Journal of Health Informatics in Africa

12th Health Informatics in Africa Conference (HELINA 2019)
Peer-reviewed and selected under the responsibility of the Scientific Programme Committee

A Fraud Detection System for Health Insurance in Nigeria

Terungwa Simon Yange[a*], Oluoha Onyekware[b] and Hettie Abimbola Soriyan[c]

[a]Department of Mathematics/Statistics/Computer Science, University of Agriculture, Makurdi, Nigeria
[b]Department of Computer Science, University of Nigeria, Nsukka, Nigeria
[c]Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

Background and Purpose: This research developed a Fraud Detection System for National Health Insurance Scheme (NHIS) in Nigeria. This was with a view to addressing the fraudulent activities of some stakeholders in NHIS; as many researches have proven that the lack of appropriate tools to do this has negatively affected service providers as well as the beneficiaries of this Scheme.
Methods: In order to achieve the aim of this research, an inspection of organizational documents, direct observation and collection of existing data from NHIS accredited health facilities and Health Maintenance Organizations in Nigeria were carried out. The system was designed using Unified Modelling Language (UML) tools. The implementation of the system was done using MongoDB as the big data storage mechanism for the input, Comma Separated Values (CSV) files as a storage facility for the intermediate results generated during processing and MySQL as the storage mechanism for the final output, Apache MapReduce as the big data processing platform, Association Rule Mining as the data analytics algorithm, and Java programming language as the implementation technology.
Results: The system modules of comprised of four modules: user management, enrollment processing, referral processing and claims processing. With this, it identified different types of frauds in NHIS such as double billing, billing for services not provided, ghost patients, identity theft, self-referral, collusion with providers and kickback schemes.
Conclusions: This paper developed a system for the detection of the fraudulent activities of the actors of NHIS. This system employed data from the Nigerian NHIS which was categorised into: enrollment, referral and claim data with different file formats: pdf, jpg, png, csv and excel.

Keywords: Fraud, Detection, Health, Insurance, Prevention, NHIS.

1 Introduction

Health insurance covers the entire or fragment of the risk of a patient incurring medical expenses, spreading the risk over a large number of persons [1]. By estimating the total risk of healthcare and other expenses in the healthcare system over the pool of risk, an insurer can develop a routine finance structure, such as a monthly premium or payroll tax, to provide the money to pay for the healthcare benefits specified in the insurance treaty. The benefit is administered by a central body such as a government agency, private business or charity organisations [2]. Healthcare insurance has attracted substantial interest in the past. It is a social security plan that promises the delivery of the needed healthcare services to a person on the contribution of a token to provide economic protection to the participants. It is also, a mechanism for protecting families against the unforeseen exorbitant costs of illness by sharing the risks of future costs among healthy and sick populations in the form of regular predictable payments [3].

The National Health Insurance Scheme (NHIS) is a scheme put in place by the Federal Government of Nigeria to provide full access to quality healthcare service in the Country [4]. The scheme covers civil

*Corresponding author: Address: University of Agriculture, Makurdi, Nigeria. Email: lordesty2k7@ymail.com Tel: +2348064067803, +2347056808523
 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

servants, the armed forces, the police, the organized private sector, students in tertiary institutions, self-employed, vulnerable persons, and the unemployed among others [5]. Social security is a human right, as well as an economic and political necessity; it is an indispensable part of an efficient market economy. Social security programmes are usually established as a means of improving the well-being of the poor, reduce inequality within society and conciliate different social demands, thus avoiding the social and political conflicts [6]. Adequate social security policies could be an important factor in the process of socio-political development and economic growth of our dear Country, Nigeria.
Conceptually, NHIS is a welcomed innovation and development in the Nigeria health sector given its objective. Although at present, only public servants in Federal establishments benefit from this scheme since the Federal Government is the only public-sector that has implemented the programme. On reduction of dependence on government for funding health services, [5] acknowledged that the scheme has reduced the burden on the government and improved the funding of health service through its contributory strategy. The 15 percent deduction from basic salaries of workers, which is remitted to the NHIS and the co-payment system, has increased healthcare funding.

Currently, the implementation of NHIS has not been easy due to inadequate physical health facilities and personnel, administrative and logistics bottlenecks [7]. The nation does not have enough healthcare providing institutions with adequate medical facilities and personnel for effective implementation of the Scheme. Besides, the administration of the Scheme has not been easy given the delays in processing document of registered beneficiaries and remitted contributions to the NHIS, and Health Maintenance Organisations (HMOs) and Heath Provides (HPs). Furthermore, the informal sector is very difficult to organized for the Scheme. Even private hospital and clinic are becoming unwilling to embrace the scheme. This has made NHIS prone to different types of fraudulent activities.

