REVOLUTIONIZING HEALTH SYSTEMS EFFICIENCY: A NOVEL SOFTWARE ENGINEERING APPROACH LEVERAGING MACHINE LEARNING

[1]Dr. O. Sampath, [2]Sanikala AnilKumar
[1]Associate Professor, [2]MCA Student
Department of Master of Computer Application,
Rajeev Gandhi Memorial College of Engineering and Technology
Nandyal, 518501, Andhra Pradesh, India.

ABSTRACT

Recently, machine learning has become a hot research topic. Therefore, this study investigates the interaction between software engineering and machine learning within the context of health systems. We proposed a novel framework for health informatics: the framework and methodology of software engineering for machine learning in health informatics (SEMLHI). The SEMLHI framework includes four modules (software, machine learning, machine learning algorithms, and health informatics data) that organize the tasks in the framework using a SEMLHI methodology, thereby enabling researchers and developers to analyze health informatics software from an engineering perspective and providing developers with a new road map for designing health applications with system functions and software implementations. Our novel approach sheds light on its features and allows users to study and analyze the user requirements and determine both the function of objects related to the system and the machine learning algorithms that must be applied to the dataset. Our dataset used in this research consists of real data and was originally collected from a hospital run by the Palestine government covering the last three years. The SEMLHI methodology includes seven phases: designing, implementing, maintaining and defining workflows; structuring information; ensuring security and privacy; performance testing and evaluation; and releasing the software applications.

INDEX TERMS: Health dataset analysis, machine learning, methodology, software development management, software engineering.

1. INTRODUCTION

The field of health informatics (HI) aims to provide a large-scale linkage among disparate ideas. Normally, a healthcare dataset is found to be incomplete and noisy; as a result, reading data from dataset linkage traditionally fails within the discipline of software engineering. Machine learning (ML) is a rapidly maturing branch of computer science since it can store data on a large scale. Many ML tools can be used to analyze data and yield knowledge that can improve the quality of work for both staff and doctors; however, for developers, there is currently no methodology that can be used. Regarding software engineering, there has been a lack of approaches to evaluating which software engineering tasks are better performed by automation and which require human involvement or human-in-the-loop approaches [1].

Big data has many challenges regarding analysis challenges for real-world big data [2], including OLAP mass data, mass data protection, mass data survey and mass data dissemination.

Recently, a set of frameworks have been used to develop data analysis tools such as Win-CASE [3] and SAM [4]. The market has vast data analysis tools that can discover interesting patterns and hidden relationships to support decision makers [5]. BKMR used the R package as a statistical approach on health effects to estimate the multivariable exposure-response function [6].

Augmentor included the Python image library for augmentation [7], while for the visualization of medical treatment plans and patient data, CareVis was used [8], as it was designed for this task. Other applications require a visual interface using COQUITO [9]. For health-care data analytics, the widely known 3P tools [10] were used. Many simple applications, such as WEKA, which provided a GUI for many machine learning algorithms [11], while Apache Spark was used for the cluster computing framework [12], are powerful systems that can used in various applications for solving problems using big data and machine learning [13]. Table 1 summarizes the main tools used for big data in analytics according with respect to the task. Software engineering for machine learning applications (SEMLA) discusses the challenges, new insights, and practical ideas regarding the engineering of ML and artificial engineering (AI) [14]. NSGA-II proposed algorithms for real-world applications that include more than one objective function for enhancing performance in terms of both diversity and convergence [15]. ML algorithms in clinical genomics generally come in three main forms: supervised, unsupervised and semi-supervised [16]. Interflow system requirement analysis (ISRA) has been used to determine the system requirements [17].

Electronic healthcare (eHealth) frameworks have replaced traditional medical frameworks to improve mobile healthcare (mHealth) and enable patient-to-physician and patient-to-patient interactions to achieve improved healthcare and quality of life (QoL) . Big data and IoT have been used for improving the efficiency of m-health systems by predicting potential life-threatening conditions during the early stages. Intelligent IoT eHealth solutions enable healthcare professionals to monitor health-related data continuously and provide real-time actionable insights used to support decision making.

Machine learning is a field of software engineering that frequently utilizes factual procedures to enable PCs to “learn” by using information from saved datasets. Unsupervised or information mining focuses more on exploratory information investigation and is known as learning supported by data analytics. Patient laboratory test queue management and wait time prediction are a challenging and complicated job. Because each patient might require different phase operations (tasks), such as a check-up, various tests, e.g., a sugar level test or blood test, X-rays or surgery, each task can consider different medical tests, from 0 to N , for each patient according to their condition.

In this article, based on a grounded theory methodology [21], the researchers proposed a novel methodology, SEMLHI, in developing a framework by defining the research problem and methodology for the developers. The SEMLHI framework includes a theoretical framework to support research and design activities that incorporate existing knowledge. The SEMLHI framework was composed of four components that help developers observe the health application flow from the main module to submodules to run and validate specific tasks. This enables multiple developers to work on different modules of the application simultaneously. The SEMLHI framework supports the methodological approach to conducting research on health informatics. It also supports a structure that presents a common set of ML terminology to use, compare, measure, and design software systems in the area of health. This creates a space whereby SE and ML experts can work on a specific methodological approach to enable health informatics software development teams to integrate the ML model lifecycle. Our methodology was applicable to current systems or in the development of new systems that use the ML module for current systems, which can be used in regular updates to add data to the system, to perform irregular updates and to add new features such as new versions of ICD diagnosis codes, minor model improvements for bug fixes, new functionalities required by the client, and new hardware or architectural constraints.

