ML Based Smart Health Insurance System

1st Shruthi G R
Information Science and Engineering
New Horizon College of Engineering
Bengaluru, India
shruthigr.nhce@newhorizonindia.edu

2nd Sanjana Hombal
Information Science and Engineering
New Horizon College of Engineering
Bengaluru, India
sanjanahombal3@gmail.com

3rd Sanchitha B S
Information Science and Engineering
New Horizon College of Engineering
Bengaluru,India
sanchithashetty95@gmail.com

4th Shreya L
Information Science and Engineering
New Horizon College of Engineering
Bengaluru, India
shreyalingareddy81@gmail.com

5th Sharanya G
Information Science and Engineering
New Horizon College of Engineering
Bengaluru, India
sharu.g.2000@gmail.com

***Abstract*— The healthcare industry is an intricate system with numerous moving components. It is expanding at a perditions pace. At the same time, fraud in this industry is turning into a critical problem. One of the issues is the misuse of the medical insurance systems. Manual detection of frauds in the healthcare industry is strenuous work. Healthcare fraud is increasingly apperceived as one of the serious social concerns. Clearly, healthcare fraud is a problem for the government and there is a need for more effective detection methods. The main objective is to detect whether the claims done by the doctors in the hospital are genuine or fraudulent. To overcome the misuse of health insurance, our team has proposed a fraud detection system using Machine Learning. This verification is done by the insurance company by checking the documents uploaded by the hospital. Our model will be trained to detect whether it is possible for the patient to have the claimed medical condition. It also includes a module that has all the costs of the treatments. This is compared with the budget claimed by the hospital. To implement this project we are building a portal that can be accessed by the Hospital and the Insurance Company. Through this portal the hospital can claim for approval of insurance for each patient, by uploading the medical background and test reports along with the required budget estimation.**

Keywords—component, formatting, style, styling, insert

I. INTRODUCTION

Health Insurance is bond given by the insurance company where the company assures to take care of all the medical expenses and pays bill anytime when the customer encounters any uncertainty in their health or meets with accident. Insurance companies are like third party where they provide help to insured customers. They will be tie-upped with numerous hospitals. So, when anyone who has taken the insurance gets hospitalized due to any reason, they provide cashless treatment to them as they own insurance. If the person decides to admit to a hospital which is not tie-upped with that insurance company, them the amount which is essential for the surgery and hospitalization will be compensated to their account based on the submissions. There is even little insurance which government provides where the premium amount will be deducted in form of taxes from their income.

These days since health is being a major concern and uncertainty is increasing, it's important for one to own any health and life insurance. When we buy any insurance, we can add dependents onto the plan so that the insurance plan can be covered to ourselves and to our family as well. By doing so we can save a lot of money as we all know medical industry is too expensive. Even though people feel visiting hospital is more like burning a hole in our own pockets as it eats up all our savings and income. Health insurance play's crucial character in each and every family but also plays even big crucial part in those families only one man is earning and running the family. In a scenario where the family man is itself hospitalized, with insurance cannot offered the price. So, it's better everyone own at least one insurance and pay the annual premium which will reduce their burden and stress at any emergencies.

The coverage of medical insurance various from the policy to company from whom we buy it. This coverage includes hospitalization, the necessary tests, medicine, pre and post treatment doctor consultation and hospital visits until the patient is recovered. Compound of hardware and software technology that process and procedures all the maintenances of stocked products, whether products are sent to vendors or if there are enough raw material supplies and so on.

As we speak about insurance, it's even important in choosing a good policy from a good insurance company. We need to investigate well before investing on any insurance. Few benefits a good insurance company provides us:

- Have better network with all the hospitals which provides us better treatment and with upgraded facility.
- The policies which are in long term and are covered even in our old age.
- No matter how critical condition we are in, they should be no increase in the policy premium during the term.
- A policy protects the entire family at critical illness.
- Should provide us the flexibility to choose health plan.

II. LITERATURE SURVEY

In paper [1], " Fraud Detection of Medical Insurance Employing Outlier Analysis ", Fraud detection is an important issue in the area of data science, and it has a lot of practical applications in related fields, such as business, health, and environment. Most traditional methods detect fraud based on rulemaking. Unfortunately, it is not always useful in the medical field since the boundary of fraud detection is vague. As a result, outlier detection is a promising method..

In paper[2] " Healthcare Fraud Detection Based on Trustworthiness of Doctors ", Big data is now rapidly expanding into various domains such as banking, insurance and e-commerce. Data analysis and related studies have attracted more attentions. In health insurance, abuse of

diagnosis is one of the key fraud problems, which damages the interests of insured people. To address this issue, numbers of studies have focused on this topic. This paper develops a healthcare fraud detection approach based on the trustworthiness of doctors to distinguish fraud cases from normal records.

