International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 20, Issue 5 (May. 2024), PP. 117-128

Medicare Hospitality Fraud Detection

[1.]AMGOTH NARESH [2.]Pavani Manaswi [3.]P.V.M Rohan
B-Tech,4th yearstudentsDepartmentofCyber Security (CSE)SphoorthyEngineeringCollege.
,Under the Supervision of Mr. Nasair Uddin Khan (Assistant professor)

ABSTRACT
With the overall increase in the elderly population come additional, necessary medical needs and costs. Medicare is a U.S. healthcare program that provides insurance, primarily to individuals 65 years or older, to offload some of the financial burden associated with medical care. Even so, healthcare costs are high and continue to increase. Fraud is a major contributor to these inflating healthcare expenses. Our paper provides a comprehensive study leveraging machine learning methods to detect fraudulent Medicare providers. We use publicly available Medicare data and provider exclusions for fraud labels to build and assess three different learners. In order to lessen the impact of class imbalance, given so few actual fraud labels, we employ Logistic Regression creating two class distributions. Our results show that the other algorithms have poor performance compared with Logistic Regression. Learners have the best fraud detection performance, particularly for the 80:20 class distributions with average AUC scores, respectively, and low false negative rates. We successfully demonstrate the efficacy of employing machine learning Models to detect Medicare fraud.

Date of Submission: 06-05-2024 Date of Acceptance: 19-05-2024

I. INTRODUCTION

Healthinsurersreceivemillionsofclaimsperyear.Giventhatinformationas symmetriesbetweentheprincipal(insurer)andtheagents(healthcareprovidersandtheinsured)canleadtomoralhazard,insurancecompaniesfacethechoiceofeitherpayingoutinsuranceclaimsimmediatelywithoutanyadjustmentsorreviewingclaimsthataresuspicious.Themostcommonmethodfor undertakingthelatterinvolvesmanuallyauditingclaimsdata,whichisatime-consumingandexpensiveprocess.Machinelearningmodelscangreatlycutauditingcostsbyautomaticallyscreeningincomingclaimsandflaggingupthosethataredeemedtobesuspicious–i.e.,potentiallyincorrect– forsubsequentmanualauditing.Insurance fraud is a widespread and high-priced problem for each policyholder and insurance businesses in all sectors of the coverage industry. India is one of the quickest developing economies in the international, has a burgeoning middle class, and has witnessed a giant upward push within the demand for medical insurance products.

PROJECT OBJECTIVES

Provider Fraud is one of the biggest problems facing Medicare. According to the government, the total Medicare spending increased exponentially due to frauds in Medicare claims.

Healthcare fraud is an organized crime which involves peers of providers, physicians, beneficiaries acting together to make fraud claims.

Healthcarefraudandabusetakemanyforms.Someofthemostcommontypesoffraudsbyprovidersare:

- Billingforservicesthatwerenotprovided.
- Duplicatesubmissionofaclaimforthesameservice.
- Misrepresentingtheserviceprovided.
- Chargingforamorecomplexorexpensiveservicethanwasprovided.
- Billingforacoveredservicewhentheserviceprovidedwasnotcovered.

PROBLEM STATEMENT

The goal of this project is to "predict the potentially fraudulent providers "based on the claims filed by them, along with this, we will also discover important variables helpful in detecting the behavior of potentially fraud providers.

Introduction to the Dataset For the purpose of this project, we are considering Inpatient claims, Outpatient claims and Beneficiary details of each provider.