According to [8], healthcare fraud is defined as "an intentional deception or misrepresentation made by a person or an entity, with the knowledge that the deception could result in some kinds of unauthorized benefits to that person or entity". The NHCAA projected conservatively that at least 3%, or more than 60 billion dollars, of the United State of America's yearly healthcare expenses was lost to fraud in 2010. This loss which was also estimated by other government and law enforcement agencies was as high as 10%. Besides the monetary loss, fraud has severely hampered the healthcare system from providing quality care to authentic beneficiaries. Hence, the need for an effective fraud detection is imperative for providing easy access to healthcare services, improving the quality and reducing the cost of healthcare services. In any discussion concerning fraud, it is important to state clearly the distinction between fraud prevention and fraud detection. Fraud prevention define procedures to stop fraud from taking place. Fraud detection on the other hand, encompasses recognizing fraud as quickly as possible once it has been committed. Several fraud detection cases involve huge datasets that are continuously changing. In the nutshell, fraud detection surfaces once there is failure in fraud prevention [9].

Instead of prioritising quality healthcare delivery, fraudulent activities are prevalent among professionals in NHIS. These fraudulent activities include unrealistic bills from hospitals, cooking up cases in order to extort the Scheme by the hospitals, collusion of patient with providers or providers with HMOs, **ghost patients,** managed care fraud, reverse false claim cases, lying about eligibility, scheduling extra visits for patients, billing for services rendered by unqualified personnel, ganging, using the wrong diagnosis, false negotiation cases, providing unnecessary care and maximizing care, billing for services not provided, submitting double bills *etc.* [1][6][10]. The absence of a robust and functional health information system and lack of adequate modern information technology infrastructure has hindered the detection of fraud, sharing of information, prompt and timely data processing, creation of a database between various stakeholders in the scheme [6][7].

This paper developed a system for the detection of fraudulent activities in the health insurance scheme in Nigeria, NHIS. The system employed Association Rule Mining algorithm, MapReduce Framework, MongoDB, MySQL and Java Programming Language.

1.1 Health Insurance Fraud

The substantial increase in medical expenditures required to satisfy the quest for high quality and high-technology services has given birth to a greater demand for health insurance schemes. Most people now bank on health insurance systems, which are either funded by the government or managed by the private sector, to share the high cost of healthcare. With this intensive need for health insurance, fraudulent

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

behaviours become a serious problem. For instance, [11] reported that 10% of United States annual spending on healthcare is loss to fraud. The health insurance programmes of other countries are also faced with similar challenges [12]. The other healthcare crimes (*i.e.,* medical and drug) which involve surgeries, invasive testing, certain drug therapies *etc.,* even place their patients at significant physical risk and affect patients' health.

Fraud in the healthcare insurance involve three parties [12][13]: the healthcare service provider (*i.e.,* the physician, pharmacist, laboratory scientist, health centre, pharmacy, laboratory, and even ambulance companies) which render healthcare services; the healthcare service consumer or beneficiary or insurance subscriber (*i.e.*, patient) which receive healthcare service from the provider; and the healthcare insurance carrier which collect steady premiums from subscribers and make the commitment to pay healthcare cost on their behalves. These parties exchange information amongst them in the course of care delivery. This is basically in the form of service requested by the subscriber (patient visit) to the provider, explanation of benefits which contain the detail services rendered by the provider to the subscriber, claim/bill which is sent to carrier for the services rendered to the subscriber by the provider, and the payment to the provider based on the claim submitted to the carrier [11]. As the number of beneficiaries (patients) of this scheme increases, high volume of data is generated by both the providers and the carriers; and consequently, some fraudulent activities (such as billing services that were never rendered, performing medically unnecessary services, misrepresenting non-covered treatments as medically necessary covered treatments, and misrepresenting applications for obtaining lower premium rate) are carried by these actors (beneficiary, provider and insurer) which give rise to the need to investigate such acts in an attempt to identify perpetrators, and this requires a proper analysis tool for the purpose [13][14].

The reasons why health insurance fraud has become a prevalent practice is that majority of those involved find it beneficial in diverse ways. Several surgeons see it as necessary to provide quality care for their patients [15]. Most patients, although disapproving of the idea of fraud, are occasionally more eager to admit it when it affects their own medical care. Programme Administrators are often compassionate on the issue relating to healthcare insurance fraud as they seem to take full advantage of the services of their providers [14]. A summary of healthcare frauds is highlighted below.

Kickback schemes: One of the widespread and discussed form of healthcare fraud is kickbacks [11]. These occur in many forms. For instance, pharmacists can fill a prescription with a specific brand of drugs instead of the other that yields a bonus from the pharmaceutical company. Aside, financial implications, these drugs might also be harmful to the patient's health. Physicians themselves can fraudulently write prescriptions for money, essentially a kickback from the downstream illegal sale of these drugs.

Self-referral: This refers to the transfer of a patient to a facility with which the referring healthcare personnel has a financial relationship [12]. This might involve a kickback scheme if the facility where the patient is referred to, pays a certain amount of money back to the physician, but other financial relationships are conceivable [16].