2. LITERATURE SURVEY

Interactive machine learning: Experimental evidence for the human in the algorithmic loop

Recent advances in automatic machine learning (aML) allow solving problems without any human intervention. However, sometimes a human-in-the-loop can be beneficial in solving computationally hard problems. In this paper we provide new experimental insights on how we can improve computational intelligence by complementing it with human intelligence in an interactive machine learning approach (iML). For this purpose, we used the Ant Colony Optimization (ACO) framework, because this fosters multi-agent approaches with human agents in the loop. We propose unification between the human intelligence and interaction skills and the computational power of an artificial system. The ACO framework is used on a case study solving the Traveling Salesman Problem, because of its many practical implications, e.g. in the medical domain. We used ACO due to the fact that it is one of the best algorithms used in many applied intelligence problems. For the evaluation we used gamification, i.e. we implemented a snake-like game called Traveling Snakesman with the MAX–MIN Ant System (MMAS) in the background. We extended the MMAS–Algorithm in a way, that the human can directly interact and influence the ants. This is done by “traveling” with the snake across the graph. Each time the human travels over an ant, the current pheromone value of the edge is multiplied by 5. This manipulation has an impact on the ant’s behavior (the probability that this edge is taken by the ant increases). The results show that the humans performing one tour through the graphs have a significant impact on the shortest path found by the MMAS. Consequently, our experiment demonstrates that in our case human intelligence can positively influence machine intelligence. To the best of our knowledge this is the first study of this kind.

Big data challenges and achievements: Applications on smart cities and energy sector

In this paper, the Big Data challenges and the processing is analyzed, recently great attention has been paid to the challenges for great data, largely due to the wide spread of applications and systems used in real life, such as presentation, modeling, processing and large (often unlimited) data storage. Mass Data Survey, OLAP Mass Data, Mass Data Dissemination and Mass Data Protection. Consequently, we focus on further research trends and, as a default, we will explore a future research challenge research project in this area of research.

CASE: A framework for computer supported outbreak detection

In computer supported outbreak detection, a statistical method is applied to a collection of cases to detect any excess cases for a particular disease. Whether a detected aberration is a true outbreak is decided by a human expert. We present a technical framework designed and implemented at the Swedish Institute for Infectious Disease Control for computer supported outbreak detection, where a database of case reports for a large number of infectious diseases can be processed using one or more statistical methods selected by the user. Based on case information, such as diagnosis and date, different statistical algorithms for detecting outbreaks can be applied, both on the disease level and the subtype level. The parameter settings for the algorithms can be configured independently for different diagnoses using the provided graphical interface. Input generators and output parsers are also provided for all supported algorithms. If an outbreak signal is detected, an email notification is sent to the persons listed as receivers for that particular disease. The framework is available as open source software, licensed under GNU General Public License Version 3. By making the code open source, we wish to encourage others to contribute to the future development of computer supported outbreak detection systems, and in particular to the development of the CASE framework.

3. EXISTING SYSTEM:

Normally, a healthcare dataset is found to be incomplete and noisy; as a result, reading data from dataset linkage traditionally fails within the discipline of software engineering. Machine learning (ML) is a rapidly maturing branch of computer science since it can store data on a large scale. Many ML tools can be used to analyze data and yield knowledge that can improve the quality of work for both staff and doctors; however, for developers, there is currently no methodology that can be used. Regarding software engineering, there has been a lack of approaches to evaluating which software engineering tasks are better performed by automation and which require human involvement or human-in-the-loop approaches.

DISADVANTAGES OF EXISTING SYSTEM:

- ❖ Regarding software engineering, there has been a lack of approaches to evaluating which software engineering tasks are better performed by automation and which require human involvement or human-in-the-loop approaches.

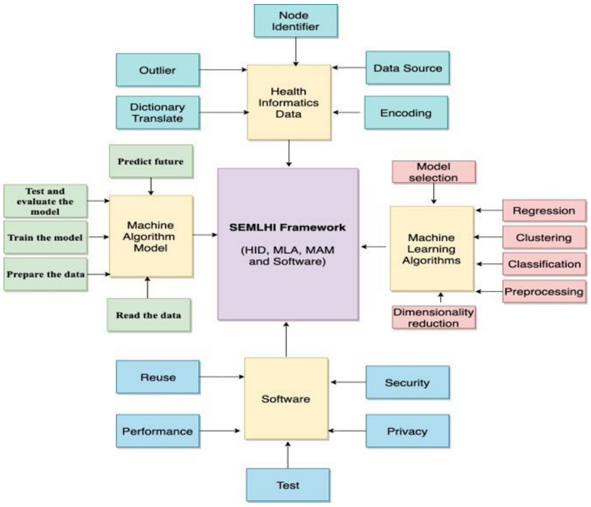
4. PROPOSED SYSTEM:

In propose paper author is combining Software Engineering and Machine Learning algorithms to improve disease prediction in health care systems and to minimize time taken to predict disease as we don't have enough hospitals or bed to accommodate growing number of patients and we can solve this problem of predicting disease with less time by employing software and machine learning algorithms. Propose paper concept is known as SEMLHI (where SE refers to software and ML refers to machine learning and HI refers to health data).

