Claim categorization using Artificial Intelligence: a proof of concept

Two-Sentence Overview

The goal of this project is to develop an automatic claim categorization module for openIMIS based on state-of-the-art Artificial Intelligence (AI) algorithms and methodologies to drastically reduce the manpower, resources and time required to review the claim for reimbursement.
As a part of its strategy to contribute to an equitable society, Swiss Tropical and Public Health Institute (Swiss TPH) designed and developed openIMIS and works with the reputable IT firm SolDevelo who have extended experience in health informatics. Both institutions are already working in close partnership on the development and maintenance of openIMIS and, with the help of an Implementer partner, can provide access to categorized claims data and validate the proposed solution.

Executive Summary
This project aims to develop an AI-based claim categorisation proof-of-concept to automatically update the claim's status to accepted, rejected or to be further analysed by a medical expert. It will be divided into three phases:

- Research phase: Swiss TPH will work closely with the Implementer to undertake the research and development of the AI algorithm for claim categorisation based on anonymized openIMIS claim data and associated entities (insuree, health facility, diagnostics, medical items and services); it is mandatory to have access to a database of already categorized digitized claims (by a medical officer) and to define the most influential features to be considered by the AI algorithm; the objective of the AI algorithm will be to have a performance similar to the human expert; several (supervised and unsupervised) methods can be implemented to meet this goal;
- Development phase: SolDevelo will develop an openIMIS module that will integrate the AI algorithm and will make the necessary links and transitions for its activation;
- Testing phase: the Implementer will run the AI module in real case situations to validate the previous developments.

The Digital Square investment will allow us to undertake the activities defined by these phases.

Consortium Team

Swiss TPH (prime organisation)

Swiss TPH is a leading institute in global health with a particular focus on low- and middle-income countries with a staff strength of over 850 staff from 80 different nations, currently active in 300 projects across 100 countries. Swiss TPH will be responsible for technical project management, the expertise on health financing, openIMIS, the development of the AI algorithm, drafting of business and technical specifications, supporting the system and architecture design.
Swiss TPH's relevant experience includes involvement in the design and implementation of the Insurance Management Information System (which is the genesis of the openIMIS Initiative) since its inception in Tanzania and has supported its implementation in a number of countries. Swiss TPH is currently implementing two projects at scale for the deployment of insurance schemes through openIMIS, in Tanzania and Cameroon, and two openIMIS pilots in Chad and Democratic Republic of Congo. In addition, Swiss TPH is actively involved in the development of openIMIS as part of the Implementers and

Developers Committees of the openIMIS Initiative. Further, Swiss TPH is also implementing projects on AI with supervised machine learning for Clinical Decision Support System improvements.

SolDevelo

SolDevelo is a dynamic IT company (+80 staff) focused on delivering high-quality software and innovative solutions. SolDevelo will be responsible for software integration of the AI module into openIMIS. SolDevelo is currently involved in several openIMIS projects, including the 'maintenance and support project', HL7 FHIR module development, openIMIS integration with OpenMRS and enhancing the security of the legacy system. SolDevelo has been involved in many opportunities that required skill sets relevant to this particular project, especially through opportunities like OpenMRS (core contributors), HL7 FHIR (OpenMRS Sync 2.0 module), nationwide micro-service based implementations (OpenLMIS), nationwide OpenHIE architecture based implementations (National Health Infrastructure project with such components like OpenELIS, DHIS2, OpenMRS and many other HIE compatible applications, health standards-based workflows for the Client Registry, Facility Registry, Health Management Information System, Shared Health Record, and Interoperability Layer).

openIMIS Implementer

The partnering with an openIMIS Implementer (governmental entity currently productively using openIMIS, e.g. Ministry of Health Tanzania, Ministry of Health Nepal) is required. The Implementer will be responsible for the already categorized claim data, the validation of the anonymized data set for AI analytics and the testing of the openIMIS AI module developed under this project.
The Implementer should have implemented openIMIS for several years and should have thousands of categorized claims already entered into their database. They should be familiar with openIMIS database structure and share with the partners' anonymized data that will be used for the development of the AI algorithm.