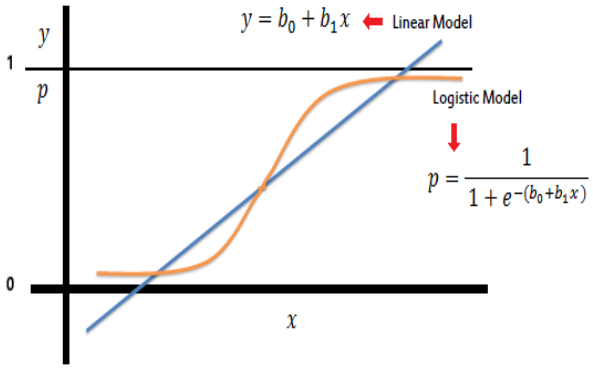
II. LITERATURE SURVEY

Theexistingproposedhealthcarefrauddetectiontacticswithintheliteraturecanbeclassifiedasthreeclasses:supervisedapproachwhichincludedecisiontreeandneuralcommunity,usedwhilehistoricfraudfactsistobehadandlabelled;unsupervisedmethodalongwithclustering,usedwhenthere'sno labelled ancientfraudrecords;andhybridtechniquewhichcombinesupervisedandunsupervisedmethodsandnormallyuseunsupervisedprocessestoenhancetheoverallperformanceofsupervisedapproach.
Thispaperobjectiveistoperceivehealthcarefraudulentbehaviour,analysethecharacteristicsofhealthcarestatisticsandoverviewandcomparecurrentlyproposedfrauddetectiontacticstheusageofhealthcarefactsaswellastheircorrespondingfactspreprocessandtalkthefutureresearchdirectionsSpecially,thispaperstartsoffevolvedwithaheritageinformation adventpeoplefitnesscaremachineanditsfraudbehaviour.

EXISTING SYSTEM

Logistic Regression

Logistic regression predicts the probability of an outcome that can only have two values (i.e. a dichotomy). The prediction is based on the use of one or several predictors (numerical and categorical). A linear regression is not appropriate for predicting the value of a binary variable for two reasons:



- A linear regression will predict values outside the acceptable range (e.g. predicting probabilities outside the range 0 to 1)
- Since the dichotomous experiments can only have one of two possible values for each experiment, the residuals will not be normally distributed about the predicted line.
- On the other hand, a logistic regression produces a logistic curve, which is limited to values between 0 and 1. Logistic regression is similar to a linear regression, but the curve is constructed using the natural logarithm of the "odds" of the target variable, rather than the probability. Moreover, the predictors do not have to be normally distributed or have equal variance in each group.

DRAWBACKS OF EXISTING SYSTEM

- If the number of observations is lesser than the number of features, Logistic Regression should not be used,

otherwise, it may lead to overfitting.

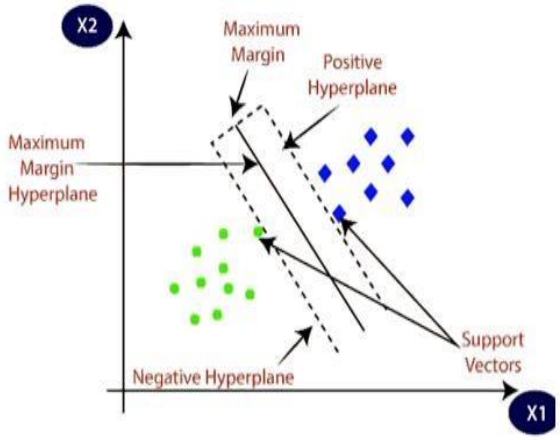
- It can only be used to predict discrete functions. Hence, the dependent variable of Logistic Regression is bound to the discrete number set.
- Non-linear problems can't be solved with logistic regression because it has a linear decision surface. Linearly separable data is rarely found in real-world scenarios.
- In Linear Regression independent and dependent variables are related linearly. But Logistic Regression needs that independent variables are linearly related to the log odds (log(p/(1-p)).