ADVANTAGES OF PROPOSED SYSTEM:

- Advance machine learning algorithm called EXTREME LEARNING MACHINE (EML) and this EML algorithm is giving better prediction result compare to propose paper algorithms.

SYSTEM ARCHITECTURE



5. MODULES DESCRIPTION

Health Informatics Data:

To predict any disease we need to build Machine Learning models by using datasets and this datasets often contains missing data, null and non-numeric data and this type of data could degrade ML prediction accuracy and to overcome from this problem author is applying PREPROCESSING on health care data to remove all missing and null values and then convert non-numeric data to numeric data by applying python SKLEARN PREPROCESSING classes. Often this dataset may contains unnecessary columns or attributes and to remove this attributes author applying dimensionality reduction algorithm called PCA. PCA (principal component analysis) remove unnecessary attributes from dataset and maintain only important attributes necessary to make correct prediction.

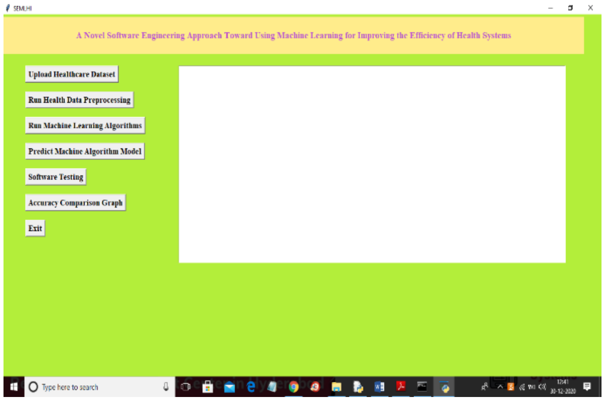
ML Algorithms: In this module author using various machine learning algorithms such as Linear SVC, Multinomial Naïve Bayes, Random Forest, Logistic Regression and KNN. This algorithm train itself with available datasets and then generate a train model and then this train model will be applied on new test data to perform prediction. By using above algorithms we can make machine to learn and perform prediction without any human supports.

Machine Algorithm Model: Once after building above models then we can apply new test data on this model to predict whether patient lab reports are positive or negative.

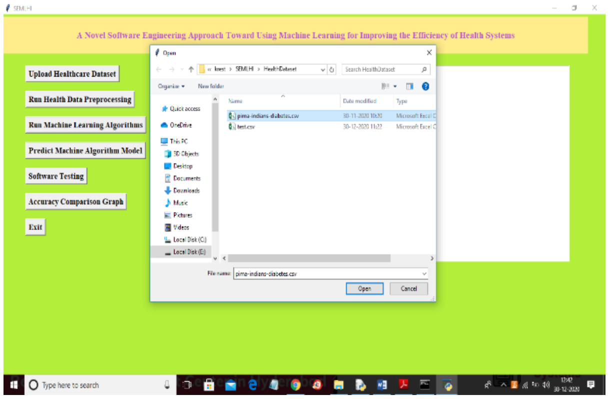
Software: This module used by developers to check reliability of above modules by applying software quality check, UNIT TESTING and software verification.

6. SCREEN SHOTS

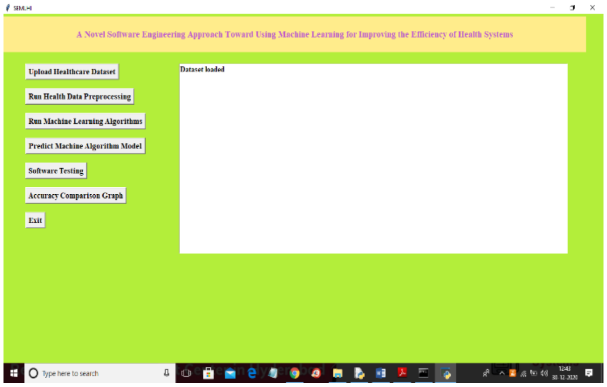
To run project double click on run.bat file to get below screen shots


SEMLHI
A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems
Upload Healthcare Dataset
Run Health Data Preprocessing
Run Machine Learning Algorithms
Predict Machine Algorithm Model
Software Testing
Accuracy Comparison Graph
Exit
Type here to search

In above screen click on 'Upload Healthcare Dataset' button to upload dataset

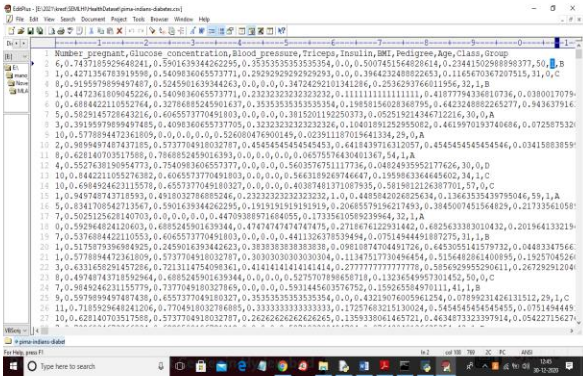

SEMLHI
A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems
Upload Healthcare Dataset
Run Health Data Preprocessing
Run Machine Learning Algorithms
Predict Machine Algorithm Model
Software Testing
Accuracy Comparison Graph
Exit
Open
HealthDataset
pima-indians-diabetes.csv
test.csv
File name: pima-indians-diabetes.csv
Open
Cancel
Type here to search