In paper [3], " Healthcare Insurance Frauds: Taxonomy and Blockchain-Based Detection Framework ", Medical health insurance fraud has been a major concern for the healthcare industry and governmental institutions. In the United States, the health insurance companies recorded a loss of tens of billions yearly due to healthcare fraud. Some types of fraud are at the risk of the patient's health.

In paper [4], " Research on Recognition of Medical Insurance Fraud Based on Modified Support Vector Machine", With the rapid development of China's medical insurance industry, fraud irregularities are also constantly increasing with renovation of forms and means. Then the way of detecting the fraud automatically and efficiently has become the major issue on which much attention is paid. This paper selects genetic algorithm (GA) and particle swarm optimization (PSO) as the basic algorithm..

In paper [5]," Heath care fraud detection methods and new approaches ", Fraud and abuse have caused significant cost in the health care systems of many countries. This paper aims to provide a review on health care fraud detection, based on big data approach. For this reason, available articles in the literature were reviewed with focuses on classifying techniques, identifying the major sources and characteristics of the data. The recent solutions and challenges in health care fraud and big data analytics were evaluated within some certain criteria.

In paper [8], " Big Data-Driven Abnormal Behavior Detection in Healthcare Based on Association Rules" Healthcare insurance frauds are causing millions of dollars of public healthcare fund losses around the world in various ways, which makes it very important to strengthen the management of medical insurance in order to guarantee the steady operation of medical insurance funds.

In paper [6], " A Fraud Detection Approach with Data Mining in Health Insurance" this paper aims to extend the platform by integrating wearable and unobtrusive sensors to monitor patients with coronavirus disease.

In paper [7], In this paper, Health claim frauds are affecting the economic status of developing as well as developed countries. Health care fraud detection is now becoming more and more important. In order to detect and avoid fraud we are going to use data mining techniques. We have proposed a Hybrid model system consisting of classification and clustering.

In paper [9], " Naïve Classification Approach for Insurance Fraud Prediction" An approach which can be used for the prediction of future potentials on the basis of present information is known as prediction analysis. This study is relied on the fraudulent discovery in the insurance business. A number of approaches have been projected up to now for the fraudulent discovery in insurance sector. These approaches mainly rely on machine learning algorithms. The insurance fraud detection is the major issue of prediction analysis.

In paper [10], Fraud can be spread broadly and it is extremely costly to the therapeutic protection framework. Unscrupulous protection might be a case created to cover up or twist information that is intended to deliver social insurance edges. Cheats might be of the numerous sorts and submitted by the protection guarantor or the safeguarded. The unscrupulous social insurance providers are the reason for extortion in the wellbeing segment. The commitment of this case misrepresentation discovery is Associate in nursing trial study on extortion recognizable proof and exploitative examples.

In paper [11], " A Survey Paper on Fraud Detection and Frequent Pattern Matching in Insurance claims", Fraudulent insurance claims increase the burden on society. Frauds in health care systems have not only led to additional expenses but also degrade the quality and care which should be provided to patients. Insurance fraud detection is quite subjective in nature and is fettered with societal need. This empirical study aims to identify and gauge the frauds in health insurance data. The contribution of this insurance claim fraud detection experimental study untangle the fraud identification frequent patterns underlying in the insurance claim data using rule based pattern mining.

In paper [12], " Effective Fraud Detection in Healthcare Domain using Popular Classification Modeling Techniques." Fraud is any activity with malicious intentions resulting in personal gain. In the Present Day scenario, every sector is polluted by such fraudulent activities to fetch unauthorized benefits. In HealthCare, an increase in fraudulent insurance claims has been observed over the years which may constitute around 3-5% of the total cost.

In paper [20], " Effective Fraud Detection in Healthcare Domain using Popular Classification Modeling Techniques", Fraud is any activity with malicious intentions resulting in personal gain. In the Present Day scenario, every sector is polluted by such fraudulent activities to fetch unauthorized benefits. In HealthCare, an increase in fraudulent insurance claims has been observed over the years which may constitute around 3-5% of the total cost. Increasing healthcare costs along with the hike in fraud cases have made it difficult for people to approach these services when required.