PROPOSED SYSTEM

Support Vector Machine Algorithm

- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is usedfor Classification as well as Regression problems. However, primarily, it is used for Classification problemsin Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.
- SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are

called as support vectors, and hence algorithm is termed as Support Vector Machine.Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



APPLICATIONS OF SVM ALGORITHM

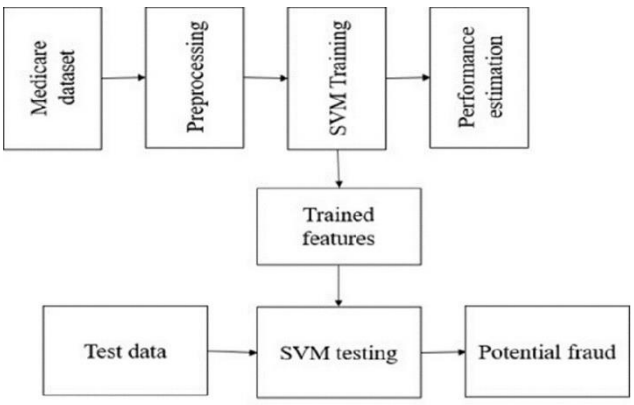
- Face recognition
- Weather prediction
- Medical diagnosis
- Spam detection
- Age/gender identification
- Language identification
- Sentimental analysis
- Authorship identification
- News classification

ADVANTAGES OF PROPOSED SYSTEM

- SVM works relatively well when there is a clear margin of separation between classes.
- SVM is more effective in high dimensional spaces.
- SVM is effective in cases where the number of dimensions is greater than the number of samples.
- SVM is relatively memory efficient.

IMPLIMANTATION

DATA PRE-PROCESSING IN MACHINE LEARNING ARCHETECTURE



Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

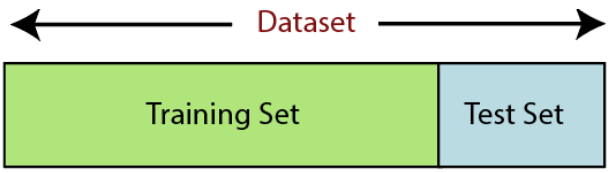
When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

WHY DO WE NEED TO PRE-PROCESSING?

A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

- Getting the dataset
- Importing libraries
- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set
- Feature scaling

Splitting the Dataset into theTraining set and Test set



Training Set:

A subset of dataset to train the machine learning model, and we already know the output.

Test set:

A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

SYSTEM DESIGN:

UML Diagrams Overview

UML combines best techniques from data modeling (entity relationship diagrams), business modeling (workflows), object modelling, and component modelling. It can be used with all processes, throughout the software development life cycle, and across different implementation technologies. UML has synthesized the notations of the Booch method, the Object-modeling technique (OMT) and Object-oriented software engineering (OOSE) by fusing them into a single, common, and widely usable modelling language. UML aims to be a standard modelling language which can model concurrent and distributed systems.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modelling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

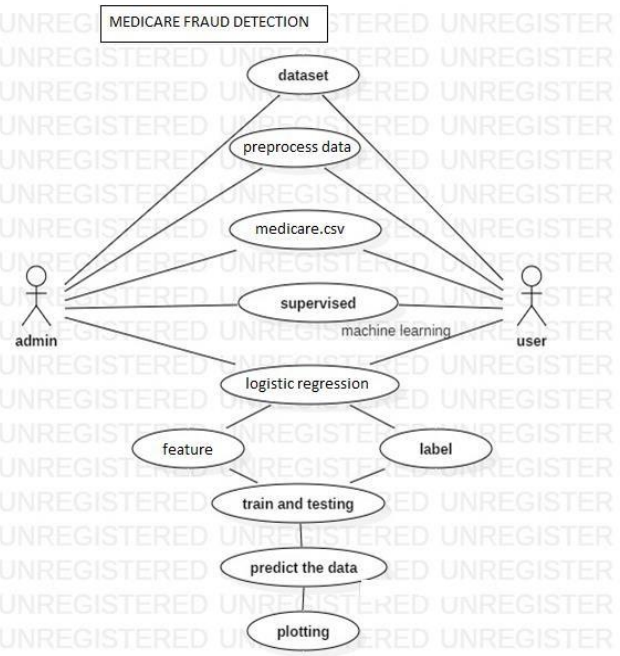
GOALS

The Primary goals in the design of the UML are as follows:

- Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
- Provide extendibility and specialization mechanisms to extend the core concepts.
- Be independent of programming languages and development process.
- Provide a formal basis for understanding the modelling language.
- Encourage the growth of tools market.
- Support higher level development concepts such as collaborations, frameworks,patterns, and components.
- Integrate best practices.