In above screen selecting and uploading diabetes dataset and then click on 'Open' button to load dataset and to get below screen


SEMLHI
A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems
Upload Healthcare Dataset
Run Health Data Preprocessing
Run Machine Learning Algorithms
Predict Machine Algorithm Model
Software Testing
Accuracy Comparison Graph
Exit
Type here to search

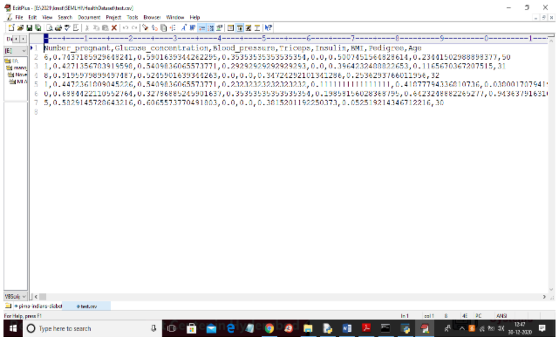
Dataset loaded

In above screen dataset loaded and in below screen of dataset we can see there is last label 'Class' which contains values as 0 and 1 where 0 means that lab values contains no disease and 1 means that lab values contains disease


EditPlus - [E:\2021\kreet\SEMLHI\HealthDataset\pima-indians-diabetes.csv]
File Edit View Search Document Project Tools Browser Window Help
pima-indians-diabet
For Help, press F1
Type here to search

```
Number_pregnant,Glucose_concentration,Blood_pressure,Triceps,Insulin,BMI,Pedigree,Age,Class,Group
6,0.7437185929648241,0.5901639344262295,0.35353535353535354,0.0,0.5007451564828614,0.23441502988898377,50,1,B
1,0.4271356783919598,0.5409836065573771,0.29292929292929293,0.0,0.3964232488822653,0.1165670367207515,31,0,C
8,0.9195979899497487,0.5245901639344263,0.0,0.0,0.34724292101341286,0.2536293766011956,32,1,B
1,0.4472361809045226,0.5409836065573771,0.23232323232323232,0.1111111111111111,0.41877794336810736,0.0380017079
0,0.6884422110552764,0.32786885245901637,0.35353535353535354,0.19858156028368795,0.6423248882265277,0.943637916
5,0.5829145728643216,0.6065573770491803,0.0,0.0,0.3815201192250373,0.05251921434671216,30,0,A
3,0.3919597989949745,0.4098360655737705,0.32323232323232326,0.10401891252955082,0.4619970193740686,0.072587532
10,0.5778894472361809,0.0,0.0,0.0,0.526080476900149,0.02391118701964134,29,0,A
2,0.9899497487437185,0.5737704918032787,0.45454545454545453,0.6418439716312057,0.454545454545454546,0.03415883859
8,0.628140703517588,0.7868852459016393,0.0,0.0,0.0,0.0657557643040137,54,1,A
4,0.5527638190954773,0.7540983606557377,0.0,0.0,0.5603576751117736,0.04824935952177626,30,0,D
10,0.8442211055276382,0.6065573770491803,0.0,0.0,0.5663189269746647,0.1959863364645602,34,1,C
10,0.6984924623115578,0.6557377049180327,0.0,0.0,0.40387481371087935,0.5819812126387701,57,0,C
1,0.949748743718593,0.4918032786885246,0.23232323232323232,1.0,0.4485842026825634,0.1366353543979504,59,1,A
5,0.8341708542713567,0.5901639344262295,0.19191919191919191,0.20685579196217493,0.3845007451564829,0.21733561058
7,0.5025125628140703,0.0,0.0,0.0,0.44709388971684055,0.1733561058923964,32,1,A
0,0.592964824120603,0.6885245901639344,0.47474747474747475,0.2718676122931442,0.6825633383010432,0.201964133219
7,0.5376884422110553,0.6065573770491803,0.0,0.0,0.4411326378539494,0.07514944491887275,31,1,B
1,0.5175879396984925,0.2459016393442623,0.38383838383838383,0.09810874704491726,0.6453055141579732,0.04483347566
1,0.5778894472361809,0.5737704918032787,0.30303030303030304,0.11347517730496454,0.5156482861400895,0.192570452
3,0.6331658291457286,0.7213114754098361,0.41414141414141414,0.2777777777777778,0.5856929955290611,0.267292912040
8,0.49748743718592964,0.6885245901639344,0.0,0.0,0.5275707898658718,0.13236549957301452,50,0,C
7,0.9849246231155779,0.7377049180327869,0.0,0.0,0.5931445603576752,0.15926558497011,41,1,B
9,0.5979899497487438,0.6557377049180327,0.35353535353535354,0.0,0.4321907600596125,0.07899231426131512,29,1,C
11,0.7185929648241206,0.7704918032786885,0.3333333333333333,0.17257683215130024,0.5454545454545455,0.0751494449
10,0.628140703517588,0.5737704918032787,0.26262626262626265,0.1359338061465721,0.4634873323397914,0.054227156270
```