Identity fraud: This kind of fraud occurs when an uninsured individual assumes the identity of an insured person so as to benefit from the services packaged in the scheme or to hide a particular illness. The healthcare services eventually provided to the person 'lending' their identity could be negatively affected, since this will be at par with their actual health records. This kind of fraud can also be perpetrated without the consent of the actual owner [14] [16].

Double bills: Many care providers submit the same claim multiple times, in order to get paid different times for performing the same action (*i.e.,* submitting a particular claim multiple time for the same service) [10] [17]. This is known as double billing and it is also a fraud.

Billing for services not provided: Here, claims are generated and submitted to insurance companies for healthcare services that are not provided or for drugs that have not been delivered to the patient. This is known as phantom billing [9] [13].

Ghost Patients: The submission of a claim for healthcare services provided to a patient who either does not exist or who never received the service or item billed in the claim or patients that are dead or have changed their provider [13].

Collusion with providers: Both provider and the member collude to submit false claims where the physician receives the benefits from the false claims [18].

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

Among these frauds, the ones committed often by health service providers accounts for the greatest proportion of the total healthcare fraud. The reason for this is that the historically prevailing attitude in the medical profession is one of "fidelity to patients". Although the vast majority of service providers are honest and ethical, the few dishonest ones may have various possible ways to commit fraud on a very broad scale, thus posing great damage to the healthcare system. Some service providers' fraud, such as that involving medical transportation, surgeries, invasive testing, and certain drug therapies, even places patients at a high physical risk.

1.2 Investigation of Healthcare Insurance Fraud

[19] developed a fraud detection system using data mining techniques. In all the categories of insurance including health insurance, fraud is a major issue. According to the researchers, fraud in health insurance is perpetrated via the intentional deception or misrepresentation of facts for gaining some shady benefit in the form of healthcare expenses. Data mining tools and techniques are used to identify fraudulent activities in large insurance claim datasets. Based on some identified cases of fraud from a sample dataset, the anomaly detection technique calculates the likelihood or probability of each record to be fraudulent by analysing the past insurance claims. The analysts can then have a closer investigation for the cases that have been marked by data mining software.

[10] developed a data mining approach for the investigation of fraud in health insurance scheme using knee-point k-means algorithm. NHIS was considered as the case study for the work. The research focuses on the use of some computer-based methods that could help to properly target investment in the healthcare sector and also drastically reduce fraud in health insurance by healthcare providers. To this effect, they applied the knee-point k-means clustering method, which was capable of detecting fraudulent claims by health service providers. Cluster-based outliers were examined. Health providers' claims submitted to HMO were grouped into clusters. Claims with similar characteristics were grouped together. The claims were grouped into two clusters: fraudulent and non-fraudulent.

In a survey of hybrid methods for the uncovering fraud in healthcare insurance by [20], acts carried out with the aim of obtaining a fraudulent outcome from health insurance processes were carefully examined. According to the researchers, when providers tries to enjoys some benefits or advantages to which they are not entitled then that attempt is considered as insurance fraud and it has become a major concern for health insurance companies. They proposed a hybrid framework that applied some data mining techniques to detect frauds. This framework considered the analysis of the characteristics of healthcare insurance data, some preliminary knowledge of healthcare system and the fraudulent behaviours. The framework harnessed the advantages of both the supervised, semi-supervised and unsupervised learning methods to identify fraudulent claims.

[21] investigated the benefits of applying big data techniques in the detection of fraud in public health insurance system in Romania. They outlined the benefits of using big data technology in combating crime in the healthcare industry and came up with the following conclusions:

i. that big data technology and its distributed processing power has taken fraud detection in healthcare insurance to a higher level. Few years ago, insurance fraud detection was not considered cost-effective as it takes time and was also too expensive, and hence, so many organisations prefer to pay claims without proper investigation.

ii. that using big data analytics approaches can culminate into speedy identification of fraudulent claims, and also generates a new set of tests to automatically reduce the section that was potentially fraudulent or detect new patterns of fraud, previously not known.

iii. that the processing of complex data using sophisticated tools reveals its huge potential, and also shows that orthodox tools for data analysis cannot handle them. The analytics tools applied in the field of healthcare insurance were briefly described, each of them being effective for a particular type of fraud or a particular stage of the fraud detection process. All these culminate into the deduction that the best solution for detecting fraud in the health insurance system is, at present, a hybrid solution, both in terms of technologies and in terms of models of analysis.

iv. that in healthcare insurance frauds are country-specific, usually based on gaps or weaknesses in the country's rules and regulations. Models are continuously changing fraud as dubious individuals are always seeking for new ways to evade the law.

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

Accordingly, approaches for detecting and preventing fraud must always be adjusted and ready to rediscover the fraudulent actions [22][23][24]. To add to the lapses in the country's constitution, each country has unique economic, political, social, and institutional opportunities for and barriers which makes fraud examination different amongst countries. A crucial and peculiar issue in the NHIS is the high level of corruption in the sector, lack of accountability and clear sense of irresponsibility [5][22].