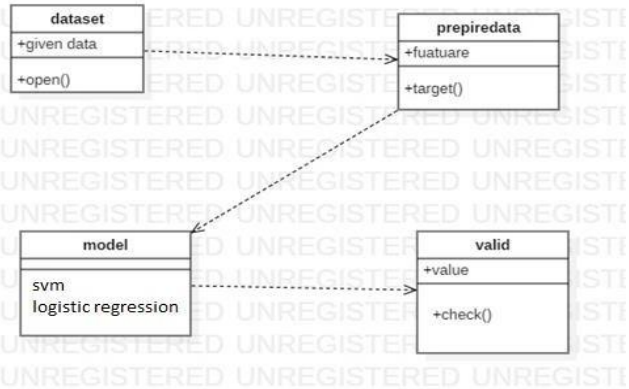
USE CASE DIAGRAM

A use case diagram in the Unified Modelling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



CLASS DIAGRAM

The class diagram is used to refine the use case diagram and define a detailed design of the system. The class diagram classifies the actors defined in the use case diagram into a set of interrelated classes. The relationship or association between the classes can be either an "is-a" or "has-a" relationship. Each class in the class diagram may be capable of providing certain functionalities. These functionalities provided by the class are termed "methods" of the class. Apart from this, each class may have certain "attributes" that uniquely identify the class.

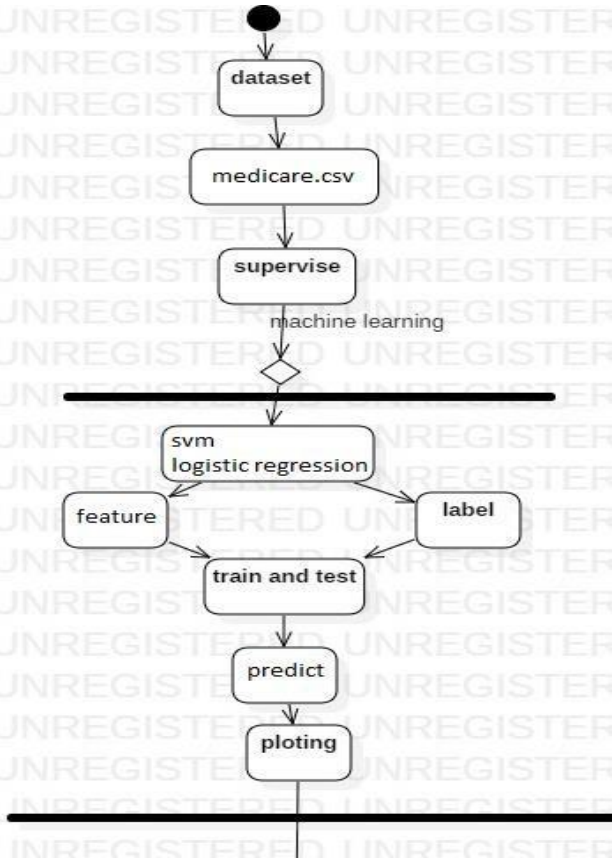


ACTIVITY DIAGRAM

Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

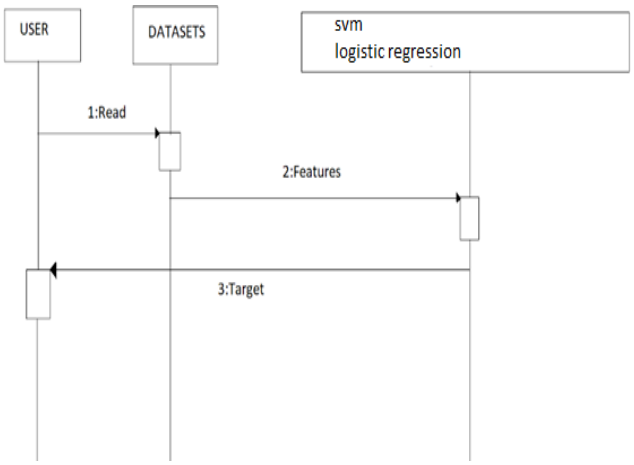
Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams can be used to describe