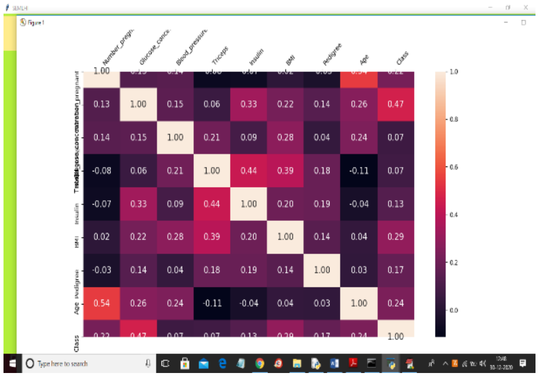
In above dataset screen all values are the lab report values and 'Class' value contains 0 or 1 and ML algorithm will train with above lab report values and Class Value and then generate a model. Generated train model we will apply on below test data to predict class label. In below test dataset we can see there is no Class label column and ML will predict Class label by using alone lab values. See below test values



EditPlus - [E:\2021\Jenst\SEMLHI\HealthDataset\test.csv]
File Edit View Search Document Project Tools Browser Window Help

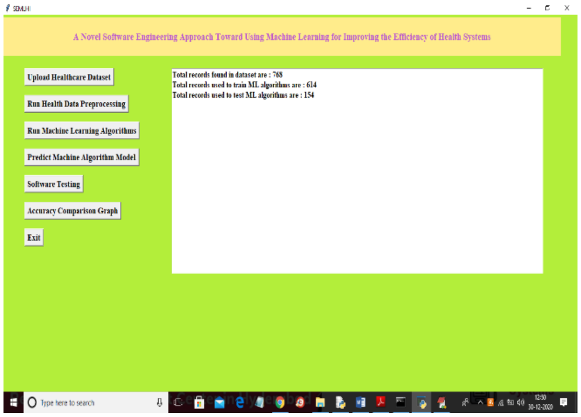
```
Number_pregnant,Glucose_concentration,Blood_pressure,Triceps,Insulin,BMI,Pedigree,Age
6,0.7437185929648241,0.5901639344262295,0.3535353535353535354,0.0,0.5007451564828614,0.23441502988898377,50
1,0.4271356783919598,0.5409836065573771,0.29292929292929293,0.0,0.3964232488822653,0.11656703672075115,31
8,0.9195979899497487,0.5245901639344263,0.0,0.0,0.34724292101341286,0.25362937660119556,32
1,0.4472361809045226,0.5409836065573771,0.23232323232323232,0.11111111111111111,0.41877794336810736,0.038001707941
0,0.6884422110552764,0.32786885245901637,0.35353535353535354,0.19858156028368795,0.6423248882265277,0.943637916310
5,0.5829145728643216,0.6065573770491803,0.0,0.0,0.38152011922503731,0.05251921434671221,30
```

In above test lab report values there is no Class label. Now go back to output screen and then click on ‘Run Health Data Preprocessing’ button to remove missing values and then apply PCA dimensionality algorithm to get below graph



	Number_pregn...	Glucose_conce...	Blood_pressure	Triceps	Insulin	BMI	Pedigree	Age	Class
Number_pregnant	1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Glucose_concentration	0.13	1.00	0.15	0.06	0.33	0.22	0.14	0.26	0.47
Blood_pressure	0.14	0.15	1.00	0.21	0.09	0.28	0.04	0.24	0.07
Triceps	-0.08	0.06	0.21	1.00	0.44	0.39	0.18	-0.11	0.07
Insulin	-0.07	0.33	0.09	0.44	1.00	0.20	0.19	-0.04	0.13
BMI	0.02	0.22	0.28	0.39	0.20	1.00	0.14	0.04	0.29
Pedigree	-0.03	0.14	0.04	0.18	0.19	0.14	1.00	0.03	0.17
Age	0.54	0.26	0.24	-0.11	-0.04	0.04	0.03	1.00	0.24
Class	0.22	0.47	0.07	0.07	0.13	0.29	0.17	0.24	1.00

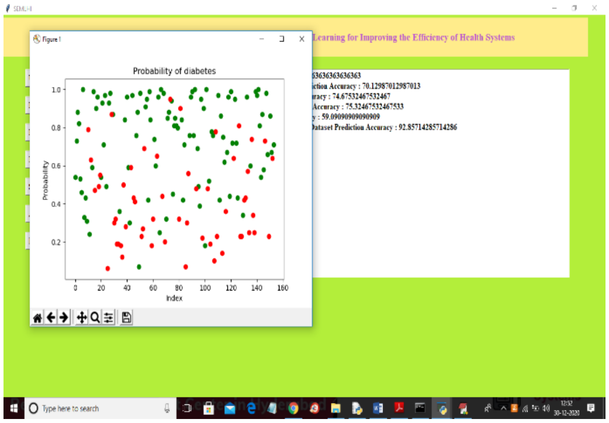
In above graph in top we can see names of columns and in boxes values with minus symbols are not important and only positive column values are important and ML algorithm will train only with positive values and now close above graph to get below screen



