

Sprint #2: Project Planning

Enoch Kang

ekang79@gatech.edu

1 PROJECT TOPIC

Many studies highlight the importance of socioeconomic factors on health. Jemal et al. concluded that almost half of all deaths among working-aged adults in the U.S. were accounted for by potentially avoidable factors associated with lower educational status (Jemal, et al. 2008). Meta-analysis of Galea and colleagues demonstrated that comparable deaths occur due to low education, racial segregation, and low social support as to myocardial infarction, cerebrovascular disease, and lung cancer, respectively (Galea, et al. 2011). More recent studies and reviews only reverify the importance of social factors on health (Braveman & Gottlieb 2014). However, such importance is yet to be reflected by most medical systems. Most information regarding significant social factors are often sporadically recorded into different domains of electronic health records (EHRs) while systematic evaluation of an individual's social determinants of health (SDoH) is rarely possible.

Recent studies suggest that integration of individual-level SDoH into EHRs holds promise for multiple benefits, including assistance in risk assessment and more accurate prediction of health outcomes (Chen, et al. 2020). With a more systematic approach that enables a collective organization and methodical evaluation of an individual's SDoH, one could expect a higher quality healthcare. However, based on currently available information, no software exists that assist individual-level SDoH collection or evaluation as an integrated part of EHRs yet.

This project plans to build a program to collect and evaluate an individual's SDoH. With the use of natural language processing, the program will automatically screen an individual's existing medical records to collect the sporadically existing information regarding SDoH. Then, the program will evaluate or classify the integrated information to be conveyed to the clinician for use in better risk assessment and medical decision-making.

2 TECHNICAL DESIGN

2.1 Tools/Technology

The following tools will be used: Python, Python machine learning libraries, Java, SMART on FHIR, ClarityNLP

2.2 Datasets and Data Sources

Patient datasets that are already mapped to the FHIR standards will be found and used. With the purpose of developing a program, synthetic datasets may be used. The SMART Health IT Launcher and HAPI FHIR demo servers will be used.

