# 

# Student Project Proposal

|  |  |
| --- | --- |
| Project Title | Healthcare Provider - Fraud Anomaly Detection |
| Industry Sponsorship (if Any) | None, n/a |

## Project Description

**Problem definition**

*[50-100 word description of the problem which you will solve]*

|  |
| --- |
| Healthcare fraud is a lucrative white-collar crime. The US National Health Care Anti-Fraud Association estimates that 3 percent of healthcare spending is lost to fraud (conservatively $300 billion/yr).  A digital fraud detection system is ideally suited to combat this threat to our fundamental right to accessible and affordable healthcare. |

**Key Research Questions/ Technological constraints that the Project will Answer**

|  |
| --- |
| 1. Can ML detect financial anomalies in Provider data? 2. Can anomalies be categorized as fraud? 3. What types of fraud is the model best suited for? |

**Final deliverables at the end of the project**

*[Please list the desired technical deliverables from the project team in as much detail as possible]*

|  |
| --- |
| 1. Tested supervised/unsupervised model, along with jupyter notebook for reproducibility 2. Admin tools/utilities for model monitoring and re-training (pipelines) 3. User tools/utilities for user presentation / review of findings (PowerBI - monthly / quarterly reports) |

**Key activities/ technologies the project team may be expected to undertake/ work with**

*[E.g. What kind of technology stack will you work with, the datasets you may need to work on, what kind of analysis you may be expected to undertake, etc.]*

|  |
| --- |
| Key Activities:   * Scoping: Select/confirm target data set (Providers? Timeline? ) * Data Eng:   + Define Baseline Data set   + Label and Organize Data * Modelling:   + Select and Train the Model   + Testing and Error Analysis * Stage / Deploy:   + Stage to QA, SIT   + Stage to Prod, Demo   + Monitor, Optimize, Maintain   Key Technologies:   * Infrastructure (OS) * Data Eng * Modelling * CI/CD Pipeline |

**Expected learning outcomes**

*[What do you expect to learn from the project? Please mention the technical skills you will imbibe over the project.]*

|  |
| --- |
| 1. Data Engineering 2. Feature Selection / Engineering, Unsupervised Learning Methods 3. Solution Architecture, API development 4. CI/CD Pipeline development |

|  |  |
| --- | --- |
| Team Size: | 3 |
| Member names: | Monika Sharma  David Lederer  Iain McKone |

### 

## Tentative Time plan

Submit a tentative time plan (table/chart or text) regarding breakdown of the work that will be conducted between in the second half of your cohort, from week 6 onward.

|  |  |  |
| --- | --- | --- |
| Week# | Key Tasks | Key Deliverables |
| 5 | Data Scoping |  |
| 6 | Solution Architecture, System Design |  |
| 7 | Feature Selection, Eng |  |
| 8 | Feature Eng |  |
| 9 | Data Labelling |  |
| 10 | Model Training |  |
| 11 | Model Validation |  |
| 12 | Model Testing |  |
| 13 | MLE Iteration, Optimization |  |
| 14 | Demo Workflow; UI Development |  |
| 15 | Dry Run; Demo Finalization |  |
| 16 | Preso / Demo |  |

## System Design

From the System design perspective, outline the following:

* Data
* Process (Models, iterations)
* Outcome (output and recommendations)

What are the system design considerations for your deployable ML model? Describe the iterations, delivery formats and limitations you may face and some solutions to overcome the limitations

* Should the model be deployed to run in batch, or to be hit from an api or some sort of streaming process as events are generated?
* What sort of infrastructure will be required for training? If it is a model that requires a lot of resources, where is the best place to train?

## Ethical Considerations

Are there any ethical considerations of your project? Consider the data source, the intended outcome, and/or the eventual use cases.

* Did you modify anything about your plan based on these considerations?
* Can you anticipate any issues that might arise during the process?

Accuracy:

Bias:

Fairness:

Privacy:

Transparency: Traceability, Communication, Intelligibility

Accountability: there are legal and financial ramifications to

Human Control and Decision Making:

Sources of potential harm:

* Innocent until proven guilty
* %confidence versus absolutes
* Reputation, unfair allocation, unfair representation