SEMLHI
A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems
Upload Healthcare Dataset
Run Health Data Preprocessing
Run Machine Learning Algorithms
Predict Machine Algorithm Model
Software Testing
Accuracy Comparison Graph
Exit
Type here to search

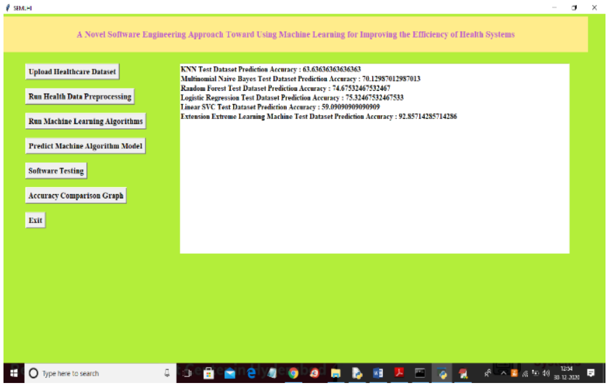
```
Total records found in dataset are : 768
Total records used to train ML algorithms are : 614
Total records used to test ML algorithms are : 154
```

In above screen after applying pre-processing and PCA we got total records as 768 and application using 614 records to train ML algorithms and to generate model and then used 154 records to test that trained model and to calculate prediction accuracy. Now both train and test data is ready and now click on ‘Run Machine Learning Algorithms’ button to start training all ML algorithms on train and test data



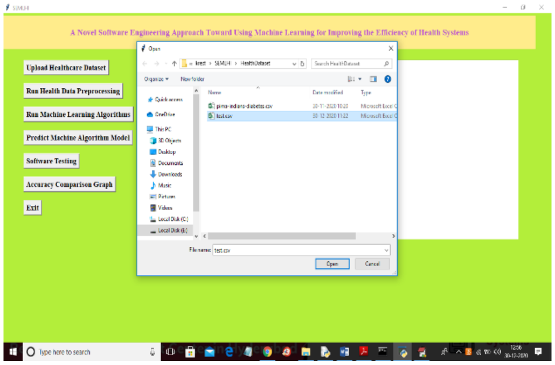
```
63636363636363
iction Accuracy : 70.12987012987013
racy : 74.67532467532467
Accuracy : 75.32467532467533
y : 59.09090909090909
Dataset Prediction Accuracy : 92.85714285714286
```

In above graph green colour dots are the records which contains no disease and red colour dots are the records which contains disease and this graph generated for all 154 test records. Now close above graph to see all ML prediction accuracy

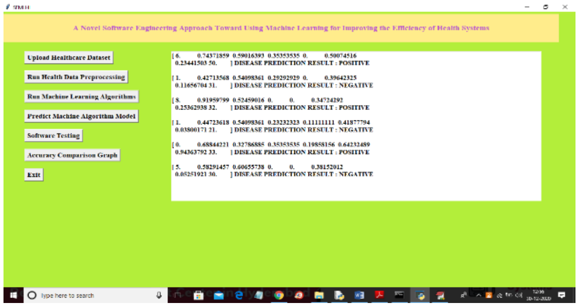

SEMLHI
A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems
Upload Healthcare Dataset
Run Health Data Preprocessing
Run Machine Learning Algorithms
Predict Machine Algorithm Model
Software Testing
Accuracy Comparison Graph
Exit
Type here to search

```
KNN Test Dataset Prediction Accuracy : 63.63636363636363
Multinomial Naive Bayes Test Dataset Prediction Accuracy : 70.12987012987013
Random Forest Test Dataset Prediction Accuracy : 74.67532467532467
Logistic Regression Test Dataset Prediction Accuracy : 75.32467532467533
Linear SVC Test Dataset Prediction Accuracy : 59.09090909090909
Extension Extreme Learning Machine Test Dataset Prediction Accuracy : 92.85714285714286
```

In above screen we can see prediction accuracy of each algorithm and from all algorithms extension Extreme Machine Learning is giving good prediction accuracy and now all ML algorithms are ready with trained model and now click on 'Predict Machine Algorithm Model' button to upload new test records and then ML will predict whether new test records contains positive or negative disease

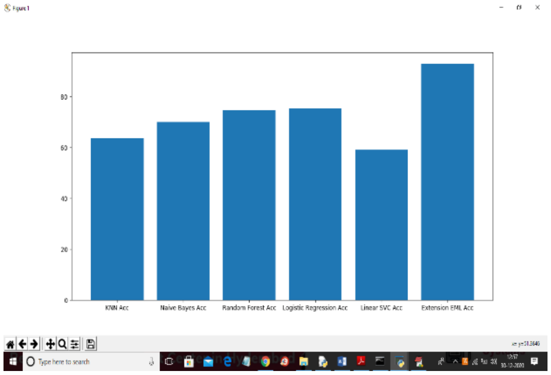

SEMLHI
A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems
Upload Healthcare Dataset
Run Health Data Preprocessing
Run Machine Learning Algorithms
Predict Machine Algorithm Model
Software Testing
Accuracy Comparison Graph
Exit
Open
krest > SEMLHI > HealthDataset
Search HealthDataset
Organize
New folder
Quick access
OneDrive
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C:)
Local Disk (E:)
Name
Date modified
Type
pima-indians-diabetes.csv
test.csv
Microsoft Excel C
File name: test.csv
Open
Cancel
Type here to search

