Project Title: Identify fraudulent patterns in healthcare data via predictive modeling

Sponsor Contact: Swathi Young <syoung@integritym.com>
Natasha Williams <nwilliams@integritym.com>

Background: Although many healthcare providers work ethically and provide high-quality patient care, there are some who do not. These types of healthcare providers often exploit federal healthcare programs, such as Medicare stealing billions of dollars each year. As a result, this negatively impacts Medicare beneficiaries' health and welfare costing taxpayers billions of dollars. There is a recurring need to effectively identify patterns in fraud, waste and abuse in these programs. To combat this issue, IntegrityM utilizes healthcare data analysis techniques to detect and prevent fraud, waste and abuse across Medicare, Managed Care and Medicaid programs.

Project Overview and Problem Definition: Emerging technologies like Artificial intelligence (AI) and Machine Learning (ML) have the potential to help the Department of Health and Human Services, Centers for Medicare and Medicaid Services (CMS) mitigate fraud, waste and abuse, expedite the claims process, reduce errors, and lower the costs of Medicare claims management. AI and ML have the potential to help government agencies like CMS to mitigate fraud, reduce errors, and lower the cost of paper-intensive processes, while enabling collaboration across multiple divisions and agencies to provide more effective and efficient services to citizens.

Project Goals: The objective of the project is to develop a predictive model prototype to identify fraudulent patterns in healthcare data. To accomplish this, a team of data scientists will develop and use machine learning algorithms to detect fraud, waste and abuse using medical claims data*. By identifying these patterns, preventative measures can be established and implemented to reduce fraud, waste and abuse which can lead to reduced patient harm and financial costs.

Potential Roadblocks and Barriers to Success: Usage of publicly available datasets for this problem can potentially lead us to less accurate results.

Preferred Methodology: No preference on the project execution. This is up to the student team.

Data Requirements and Availability: Data will be provided.

Analytics Requirements: Machine learning using R/Python interactive visualizations through Tableau may be helpful.

Preferred Tooling: The team will be providing the project sponsor with all data, visualizations, and program code. We prefer Python for the bulk of the code and potentially Tableau for visualizations.

Project Schedule: Below the template class schedule with regard to project completion and sponsor interaction.

Session	Date	Client Activity
Session 1	September 2021 1st week	1. Reach out to client. 2. Study the "Project Brief" 3. Plan a structured client ideation meeting 4. Meet with client. 5. Develop a workplan (based on CRISP-DM), schedule client meetings and comm protocols, "internal" meetings, roles and responsibilities, milestones, and activities
Session 2	September 2021 2nd week	Client meeting 2 1. Weekly client cadence 2. Present client workplan to client 3. Present client deliverable outline and strawman deliverable to client for approval 4. Discussion on "Problem Understanding" and "Data Acquisition" Client meeting 3 1. Status report 2. Focus on "Data Understanding" and "Data Preparation" (in the context of the strawman) 3. Discuss modeling methodology and finalize tooling
Session 3	September 2021 3rd week	Client meeting 4 1. Status report 2. Completion of Data Acquisition and Preparation 3. Share data understanding and visualization Client meeting 5 1. Status report 2. Preliminary analytics results
Session 4	September 2021 4th week	Client meeting 6 1. Status report 2. Visible signs of progress (progressed strawman) 3. Progress through Iterate of Data Preparation and Analytics Modeling Client meeting 7 1. Status report 2. Visible signs of progress (progressed strawman) 3. Progress through Iterate of Data Preparation and Analytics Modeling

Session 5	October 1st week 2021	Client meeting 8 1. Status report 2. Visible signs of progress (progressed strawman) 3. Progress through Iterate of Data Preparation and Analytics Modeling
Session 6	Oct. 2021 2nd week	Client meeting 9 1. Status report 2. Visible signs of progress (progressed strawman) 3. Progress of machine learning models 4. Evaluate accuracy, precision and recall
Session 7	Oct. 2021 3rd week	Client meeting 10 1. Status report 2. Visible signs of progress (progressed strawman) 3. Progress of machine learning models 4. Evaluate accuracy, precision and recall
Session 8	Oct. 2021 4th week	Client meeting 11 1. Status report 2. Discuss presentation of final results
Session 9	Nov 2021 1st week	Client meeting 12 1. Status report 2. Present final results to client
Session 10	Nov 2021 2nd week	Practice for Presentations -- Teams 1 to 4
Session 11	Nov 2021 3rd week	Practice for Presentations -- Teams 5 to n
Session 12		Thanksgiving holiday
Session 12	Dec 1st/2nd week	Present final results to clients.