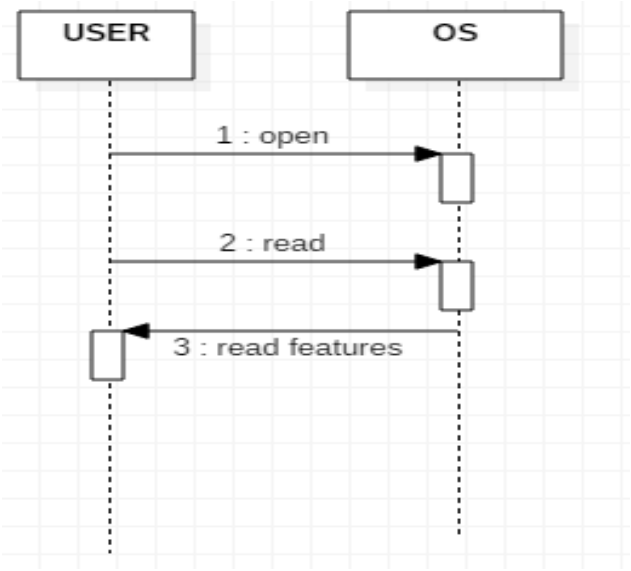
the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



SEQUENCE DIAGRAM

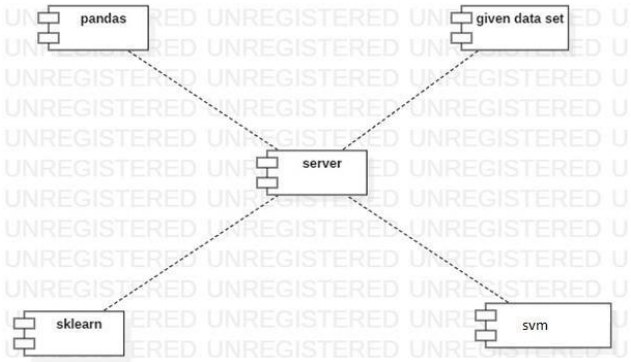
A sequence diagramin Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows, as parallel vertical lines ("lifelines"), different processes or objects that live simultaneously, and as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner.





COMPONENT DIAGRAM

The component diagram represents the high-level parts that make up the system. This diagram depicts, at a high level, what components form part of the system and how they are interrelated. A component diagram depicts the components culled after the system has undergone the development or construction phase.



SYSTEM REQUIREMENTS

SOFTWARE REQUIREMENTS

- The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.
- The appropriation of requirements and implementation constraints gives the general overview of the project in regard to what the areas of strength and deficit areand how to tackle them.

❖ Jupiter

HARDWARE REQUIREMENTS

- **HARDWARE REQUIREMENTS**
- Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

- Operating system : Windows
- Processor : intel i5
- Ram : 8 GB
- Hard disk : 512GB

MACHINE LEARNING

What is Machine Learning

Before we look at the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of building models of data.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models tunable parameters that can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain. Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here.

Categories of Machine Leaning

At the most fundamental level, machine learning can be categorized into two main types: supervised learning and unsupervised learning.

Supervised learning involves somehow modelling the relationship between measured features of data and some label associated with the data; once this model is determined, it can be used to apply labels to new, unknown data. This is further subdivided into classification tasks and regression tasks: in classification, the labels are discrete categories, while in regression, the labels are continuous quantities. We will see examples of both types of supervised learning in the following section.

Unsupervised learning involves modelling the features of a dataset without reference to any label and is often described as "letting the dataset speak for itself." These models include tasks such as clustering and dimensionality reduction. Clustering algorithms identify distinct groups of data, while dimensionality reduction algorithms search for more succinct representations of the data. We will see examples of both types of unsupervised learning in the following section.

Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate, and solve complex problems. On the other side, AI is still in its initial stage and have not surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, “to make decisions, based on data, with efficiency and scale”.

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programming logic, in the problems that cannot be programmed inherently. The fact is that we can’t do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

MODULES USED

1. TensorFlow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache

2. NumPy

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary datatypes can be defined using NumPy which allows NumPy to seamlessly and speedily

integrate with a wide variety of databases.

3. **SciKit-Learn**

- Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python.
- It is licensed under a permissive simplified BSD license and is distributed under many Linuxdistributions,encouraging academic and commercial use.

4. **Pandas**

Pandasis an open-source Python Library providinghigh-performancedata manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

5. **MatplotLib**

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object orientedinterface or via a set of functions familiar to MATLAB users.

PSUDO CODE STRUCTURE

1. Import the necessary libraries (os, numpy, pandas, sklearn)
2. Set the working directory
3. Load the "TRAIN2.csv" dataset into a Pandas dataframe
4. Extract the values of the "PotentialFraud" column and store them in a variable called "y"
5. Remove the "Provider" and "PotentialFraud" columns from the dataframe
6. Extract the remaining values in the dataframe and store them in a variable called "x"
7. Scale the values in "x" using StandardScaler
8. Split the scaled values in "x" and the values in "y" into training and testing sets
9. Train a logistic regression model using the training sets
10. Calculate and print the scores for the trained model on the training and testing sets
11. Use the model to make predictions on the testing set
12. Calculate and print the confusion matrix and classification report for the model's predictions
13. Plot the ROC curve and calculate the AUC for the model's predictions
14. Save the model for future use

FUNCTIONAL REQUIREMENTS

OUTPUT DESIGN

Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provides a permanent copy of the results for later consultation. The various types of outputs in general are:

- External Outputs, whose destination is outside the organization
- Internal Outputs whose destination is within organization and
- User's main interface with the computer.

- Operational outputs whose use is purely within the computer department.
- Interface outputs, which involve the user in communicating directly.

OUTPUT DEFINITION

The outputs should be defined in terms of the following points:

- Type of the output
- Content of the output

- Format of the output
- Location of the output
- Frequency of the output
- Volume of the output
- Sequence of the output

It is not always desirable to print or display data as it is held on a computer. It should be decided as which form of the output is the most suitable.

INPUT DESIGN

Input design is a part of overall system design. The main objective during the input design is as given below:
To produce a cost-effective method of input.
To achieve the highest possible level of accuracy.
To ensure that the input is acceptable and understood by the user.

INPUT STAGES

The main input stages can be listed as below:

- Data recording
- Data transcription
- Data conversion
- Data verification
- Data control
- Data transmission
- Data validation

INPUT TYPES

It is necessary to determine the various types of inputs. Inputs can be categorized as follows:

- External inputs, which are prime inputs for the system.
- Internal inputs, which are user communications with the system.
- Operational, which are computer department's communications to the system?
- Interactive, which are inputs entered during a dialogue.

INPUT MEDIA

At this stage choice has to be made about the input media. To conclude about the input media consideration has to be given to;

- Type of input
- Flexibility of format
- Speed
- Accuracy
- Verification methods
- Rejection rates
- Ease of correction
- Storage and handling requirements
- Security
- Easy to use
- Portability

Keeping in view the above description of the input types and input media, it can be said that most of the inputs are of the form of internal and interactive. As Input data is to be the directly keyed in by the user, the keyboard can be considered to be the most suitable input device.

ERROR AVOIDANCE

At this stage care is to be taken to ensure that input data remains accurate form the stage at which it is recorded up to the stage in which the data is accepted by the system. This can be achieved only by means of careful control each time the data is handled.

ERROR DETECTION

Even though every effort is make to avoid the occurrence of errors, still a small proportion of errors is always

likely to occur, these types of errors can be discovered by using validations to check the input data.