2.3 Architecture Diagram

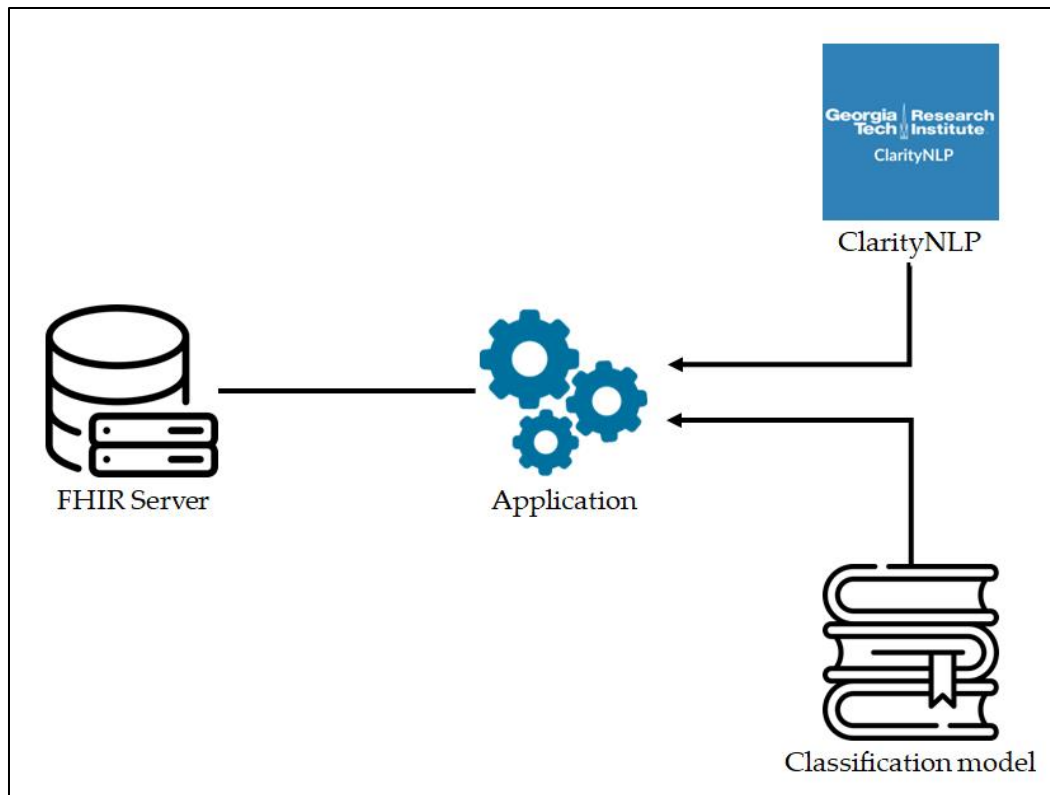


Figure 1— Architecture diagram for this project; a SMART on FHIR application developed with ClarityNLP and machine learning classification model will be launched from a FHIR server.

2.4 Screen Mock-up

The screen mock-up is a rectangular box containing the following text:

Patient Name: Enoch Kang

Social Determinants of Health for this Individual:

- Family income: no information found
- Education level: college graduate
- ...

Overall Evaluation:

Non-favorable impact (35th percentile) on health expected

Figure 2 — Screen mock-up for this project; the screen will show an organized summary of the collected information on the individual's SDoH as well as an overall evaluation.

3 IMPLEMENTATION PLAN

3.1 Project Tasks

- Preparation
 - Literature review to select relevant socioeconomic factors
 - Find appropriate datasets
- Incorporate ClarityNLP into the application
 - Learn the general workings of ClarityNLP
 - Apply ClarityNLP to organize SDoH information
- Incorporate classification model into the application
 - Select the appropriate classification algorithm for this project
 - Prepare the training and testing datasets for the learning model
 - Train and test the learning model
 - Incorporate the trained model into the final application
- Develop the functional application

- Develop an application that has access to necessary resources
- Develop a result screen
- Ensure functionality with incorporation of necessary components

3.2 Project Timeline

- Week 8
 - Literature review to select relevant socioeconomic factors
 - Find appropriate datasets
- Week 9
 - Literature review to select relevant socioeconomic factors
 - Find appropriate datasets
- Week 10
 - Develop an application that has access to necessary resources
 - Develop a result screen
- Week 11
 - Develop an application that has access to necessary resources
 - Develop a result screen
- Week 12
 - Learn the general workings of ClarityNLP
 - Select the appropriate classification algorithm for this project
 - Prepare the training and testing datasets for the learning model
- Week 13
 - Apply ClarityNLP to organize SDoH information
 - Train and test the learning model
- Week 14
 - Incorporate the trained model into the final application
- Week 15
 - Ensure functionality with incorporation of necessary components

3.3 Needs/Risks

- Access may have to be received for use of an appropriate dataset.
- ClarityNLP may not be perfectly customizable to search for the selected socioeconomic factors.
- The data may not be adequate to prepare clean training and testing datasets.
- The trained model may not perform as strongly or reliably as expected.

4 REFERENCES

1. Jemal, A., Thun, M. J., Ward, E. E., Henley, S. J., Cokkinides, V. E., & Murray, T. E. (2008). Mortality from leading causes by education and race in the United States, 2001. *American journal of preventive medicine*, 34(1), 1-8.
2. Joyner, D. A. (2017). Scaling Expert Feedback: Two Case Studies. In *Proceedings of the Fourth Annual ACM Conference on Learning at Scale*. Cambridge, Massachusetts.
3. Galea, S., Tracy, M., Hoggatt, K. J., DiMaggio, C., & Karpati, A. (2011). Estimated deaths attributable to social factors in the United States. *American journal of public health*, 101(8), 1456-1465.
4. Braveman, P., & Gottlieb, L. (2014). The social determinants of health: it's time to consider the causes of the causes. *Public health reports*, 129(1_suppl2), 19-31.
4. Chen, M., Tan, X., & Padman, R. (2020). Social determinants of health in electronic health records and their impact on analysis and risk prediction: a systematic review. *Journal of the American Medical Informatics Association*, 27(11), 1764-1773.