In above screen selecting and uploading 'test.csv' file and then click on 'Open' button to load test data and to get below prediction result

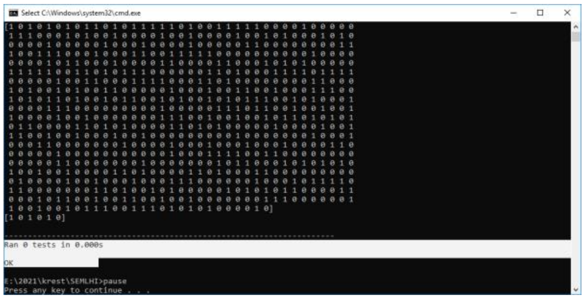

SEMLHI
A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems
Upload Healthcare Dataset
Run Health Data Preprocessing
Run Machine Learning Algorithms
Predict Machine Algorithm Model
Software Testing
Accuracy Comparison Graph
Exit
Type here to search

```
[6.         0.74371859  0.59016393  0.35353535  0.          0.50074516
 0.23441503 50.        ] DISEASE PREDICTION RESULT : POSITIVE

[1.         0.42713568  0.54098361  0.29292929  0.          0.39642325
 0.11656704 31.        ] DISEASE PREDICTION RESULT : NEGATIVE

[8.         0.91959799  0.52459016  0.          0.          0.34724292
 0.25362938 32.        ] DISEASE PREDICTION RESULT : POSITIVE

[1.         0.44723618  0.54098361  0.23232323  0.11111111  0.41877794
 0.03800171 21.        ] DISEASE PREDICTION RESULT : NEGATIVE

[0.         0.68844221  0.32786885  0.35353535  0.19858156  0.64232489
 0.94363792 33.        ] DISEASE PREDICTION RESULT : POSITIVE

[5.         0.58291457  0.60655738  0.          0.          0.38152012
 0.05251921 30.        ] DISEASE PREDICTION RESULT : NEGATIVE
```

In above screen for each test lab record ML predict whether disease is positive or negative. Now click on 'Accuracy Comparison Graph' button to get below graph



In above graph x-axis represents ML algorithm names and y-axis represents accuracy of all those algorithms and from above graph we can conclude that extension EML is giving better accuracy and now click on 'Software Testing' to check all ML are working properly and to get below screen



```
[1 0 1 0 1 0 1 0 1 1 0 1 0 1 1 1 1 1 0 1 0 0 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0
 1 1 1 0 0 0 1 0 1 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 1 0 1 0
 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1
 1 0 0 1 1 1 0 0 0 1 0 0 0 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
 0 0 0 0 1 0 1 1 0 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 1 0 1 0 1 0 0 0 0 0
 1 1 1 1 1 0 0 1 1 0 1 0 1 1 1 0 0 0 0 0 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 1
 0 0 0 0 0 1 0 0 1 1 0 0 0 1 1 1 1 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0
 1 0 1 0 0 1 0 1 0 0 1 1 0 0 0 0 0 1 0 0 0 1 0 0 1 1 0 0 1 0 0 0 1 1 1 0 0
 1 0 1 0 1 1 0 1 0 0 1 0 1 1 0 0 1 0 1 0 0 1 0 1 0 1 1 1 0 0 1 0 1 0 0 0 1
 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1
 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 1 1 0 0 1 0 0 1 0 0 1 0 1 1 0 1 0 1 0 1
 0 1 1 0 0 0 0 1 1 0 1 0 1 0 0 0 0 1 1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1
 1 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1
 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 1 1 0
 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 1 0 0 0 1 0 1 0 1 0 1 0
 1 0 0 1 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0 1 1 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0
 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 1 0 1 1 1 1 0
 1 1 0 0 0 0 0 0 0 1 1 0 1 0 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 1 1 0 0 0 0 1 1
 0 0 0 1 0 1 1 0 0 1 0 0 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1
 1 0 0 1 0 0 1 0 1 1 1 0 0 1 1 1 0 1 0 1 0 1 0 0 0 0 1 0]
[1 0 1 0 1 0]

----------------------------------------------------------------------
Ran 0 tests in 0.000s

OK

E:\2021\krest\SEMLHI>pause
Press any key to continue . . .
```

In above screen if all ML algorithms are working properly then we will get tests result as OK

7. CONCLUSION

This article addressed an important HI with ML topic in software engineering by proposing an efficient new method approach related to software engineering, identified in prior research studies, using original data sets collected during the last 3 years from a Palestine hospital. This methodology allows developers to analyze and develop software for the HI model and create a space in which software engineering and ML experts can work together on the ML model life-cycle, especially in the health field. This manuscript proposed a framework that included a theoretical framework composed of four modules (software, ML model, ML algorithms, and HI data). The new methodology was compared between three system engineering methods: Vee, Agile and SEMLHI. The results showed the delivery of the new methodology for one-shot delivery. For the MAM component on the SEMLHI framework, laboratory test results were obtained using five algorithms to test the accuracy of the ICD-10 results using equations and to evaluate the accuracy of the ML models with a sample size of 750 patients. The results for MAM showed that the SVG was approximately 0.57.

Future Enhancement:

Further We implement other ML algorithm for train data and increase prediction accuracy of each algorithm .