III. PROPOSED METHOD

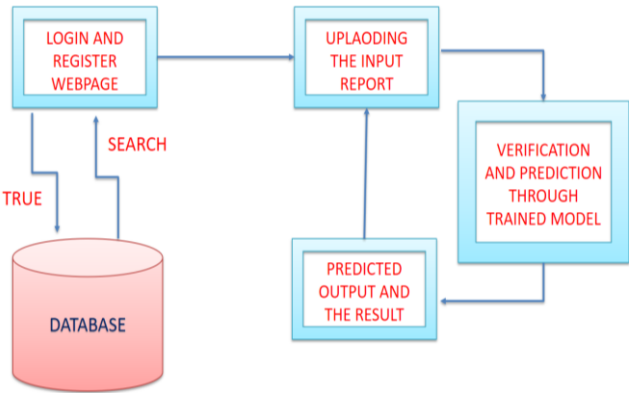


Fig 3.1 Architectural Design of the system.

To implement our project, we are building a web portal which can

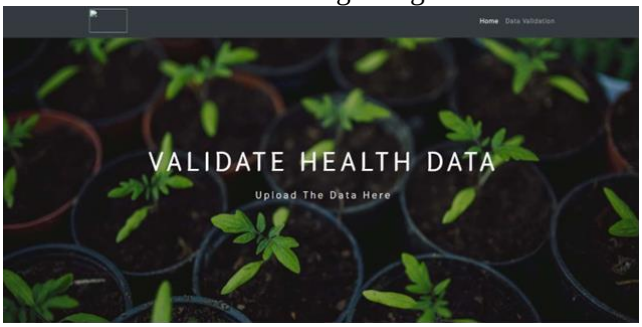
be accessible hospital staff and the insurance company team. Through the portal hospital can claim all the necessary needed and get the approval from the insurance company.
The hospital team will need to upload all the necessary documents along with excel which contains the patient's insurance ID, symptoms, and the cost estimation. Once the data is received by the insurance company, it will be feed as input to the model we have built which will detect what kind of claim it is. To verify the details uploaded it will be decided based on the following criteria:

- The patient should be hospitalized for minimum of 24 hours. The medical history, personal information, and the medical condition will be verified.
- Based on the model built by us which contains the cost estimation for each treatment, the claims budget will be cross verified.

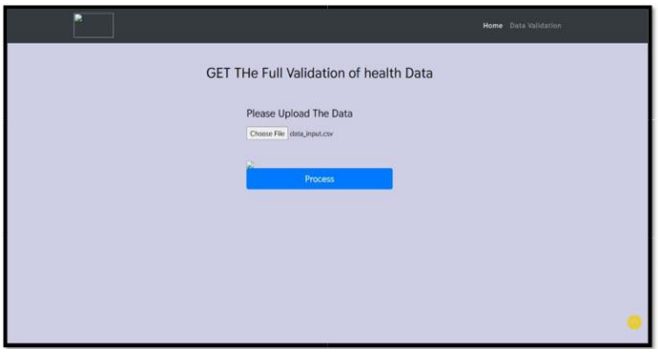
Fraud is the major concern in any industry. It happens anywhere and everywhere. So, our main aim in this project was to ensure that none of the fraud medical claims will not approved by the insurance company. This can also be financial benefit to lower class people where they value even each rupee sent on anything. If we continue supporting fraud for someone else's benefit, it will in turn be a hole in the industry. Fraud detection basically involves multiple activates which takes place to prevent income/finance being drained by fake means or claims. Even though many industries cross check before processing their business, still fraud happens. So, we need to closely monitor and double triple check and then make any move future.

IV. EXPERIMENTAL RESULTS

1. Login Page



2. Introduction Page



3. Page to upload CSV file

	A	B	C	D	E	F
1	patient_id	Symptom_1	Symptom_2	Symptom_3	Disease	Cost
2	0	itching	skin_rash	nodal_skin_eruptions	Fungal infection	29300
3	1	skin_rash	nodal_skin_eruptions	dischromic _patches	Fungal infection	28000
4	2	itching	nodal_skin_eruptions	dischromic _patches	Fungal infection	26000
5	3	itching	skin_rash	dischromic _patches	Fungal infection	44000
6	4	itching	skin_rash	nodal_skin_eruptions	Fungal infection	30000
7	5	skin_rash	nodal_skin_eruptions	dischromic _patches	Fungal infection	61599
8	6	itching	nodal_skin_eruptions	dischromic _patches	Fungal infection	35718
9	7	itching	skin_rash	dischromic _patches	Fungal infection	81870
10	8	itching	skin_rash	nodal_skin_eruptions	Fungal infection	110777
11	9	itching	skin_rash	nodal_skin_eruptions	Fungal infection	95544
12	10	continuous_sneezing	shivering	chills	Allergy	86976
13	11	shivering	chills	watering_from_eyes	Allergy	118925