[25] proposed a model using big data in investigating real time crime in the health insurance in the cloud. This approach utilizes fraud management solution to detect potential frauds in the cloud. The solution was based on massive amount of historical data, predictive statistical models and social network analytics. The model renders its services via client components like apps and web-services. Just like [21] and [20][25][26] did not implement any working system for their research. Also, as opined by [21], healthcare crimes are country specific and Nigeria has not adopted the cloud services and there are no healthcare laws relating to data on the cloud, and therefore, this model cannot be used to investigate crime in our healthcare system.

[27] developed a data analytics framework for Health Insurance data using Association Rule Mining. This was in a bid to identify fraudulent claims as deliberate cheating by concealing and omitting facts while claiming from health insurance providers has become in the health insurance domain which has led to significant amount of monetary loss to the providers. In view of the above, careful scanning of the submitted claim documents need to be conducted by the insurance companies in order to spot any discrepancy that indicates fraud. For this purpose, manual detection was neither easy nor practical as the claim documents received keep increasing and for diverse medical treatments. The researcher was able to detect fraudulent health insurance claims by identifying correlation or association between some of the attributes on the claim documents. With the application of a data mining techniques of evolving clustering method, association rule mining and support machine, this study was able to successfully determined correlated attributes to address the discrepancies of data in fraudulent claims and thus reduce fraud in health insurance. However, the study was used for structured data which made it unfit to be applied in big data which is highly unstructured. With the numerous data mining techniques implemented traditionally, it would consume more resources when applied to big data.

2 Materials and Methods

2.1 Data Collection

The data was collected from NHIS. Data collection was one of the most difficult task in this work as most of the data was collected manually in paper form by NHIS through HMOs and stored in file cabinets. This was done via document examination and observation which in either case, the data was collected from journals and NHIS databases. This data was classified into four (4) categories: enrollment, update, referral and claims. The formats of the data were: CSV, PDF, text, excel and images. The features of the data include: Enrollment: Name, Address, Date of birth, Sex, Next of Kin, Email address, Mobile, Telephone no. fixed, National ID no, Employer NHIS no., Date of NHIS registration, Nationality, Location of Posting, Photograph, Blood group, Genotype, Allergies, Relationship (Principal, Spouse, Child, Extra-dependant), Expiry date, Primary provider. Claim**:** Name, NHIS No. of patient, Name and NHIS No. of patient's primary provider, Name and NHIS No. of Secondary Provider, drug prescription sheet, Diagnosis/disease code, Treatment given, Date of treatment, Amount billed, Co-payment received (when applicable).

2.2 Design of the System

The design was done using Unified Modelling Language (UML). This was used to specified the developed model into a representative model to aid software system development. Two (2) different types of diagrams in UML were used to model the system: use case and sequence diagrams.

2.2.1 Use Case Diagram

The use case diagram for the system is shown in Figure 1. It has four (4) actors (*i.e.,* finance manager, data manager, contributory manager and the admin officer), and eight (8) use case (login /logout, upload data, analyse claims, analyse enrollment, analyse referral, analyse update, preprocess data and report). These are discussed as follows.

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

Login/Logout: The model is designed to be implemented in a web-based environment and hence, security is important. With this mindset, all users must possess valid user credentials before accessing the system. Before accessing the system, the user must provide these credentials and the system must acknowledge that these credentials are valid.

Upload Data: Uploading data needed for the system come in two folds. First, the existing data are digitized and uploaded to the system, and this is done by the data manager. The second method, the data is uploaded via the automated systems.

Analyse Claims: This is the processing of claims submitted for reimbursement. This produced two results: fraudulent and non-fraudulent.