REFERENCES

[1] A. Holzinger, ''Interactive machine learning: Experimental evidence for the human in the algorithmic loop,'' Appl. Intell., vol. 49, no. 7, pp. 2401–2414, 2019.

[2] T. A. Mohammed, A. Ghareeb, H. Al-Bayaty, and S. Aljawarneh, ''Big data challenges and achievements: Applications on smart cities and energy sector,'' in Proc. 2nd Int. Conf. Data Sci., E-Learn. Inf. Syst., 2019, p. 26.

[3] B. Cakici, K. Hebing, M. Grünewald, P. Saretok, and A. Hulth, ''CASE: A framework for computer supported outbreak detection,'' BMC Med. Inform. Decis. Making, vol. 10, no. 1, p. 14, 2010.

[4] A. J. Vickers, T. Salz, E. Basch, M. R. Cooperberg, P. R. Carroll, F. Tighe, and J. Eastham, and R. C. Rosen, ''Electronic patient self-assessment and management (SAM): A novel framework for cancer survivorship,'' BMC Med. Inform. Decis. Making, vol. 10, no. 1, p. 34, 2010.

[5] A. Ismail, A. Shehab, and I. M. El-Henawy, ''Healthcare analysis in smart big data analytics: Reviews, challenges and recommendations,'' in Security in Smart Cities: Models, Applications, and Challenges, vol. 9, A. E. Hassanien, M. Elhoseny, S. H. Ahmed, and A. K. Singh, Eds. Cham, Switzerland: Springer, Nov. 2019, pp. 27–45.

[6] J. F. Bobb, B. C. Henn, L. Valeri, and B. A. Coull, ''Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression,'' Environ. Health, vol. 17, no. 1, p. 67, 2018.

[7] B. Aribisala and O. Olabanjo, ''Medical image processor and repository– MIPAR,'' Inform. Med. Unlocked, vol. 12, pp. 75–80, Jul. 2018.

[8] W. Aigner and S. Miksch, ‘‘CareVis: Integrated visualization of computerized protocols and temporal patient data,’’ Artif. Intell.in Med., vol. 37, no. 3, pp. 203–218, Jul. 2006.

[9] J. Krause, A. Perer, and H. Stavropoulos, ‘‘Supporting iterative cohort construction with visual temporal queries,’’ IEEE Trans. Vis. Comput. Graph., vol. 22, no. 1, pp. 91–100, Jan. 2016.

[10] R. K. Pathinarupothi, P. Durga, and E. S. Rangan, ‘‘Data to diagnosis in global health: A 3P approach,’’ BMC Med. Inform. Decis. Making, vol. 18, no. 1, pp. 1–13, 2018.

[11] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques. Amsterdam, The Netherlands: Elsevier, 2016, pp. 438–441.

[12] Q.-C. To, J. Soto, and V. Markl, ‘‘A survey of state management in big data processing systems,’’ VLDB J., vol. 27, no. 6, pp. 847–872, Dec. 2018. [13] S. R. Salkuti, ‘‘A survey of big data and machine learning,’’ Int. J. Elect. Comput. Eng., to be published. Accessed: Jan. 7, 2020. [Online]. Available: http://ijece.iaescore.com/index.php/IJECE/article/view/19184/pdf

[14] F. Khomh, B. Adams, J. Cheng, M. Fokaefs, and G. Antoniol, ‘‘Software engineering for machine-learning applications: The road ahead,’’ IEEE Softw., vol. 35, no. 5, pp. 81–84, Sep. 2018.

[15] T. A. Mohammed, Y. I. Hamodi, and N. T. Yousir, ‘‘Intelligent enhancement of organization work flow and work scheduling using machine learning approach tree algorithm,’’ Int. J. Comput. Sci. Netw. Secur., vol. 18, no. 6, pp. 87–90, 2018.

[16] J. A. Diao, I. S. Kohane, and A. K. Manrai, ‘‘Biomedical informatics and machine learning for clinical genomics,’’ Hum. Mol. Genet., vol. 27, no. R1, pp. R29–R34, May 2018.

[17] P.-H. Cheng, Y.-P. Chen, and J.-S. Lai, ‘‘An interflow system requirement analysis in health informatics field,’’ in Proc. WRI World Congr. Comput. Sci. Inf. Eng., vol. 1, 2009, pp. 712–716.

[18] C. George, P. Duquenoy, and D. Whitehouse, ‘‘eHealth: Legal, ethical and governance challenges,’’ in eHealth: Legal, Ethical and Governance Challenges, C. George, D. Whitehouse, and P. Duquenoy, Eds. Berlin, Germany: Springer, 2014, pp. 1–398.

[19] K. N. Mishra and C. Chakraborty, ‘‘A novel approach towards using big data and IoT for improving the efficiency of m-health systems,’’ in Advanced Computational Intelligence Techniques for Virtual Reality in Healthcare, vol. 875. Cham, Switzerland: Springer, 2020, pp. 123–139.

[20] B. Farahani, M. Barzegari, F. Shams Aliee, and K. A. Shaik, ‘‘Towards collaborative intelligent IoT eHealth: From device to fog, and cloud,’’ Microprocessors Microsyst., vol. 72, Feb. 2020, Art. no. 102938.