DATA VALIDATION

Procedures are designed to detect errors in data at a lower level of detail. Data validations have been included in the system in almost every area where there is a possibility for the user to commit errors. The system will not accept invalid data. Whenever an invalid data is keyed in, the system immediately prompts the user and the user has to again key in the data and the system will accept the data only if the data is correct. Validations have been included where necessary.

The system is designed to be a user friendly one. In other words the system has been designed to communicate effectively with the user. The system has been designed with popup menus.

USER INTERFACE DESIGN

It is essential to consult the system users and discuss their needs while designing the user interface:

USER INTERFACE SYSTEMS CAN BE BROADLY CLASIFIED AS:

- User initiated interface the user is in charge, controlling the progress of the user/computer dialogue. In the computer-initiated interface, the computer selects the next stage in the interaction.
- Computer initiated interfaces
- In the computer-initiated interfaces the computer guides the progress of the user/computer dialogue. Information is displayed and the user response of the computer takes action or displays further information.

USER INITIATED INTERGFACES

User initiated interfaces fall into two approximate classes:

- Command driven interfaces: In this type of interface the user inputs commands or queries which are interpreted by the computer.
- Forms oriented interface: The user calls up an image of the form to his/her screen and fills in the form. The forms-oriented interface is chosen because it is the best choice.

COMPUTER-INITIATED INTERFACES

The following computer – initiated interfaces were used:

- The menu system for the user is presented with a list of alternatives and the user chooses one; of alternatives.
- Questions – answer type dialog system where the computer asks question and takes action based on the basis of the users reply.

- Right from the start the system is going to be menu driven, the opening menu displays the available options. Choosing one option gives another popup menu with more options. In this way every option leads the users to data entry form where the user can key in the data.

ERROR MESSAGE DESIGN

The design of error messages is an important part of the user interface design. As user is bound to commit some errors or other while designing a system the system should be designed to be helpful by providing the user with information regarding the error he/she has committed.

This application must be able to produce output at different modules for different inputs.

III. CONCLUSION

This work aimed at developing a novel fraud detection model for insurance claims processing based on genetic support vector machines, which hybridizes and draws on the strengths support vector machines. SVMs have been considered preferable to other classification techniques due to several advantages. With other notable advantages, it has a nonlinear dividing hyper plane, which prevails over the discrimination within the dataset. The generalization ability of any newly arrived data for classification was considered over other classification techniques.

FUTURE SCOPE

The proposed methodology provides the information that Random Forest performs better than Sequential CNN. The drawback of this methodology is that anyone would expect Sequential CNN can outperform any of the conventional ML methodologies, but it is not happening here. It may happen because the

dataset is not enough to train and identify the hidden patterns to predict the future or upcoming data and the initialization of weights was very random that might affect the training process. It can be further improved in two ways. The first way is to tune the hyperparameters through optimization, and the second method is to apply the transfer learning methodology so that the performance of the proposed methodology is improved to detect the fraud transaction through Medicare in the healthcare sector.

REFERENCES

[1]. Lakshman Narayana Vejendla and A Peda Gopi, (2019)," Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology", Revue d'Intelligence Artificielle, Vol. 33, No. 1, 2019, pp.45-48.

[2]. Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), "Classification of tweets data based on polarity using improved RBF kernel of www.jespublication.com PageNo:482 SVM". Int. j. inf. tecnol. (2020)

[3]. Lakshman Narayana Vejendla and A Peda Gopi, (2017)," Visual cryptography for gray scale images with enhanced security mechanisms", Traitement du Signal, Vol.35, No.3-4, pp.197-208. DOI: 10.3166/ts.34.197-208

[4]. Herland, M., Bauder, R.A. &Khoshgoftaar, T.M. Approaches for identifying U.S. medicare fraud in provider claims data. Health Care Manag Sci 23, 2–19 (2020). https://doi.org/10.1007/s10729-018-9460-8

[5]. Hancock, J.T., Khoshgoftaar, T.M. Hyperparameter Tuning for Medicare Fraud Detection in Big Data. SN COMPUT. SCI. 3, 440 (2022). https://doi.org/10.1007/s42979-022-01348-x