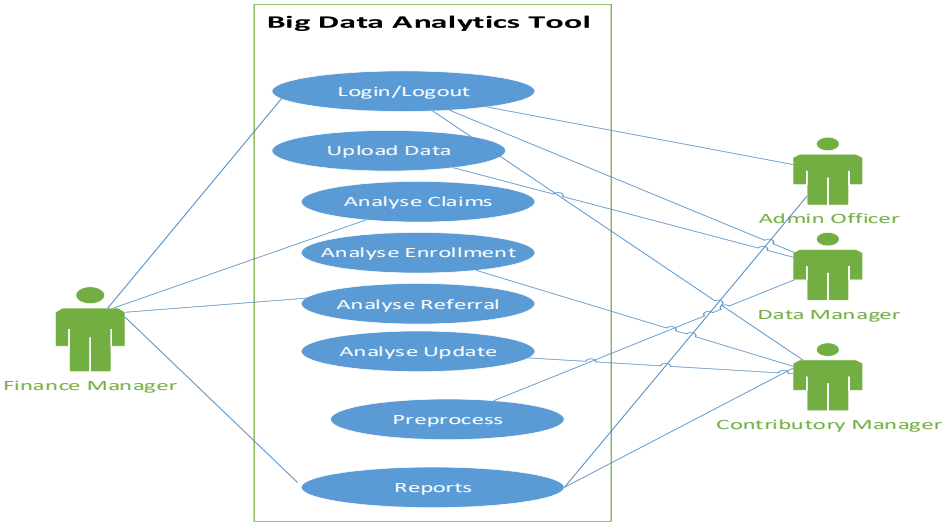

Figure 1: Use Diagram for the System

Analyse Enrollment: This is the processing of enrollment data submitted for new registration. This produced two results: fraudulent and non-fraudulent.
Analyse Update: This is the processing of update data submitted for addition of a dependant, change of primary facility or HMO *etc*. This produced two results: fraudulent and non-fraudulent.
Analyse Referral: This is the processing of referral requests made for referral of patients to higher facilities. This produced two results: fraudulent and non-fraudulent.
Preprocess Data: The cleansing of the data submitted for the processing.
Reports: These are insights generated after the processing of data.

2.2.2 Sequence Diagram

Sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the logical view of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios. Figure 2 shows the sequence diagram depicting the static features of the proposed basic flow of the model. This shows the different interactions by components of the model, the analytics and the results to be generated and their associated relationships. In the Figure 2, the data in the real world is captured and loaded into the big data warehouse (MongoDB) using the loadData() function. The data is extracted using the extractData() function and it is pre- processed and the attributes of interest are defined and captured using the dataPreprocessing() function. Also, if the data is structured, it is retrieved from the big data warehouse and preprocess directly with the aid of the dataPreprocessing() function. After preprocessing, the data is loaded into the map phase of the MapReduce where the different attributes are collated together using the collateAttributes() and rules are generated from the collated attributes using the generateRules() function. The generated rules are then pruned so as to drop the weak ones and pick the strong ones using the values of their confidence and support as computed by the pruneRules() function. These rules are used to generate insights using the generateReport() function. These insights are applied by the stakeholders using the apply() function, and

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

are also stored in the relational database (MySQL) for future reference using the storeOutput() function. To make reference to these insights, a request is sent the stakeholders using the request() function and the response to the request is sent back to the stakeholders from the database using the response() function.

3 Results

The implementation of the system was developed using Java Enterprise Edition technology. It is web-based and can be used in a distributed setting. The front-end is built using Java Server Pages, the business logic is implemented using Enterprise Java Beans. and the back-end is implemented in two folds: input storage was implemented using mongoDB and the output storage was implemented using Structured Query Language (SQL-MySQL). The input storage (MongoDB) accepts data in different formats (*e.g.,* pdf, jpeg, png, gif, csv and excel) as inputs and process them to produce report for the stakeholders which is stored in MySQL. The large input in the different formats is first divided into smaller units using the Association Rule Mining Algorithm, which is implemented in the MapReduce Framework to ease the processing of this data. The system extracts the data stored in these formats via an OCR component built in it.
This implement the parallel distributed processing model of the MapReduce using the Apriori algorithm of the association rules mining. This classes contains constants, variables and methods