4. Sample CSV file

	A	B	C	D	E	F	G	H	I	J	K
1		patient_id	Symptom_	Symptom_	Symptom_	Disease	Cost	validation_by_disease	vaildation_by_cost	final_validation	
2	0	0	itching	skin_rash	nodal_skii	Fungal infe	29300	valid	valid	valid	
3	1	1	skin_rash	nodal_skii	dischromi	Fungal infe	28000	valid	valid	valid	
4	2	2	itching	nodal_skii	dischromi	Fungal infe	26000	valid	valid	valid	
5	3	3	itching	skin_rash	dischromi	Fungal infe	44000	valid	invalid	invalid	
6	4	4	itching	skin_rash	nodal_skii	Fungal infe	30000	valid	invalid	invalid	
7	5	5	skin_rash	nodal_skii	dischromi	Fungal infe	61599	valid	invalid	invalid	
8	6	6	itching	nodal_skii	dischromi	Fungal infe	35718	valid	invalid	invalid	
9	7	7	itching	skin_rash	dischromi	Fungal infe	81870	valid	invalid	invalid	
10	8	8	itching	skin_rash	nodal_skii	Fungal infe	110777	valid	invalid	invalid	
11	9	9	itching	skin_rash	nodal_skii	Fungal infe	95544	valid	invalid	invalid	
12	10	10	continuou	shivering	chills	Allergy	86976	valid	invalid	invalid	
13	11	11	shivering	chills	watering_	Allergy	118925	valid	invalid	invalid	

5. Updated result

V. DISCUSSION

The developments that can be made on the current implementation is as stated below:

We developed a model that can detect fraud insurance claims in the healthcare area, using K-means clustering, and Logistic Regression and Decision Tree algorithms. The much more major benefits of this technique are that it needs a basic data type, has excellent viability, and produces improved consistency and recall than the standard regulation method, which reduces the workload of data analysts.

In order to increase the model's performance, we limited the lot of information as well as the amount of parameters and restrictions. As a result of the increased accuracy, this anomalous warning system can only discover roughly 71% of records. We can only validate the selected outcome and test on our 1 entries database or run enough before, which are not thorough enough to include every parameter due to the lack of limitations and the enormous amount of data.

On enhancing the model with larger datasets and constraints using various algorithms, an API can be developed for the health insurance companies so that they can use it to detect fraud claims effectively and accurately.

VI. CONCLUSION

Health unit fraud, forms of medical management fraud, sources also types of medical management data, and healthcare fraud tactics are introduced here. In the theory, various theories are discussed. It has been determined as data is a critical concern in the medical management industry. The highest information is derived from government and private insurance companies. In order to detect healthcare fraud, machine learning algorithms are commonly utilized. Machine learning approaches fall into three types: supervised, unsupervised, and semi-supervised learning. Various investigators employ semisupervised learning methods in the vast most of cases. New semi-supervised learning algorithms will be offered in a few circumstances to detect frauds in the healthcare system more efficiently.

However, there is a cover-up plan in place to hide all cases of healthcare fraud.However, there is no one-size-fits-all technique or pattern for concealing all instances of healthcare fraud. This review concludes that sophisticated techniques and recently bought sources of healthcare data will be hot topics in the near future in order to make healthcare more affordable, maximize the productivity of healthcare fraud detection, and ensure that healthcare systems are of the highest quality.

REFERENCES

[1]. Fraud Detection of Medical Insurance Employing Outlier Analysis byJinfeng Peng; Qingzhong Li; Hui Li; Lei Liu; Zhongmin Yan; Shidong Zhang

[2]. Healthcare Fraud Detection Based on Trustworthiness of DoctorsBy Haoyi Cui; Qingzhong Li; Hui Li; Zhongmin Yan

[3]. Healthcare Insurance Frauds: Taxonomy and Blockchain-Based Detection Framework (Block-HI)by Leila Ismail; Sherali Zeadally

[4]. Research on Recognition of Medical Insurance Fraud Based on Modified Support Vector Machine by Yifeng Dou; Hailing Xiong

[5]. Heath care fraud detection methods and new approachesby Ebru Aydoğan Duman; Şeref Sağıroğlu

[6]. Big Data-Driven Abnormal Behavior Detection in Healthcare Based on Association Rules by SHENGYAO ZHOU, JIE HE, HUI YANG, DONGHUA CHEN , RUNTONG ZHANG

[7]. Fraud Detection in Health Insurance using Hybrid Systemby Thotakura Lalithagayatri, Tawde Priyanka, Aruna Pavate

[8].A Fraud Detection Approach with Data Mining in Health Insuranceby Melih Kirlidog,Cuneyt Asuk

[9]. Naïve Classification Approach for Insurance Fraud Predictionby Bhavna Batra, Sheetal Kundra

[10]. A Survey Paper on Fraud Detection and Frequent Pattern Matching in Insurance claims using Data Mining TechniquesBy Pinak Patel, Siddharth Mal, Yash Mhaske