[6]. [6] Bauder, R.A., Khoshgoftaar, T.M. The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data. Health Inf Sci Syst 6, 9 (2018). https://doi.org/10.1007/s13755-018-0051-3

[7]. [7] Herland, M., Khoshgoftaar, T.M. & Bauder, R.A. Big Data fraud detection using multiple medicare data sources. J Big Data 5, 29 (2018). https://doi.org/10.1186/s40537-018-0138-3

[8]. Arunkumar, C., Kalyan, S., Ravishankar, H. (2021). Fraudulent Detection in Healthcare Insurance. In: Sengodan, T., Murugappan, M., Misra, S. (eds) Advances in Electrical and Computer Technologies. ICAECT 2020. Lecture Notes in Electrical Engineering, vol 711. Springer, Singapore. https://doi.org/10.1007/978-981-15-9019-1_1

[9]. J. Chen, X. Hu, D. Yi, J. Li and M. Alazab, "A Variational AutoEncoder-Based Relational Model for Cost-Effective Automatic Medical Fraud Detection," in IEEE Transactions on Dependable and Secure Computing, 2022, doi: 10.1109/TDSC.2022.3187973.

[10]. J. Yao, S. Yu, C. Wang, T. Ke and H. Zheng, "Medicare Fraud Detection Using WTBagging Algorithm," 2021 7th International Conference on Computer and Communications (ICCC), 2021, pp. 1515-1519, doi: 10.1109/ICCC54389.2021.9674545.

[11]. Herland, M., Bauder, R.A. &Khoshgoftaar, T.M. The effects of class rarity on the evaluation of supervised healthcare fraud detection models. J Big Data 6, 21 (2019). https://doi.org/10.1186/s40537-019-0181-8

[12]. Helmut Farbmacher, Leander Löw, Martin Spindler, An explainable attention network for fraud detection in claims management, Journal of Econometrics, Volume 228, Issue 2, 2022, Pages 244-258, ISSN 0304-4076, https://doi.org/10.1016/j.jeconom.2020.05.021.

[13]. I. Matloob, S. A. Khan, R. Rukaiya, M. A. K. Khattak and A. Munir, "A Sequence Mining-Based Novel Architecture for Detecting Fraudulent Transactions in Healthcare Systems," in IEEE Access, vol. 10, pp. 48447-48463, 2022, doi: 10.1109/ACCESS.2022.3170888.

[14]. Bauder, Richard &Khoshgoftaar, Taghi. (2017). Medicare Fraud Detection Using Machine Learning Methods. 858-865. 10.1109/ICMLA.2017.00-48.

[15]. Zhang C, Xiao X, Wu C. Medical Fraud and Abuse Detection System Based on Machine Learning. Int J Environ Res Public Health. 2020 Oct 5;17(19):7265. doi: 10.3390/ijerph17197265. PMID: 33027884; PMCID: PMC7579458.

[16]. Mayaki, Mansour Zoubeirou A., and Michel Riveill. "Multiple Inputs Neural Networks for Medicare fraud Detection." arXiv preprint arXiv:2203.05842 (2022).

[17]. Y. Yoo, D. Shin, D. Han, S. Kyeong and J. Shin, "Medicare fraud detection using graph neural networks," 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), 2022, pp. 1-5, doi: 10.1109/ICECET55527.2022.9872963.

[18]. Mansour Zoubeirou a Mayaki, Michel Riveill. Multiples inputs neural nets for Medicare fraud detection. 2021. ⟨hal-03500411⟩.

[19]. J. Yao, S. Yu, C. Wang, T. Ke and H. Zheng, "Medicare Fraud Detection Using WTBagging Algorithm," 2021 7th International Conference on Computer and Communications (ICCC), 2021, pp. 1515-1519, doi: 10.1109/ICCC54389.2021.9674545.

[20]. Sailaja, C., Teja, G. S. S. K., Mahesh, G., & Reddy, P. R. S. Detection of Fraudulent Medicare Providers using Decision Tree and Logistic Regression Models.