Project Description

Background and problem statement

Substantial manpower, time and resources needs to be deployed in the process of claims management for cost refunds or claim audits. Considering the respective insurer policies and regulations, this process can result in a) payment of the claim, b) rejection of the demand or c) for further analyzis by a Medical Officer.
Currently, static validation rules based on insurance product configuration are implemented in openIMIS. They are hardcoded in the system and applied on claims submission. However, these validation rules only allow claims to be accepted or rejected and require the subsequent review by a professional Medical Officer. Once the database of categorized claims is available, AI allows the encoding of the Medical Officer's knowledge into a model that can automatically check future claims, allowing the Medical Officer to concentrate on those claims that really need to be reviewed such as inconsistent, erroneous and fraud/abuse claims.

Objectives

As described in the previous sections, the activities will be grouped into three phases:

1. Research phase

The research phase will require an extensive, well-validated database of categorised claims to train, develop and test the AI algorithms. The first activity will be the definition of the required claim data (entity models and attributes) to feed the AI algorithms. Based on the required data and on the data privacy

concerns, the second activity will be the development of a script (code) that will anonymize the implementer data being used during the AI algorithm development.
For the AI algorithm development, several methods will be considered [1,2,3,4]:

- Supervised methods will attempt to discover relationships between claim entities (e.g. price of a drug or service, the age and gender of the patient, the diagnosis) and the output variable. Methods such as neural networks, decision trees, Support Vector Machine are applied in the insurance sector.
- Unsupervised learning methods will allow to assess one claim attributes in relation to other claims to determine the similarities between them and thus will be able to distinguish anomaly records or group of similar records using methods like ‘anomaly detection’, ‘association rules’, ‘clustering’.

The development of the AI script will cover several steps such as:

- data preparation (analysis of the existing database, visualisation, normalisation, error correction, splitting the data in ‘train/dev/test sets’)
- choosing an algorithm, training the model and evaluating its results (taking into account the objective and the existing data)
- prediction based on new un-categorized data

We will search for the most appropriate AI method and based on the anonymized data, we will develop the AI algorithm that will provide a model, the parameters will be stored into openIMIS database, allowing us to categorize future new claims.

2. openIMIS AI module development
Based on the new modular openIMIS architecture, we will develop a new openIMIS module that will integrate the proposed AI algorithm whilst taking into account the Implementers’ specific needs. During this integration, we need to decide how the algorithm will work together with the current claim review process and what event will activate it (e.g. after submission of a claim, after the static validation ...).

3. Testing in real case situations
The openIMIS AI module will be tested and validated on new submitted claims in real case situations by the Implementer.

In all these phases, Swiss TPH will closely monitor the advancement of the project and report periodically to Digital Square and to the openIMIS Initiative.

Deliverables and Schedule

The following deliverables expected from this project are:

- AI input data model structure (Jan 2020)
- SQL script to anonymize production data (Jan 2020)
- AI output data model structure (Feb 2020)
- AI script in Python (May 2020)
- openIMIS AI module in Python (Jul 2020)
- Test plans, data and results (Aug 2020)

Risk and Mitigation

The AI model thus generated will depend on the quantity and quality of categorised digitized claims. Wrong claim categorisation or not enough claims could lead to a limited model that could risk generating errors (claim rejected without reasons). We will need to define responsibilities as well as prerequisites (e.g. the minimal amount of already categorised data to be used) in the usage of the module.

References

[1] Li, J., Huang, KY., Jin, J. and Shi, J. A survey on statistical methods for health care fraud detection, Health Care Manage Science, 11 (3), pp. 275-287, 2008
[2] Ortega, P.A., Figueroa, C.J., and Ruz, G.A. A medical claim fraud/abuse detection system based on data mining: A case study in Chile, DMIN'06,The 2006 International Conference on Data Mining, Las Vegas, Nevada, USA, June 26-29, pp. 224-231, 2006
[3] Obenshain, Mary K. Application of data mining techniques to healthcare data. Infection control and hospital epidemiology, 25(8), pp. 690-695, 2004
[4] Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., and Arab, M. Using data mining to detect health care fraud and abuse: a review of literature. Global journaI of health science, 7(1), pp. 194–202, 2014