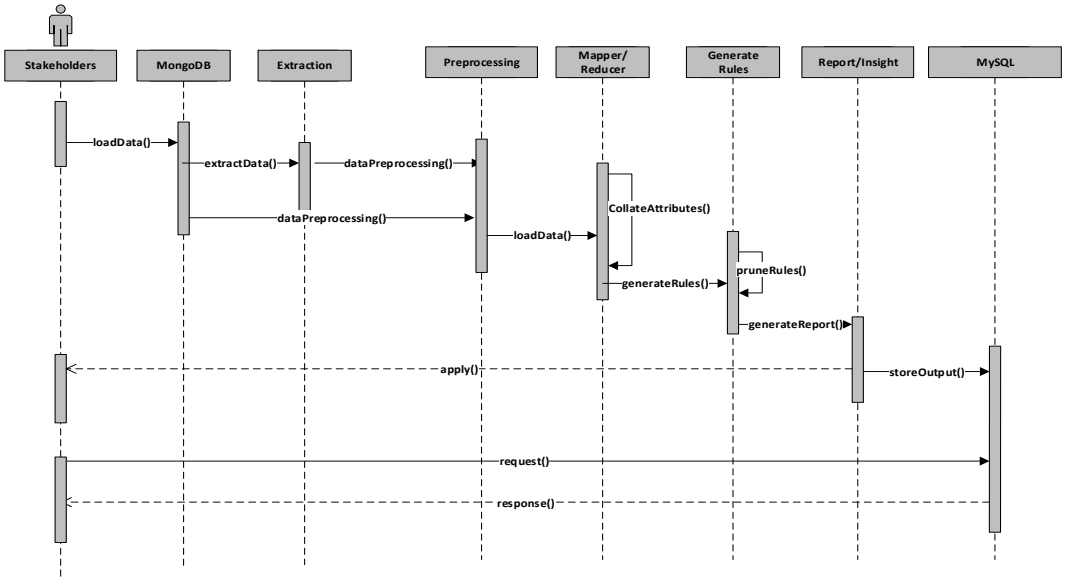

Figure 2: Sequence Diagram for the System

that are used in fraud analytics. The business logic module has a sub module known as the fraud analytics module which implement the MapReduce framework and the Apriori Association Rule Mining algorithm. Whenever a task is forwarded to this sub module, details about the task is sent to the JobTracker which coordinate, allocates and control all the activities within the analytics sub module. When these details get to the JobTracker, it assigns the TaskTrackers the job which in turns load the task unto the mapper classes. The TaskTracker also feed the JobTracker with every detail of what happens with the task, both in the mapper and the reducer classes. The relationship between the JobTracker and the TaskTracker is a master-slave one, the JobTracker is the master while the TaskTracker is the slave. The mapper and reducer classes are the kernel of the analytics process. While the mapper carryout the dirty task of the analysis, the reducer summarises whatever is produced as the outcome from the mapper. This output is sent to the relational database, MySQL, through the entity bean. It is through this same entity bean that request for retrieval of that from the database and responses to the JSPs follows.

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

The backend is designed using two different databases: MongoDB and MySQL, and temporary csv files. MySQL is a relational database management system, and it is used for storing the output of the analytics. The logic about the MySQL component of the back-end is housed in the entity bean. This component stores only processed data. The MongoDB is a NoSQL database which is used for storing big data as it is captured from the various health facilities. This is where data captured by other applications too is channelled to.

Figure 3 shows the capturing of the data into the system. Figure 4 and Figure 5 show the detection of fraud from claims and enrollment data.

4 Discussions

The modules of the application are discussed in the sections that follow below and it comprised of four modules: user management, enrollment processing, referral processing and claims processing.

User Management: This system uses single-sign-on which implies that all users must login through a single interface and are authenticated before authorizing them to access the system resources. The module is responsible for adding new staff and assigning login credentials (username and password) to them.

Claim Processing: This module handles the processing of all the claims submitted by providers for fee-for-service payment. The module verifies every item on the claim by comparing it with NHIS existing rate and anyone that fall short is either recalculated or at worse rejected. In the course of processing, all unprocessed claims in the repository are picked and process. This is to ensure that no claim spend up to fourteen days before been processed. The system was able to detect double billing, upcoding, billing for services not provided and identity theft.

