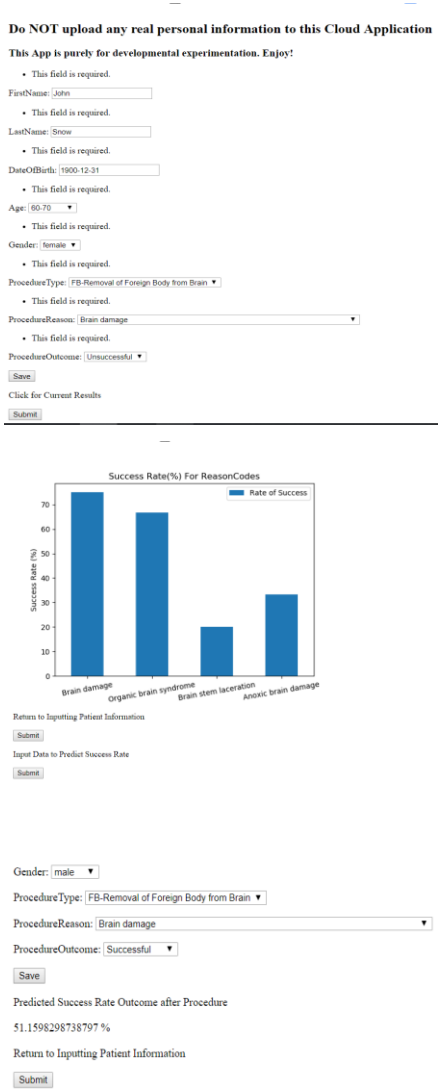
Introduction

- Analysis of medical data has been done for a very long time. In recent years, the use of electronic health data has increased drastically.
- This poses the risk of data breaches and other safety issues regarding patient data. So, data de-identification is being used to protect sensitive information.
- We make an effort to combine data de-identification with analysis of medical data through a cloud-based application.

Design

- Created a Basic Django Framework
- Enable user input to store patient information
- Used Google's HealthCare API[1] to store modular data into a dataset
- De-identify the dataset
- Export data to Google's BigQuery dataset
- Perform data cleaning and

Results



Evaluation

- Time is also a shortcoming of the application. The app runs slowly, taking nearly 10 seconds to display a plot or display a prediction as shown in the table below.

Type of Display	Average Time (sec)	Standard Deviation (sec)
Plot	7.1919782876968	1.5598146872267
Prediction	9.2735009670258	1.8648272094715

Conclusion

- Our app provides a strong first step towards combining de-identification and data processing with cloud based computing.
- There is room to improve how efficiently our app runs by restructuring the templates so that we do not perform repeated operations.
- Although we are sometimes limited by the APIs which we chose, they ultimately provide an efficient way to integrate cloud storage with healthcare data standards.

References
[1] "Projects, Datasets, and Data Stores | Cloud Healthcare API | Google Cloud." *Google*, Google, 24 Apr. 2019, cloud.google.com/healthcare/docs/concepts/projects-datasets-data-stores.

It is important to note that we are not using any real patient data: if anyone wishes to use our patient app to store real patient data, more steps must be taken to ensure the app is not violating any HIPAA policies.