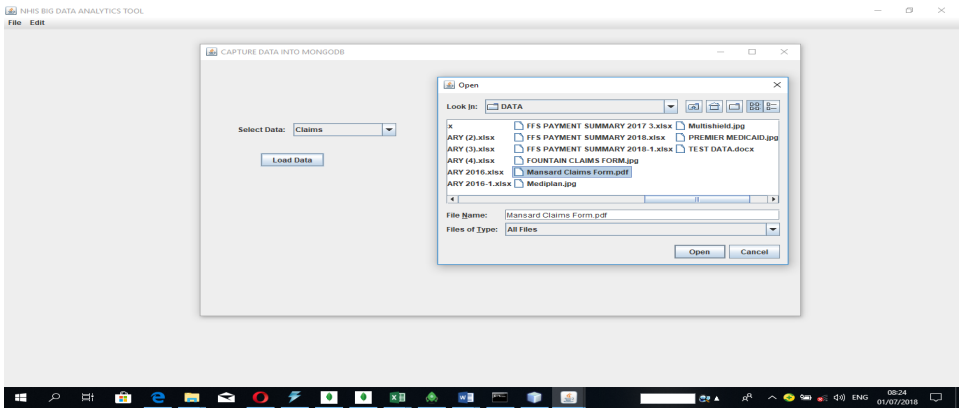

Figure 3: Data Capturing

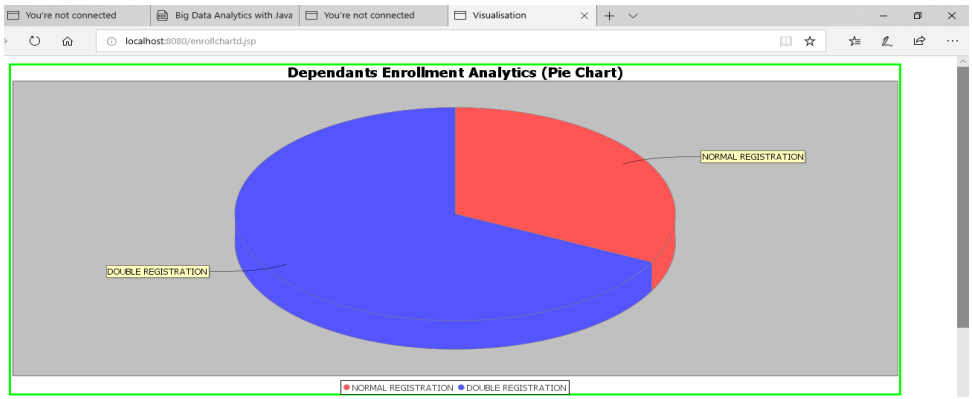

Figure. 4: Summary View of Processed Principals' Enrollment Data

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

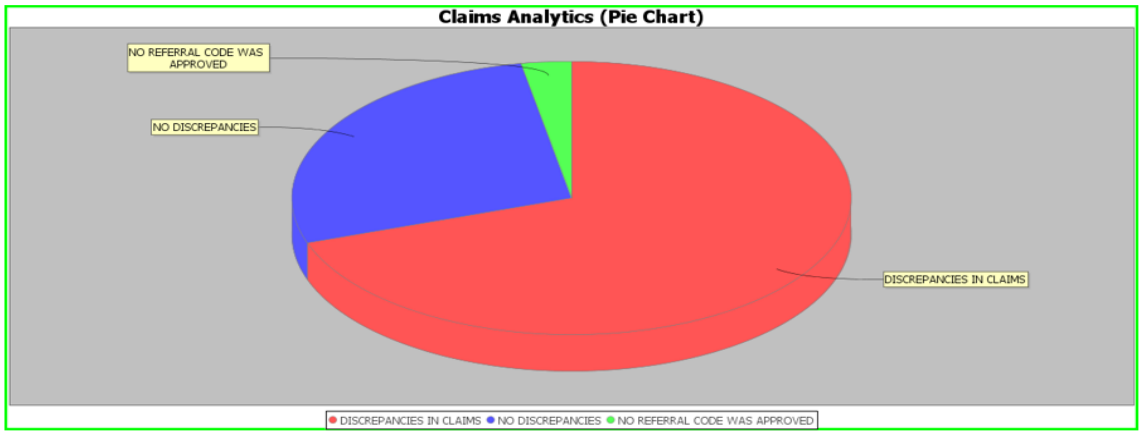

Figure. 5 Summary View of Processed Claims Data

Enrollment Processing: This module is responsible for adding or updating enrollees' information. These include personal information, next of kin information, educational information, employment information, primary providers *etc*. During enrollment, enrollees are checked to ensure that no enrollee registers twice, the number of biological children for an enrollee is not more than four. Also, that a child that is more than eighteen years is either de-register or not allowed to register. The enrollee is only allowed to register one spouse at a time. Also, there is provision to add one extra dependant which can only be allowed when the enrollee accept to be paying five thousand naira annually in addition to the normal monthly contribution. Again, enrollees who are not staying together with their dependants are allowed to have different primary facilities, otherwise, the system only allowed them to have one facility in accordance with NHIS guideline. The module was able to detect double registration of enrollees, registration more than four biological children, registration of children that are more than eighteen years and ghost enrollees.

Referral Processing: The module is responsible for the handling cases of referrals. A beneficiary is referred to another NHIS accredited hospital when the medical condition of the beneficiary is more than what the primary provider can handled. The principle of referral is that a beneficiary should be referred to the next closest NHIS accredited facility for a more specialized care. When referrals are made, the system evaluates the diagnoses in the referral request and compares it with what allowed by NHIS to be handled by the next level of healthcare. It also checks the distance between the facility they are referring the patient to and other facilities around the same location with the primary facility that is making the referral on behalf of the enrollee. In an event where the distance is more than the other facilities, such referrals are queried. Also, if the medical condition of the enrollee is not more than what should be handled at primary facility, they are also queried. Again, if the diagnoses do not tally with what the next hospital is registered to carry out, it is also queried. Otherwise, the NHIS guideline stipulates that referrals should not last for more than forty-eight hours with the HMOs. The module was able to detect self-referral, and collusion with providers.

5 Conclusion

In conclusion, this paper developed a system for the detection of the fraudulent activities of the actors of NHIS. This system employed data from the Nigerian NHIS which was categorised into: enrollment, referral and claim data with different file formats: pdf, jpg, png, csv and excel. The system was designed using UML and implemented using Association Rule Mining, MapReduce Framework, MongoDB, MySQL and Java Programming Language. The system was able to detect several fraudulent activities of providers, insurers and beneficiaries. These ranges from self-referral, collusion with providers, double registration of enrollees, registration more than four biological children, registration of children that are more than eighteen years, ghost enrollees, double billing, upcoding, billing for services not provided to identity theft. This system when adopted by NHIS, it will sanitise the entire Scheme.

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256

References

[1] Olaniyan AO. Assessment of the implementation of national health insurance scheme (NHIS) in south-western nigeria. Unpublished PhD Thesis submitted to the Department of Public Administration, Obafemi Awolowo University, Ile-Ife, Nigeria, 2017.

[2] Etobe EI, Etobe UE. The national health insurance scheme and its implication for elderly care in Nigeria. IJSR. 2015; 4(2): 128-132.

[3] Eteng FO, Ijim-Agbor U. Understanding the challenges and prospects of administering the national health insurance scheme in Nigeria. IJHSSR. 2016; 2(8): 43-48.

[4] NHIS (National Health Insurance Scheme) (2013). National health insurance scheme operational guidelines. Available from: http://www.nhis.gov.ng/images/stories/hmoregister/NHIS_OPERATIONAL_GUI DELINES.pdf.

[5] Agba MO, Ushie EM, Osuchukwu NC. National Health Insurance Scheme (NHIS) and employees' access to healthcare services in Cross River State, Nigeria. GJHSS. 2010; 10(7): 9-16.

[6] Oyegoke TO, Ikono RN, Soriyan HA. An integrated health management system for national health insurance scheme in Nigeria. JETCIS. 2017; 8(1): 30-40.

[7] Alimi OM, Binuyo OG, Gambo IP, Jimoh K. Realtime national health insurance scheme (RNHIS): Means to achieve health for all. IJCSEA. 2016; 6(2): 1-8.

[8] NHCAA (National Health Care Anti-Fraud Association) (2012). "What is health care fraud? Available from: [www.nhcaa.org/resources/health-care-anti-fraud-resources/consumer-info- action.aspx]

[9] Thornton D, Brinkhuis M, Amrit C, Aly R. Categorizing and describing the types of fraud in healthcare. PCS. 2015; 64: 713 – 720.

[10] Fashoto SG, Owolabi O, Sadiku J, Gbadeyan JA. Application of data mining technique for fraud detection in health insurance scheme using knee-point and k-means algorithm. AJBAS. 2013; 7(8): 140-144.

[11] Bagde PR, Chaudhari MS. Analysis of fraud detection mechanism in health insurance using statistical data mining techniques. IJCSIT. 2016; 7(2): 925-927.

[12] Li J, Huang K-Y, Jin J, Shi J. A survey on statistical methods for healthcare fraud detection. HCMS. 2008; 11: 275-287.

[13] Dora P, Sekharan GH. Healthcare insurance fraud detection leveraging big data analytics. IJSR. 2015; 4(4): 2073-2076.

[14] Ekin T, Ieva F, Ruggeri F, Soyer R. Applications of bayesian methods in detection of healthcare frauds. CET. 2013; 33: 151-156.

[15] Musal R. Two models to investigate medicare fraud within unsupervised databases. ESA. 2010; 37(12): 8628-8633.

[16] Liu Q, Vasarhelyi M. Healthcare fraud detection: A survey and a clustering model incorporating geo-location information. In the Proceedings of the 29th World Continuous Auditing and Reporting Symposium November 21-22, 2013, Brisbane, Australia.

[17] Jacqulin MJ, Shrijina S. Implementation of data mining in medical fraud detection. IJCA. 2013; 69(5): 1-4.

[18] Travaille P, Thornton D, Müller MR, Hillegersberg J. Electronic fraud detection in the u.s. medicaid healthcare program: Lessons learned from other industries. In the Proceedings of the Seventeenth Americas Conference on Information Systems, Detroit, Michigan August 4th-7th 2011.

[19] Kirlidog M, Asuk C. A fraud detection approach with data mining in health insurance. SBS. 2012; 62: 989 – 994.

[20] Bagul PD, Bojewar S, Sanghavi A. Survey on hybrid approach for fraud detection in health insurance. IJIRCCR. 2016; 4(4): 6918-6922.

[21] Bologa A, Bologa R, Florea A. Big data and specific analysis methods for insurance fraud detection. DSJ. 2010; 1(1): 30-39.

[22] Arodiogbu IL. Introducing social health insurance to solve problems of poor health sector financing in Nigeria. Unpublished MSc Thesis Submitted to the Department of Health Management, Planning and Policy, University of Leeds, 2005.

[23] Dutta A, Hongoro C. Scaling up national health insurance in nigeria: Learning from case studies of india, colombia, and Thailand. Washington, DC: Futures Group, Health Policy Project, 2013.

[24] Yunusa U, Irinoye O, Suberu A, Garba AG, Timothy G, Dalhatu A, Ahmed S. Trends and challenges of public health care financing system in nigeria: The way forward. IOSR. 2014; 4(3): 28-34.

[25] Konasani V, Biswas M, Koleth PK. Healthcare fraud management using big data analytics. An Unpublished Report by Trendwise Analytics, Bangalore, India, 2012.

[26] Oyegoke TO. Development of an integrated health management system for national health insurance scheme. An Unpublished M.Sc. Thesis Submitted to the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria, 2015.

[27] Kareem S, R. B. Ahmad RB, Sarlan AB. Framework for the identification of fraudulent health insurance claims using association rule mining. 2017 IEEE Conference on Big Data and Analytics (ICBDA). 2015; 99-104.

 J Health Inform Afr. 2019;6(2):64-73. DOI: 10.12856/JHIA-2019-v6-i2-256