**Abstract:-**

This report provides insights of WEKA, an open source data Mining software program programme built in Java that can be discovered [at www.cs.waikato.ac.nz/ml/WEKA/.](at%20www.cs.waikato.ac.nz/ml/weka/.%20) The workshop's purpose is to illustrate WEKA's fundamental talents, which can be utilized in undergraduate laptop era and engineering guides. WEKA gives a robust set of machine mastering algorithms for data Mining jobs, as well as a whole kind of gear for records pre-processing, analytics, and visualisation, all on hand through a simple graphical user interface. WEKA is a famous instructional software program for data mining. utilized several supervised machine learning algorithms in building a model to analyse and predict the presence of Breast cancer and Diabetes, using the dataset from Diabetes. J48 Decision Tree, IBK, PART, Zero R and Naïve Bayes algorithms were applied through WEKA machine learning software. Each model’s performance was evaluated using 10-fold cross validation and compared according to major accuracy measures, correctly or incorrectly classified instances.

Table of Contents

[Introduction 3](#_Toc105683383)

[Aim 3](#_Toc105683384)

[Objectives 3](#_Toc105683385)

[Methodology of tableau 4](#_Toc105683386)

[Methodology of WEKA 9](#_Toc105683387)

[Main body of report: 17](#_Toc105683388)

[Results and Analysis 18](#_Toc105683389)

[WEKA 24](#_Toc105683391)

[Conclusions and Future Work 25](#_Toc105683392)

[Reference list: 26](#_Toc105683393)

**Introduction of WEKA.:-**

The WEKA is a flightless bird which has an inquisitive nature which is found on the island of New Zealand. In this case of data collection method WEKA is collection of machine learning algorithms which is for tasks in data mining. It is actually used for classification, regression and association mining and visualization.

The WEKA tool gaining knowledge of Data processing which is a today's highly popular platform for computerized type, regression, clustering, and characteristic selection, which is probably common data statistics mining obligations. It includes a huge style of system getting to know algorithms and facts pre-processing techniques, in addition, it also includes the graphical character interfaces for data exploration and experimental assessment of several system learning techniques at the same topic. WEKA can paintings with information that is organized in an unmatched relational table. Its essential goals are to,

(a) help customers extract important facts from information and

(b) make it easy for them to choose the correct method for developing an powerful predictive model from it.

# Main Body:-

Early detection is vital for stopping the unfold of a ailment that would endanger human lives. COVID-19, a contagious disease that has mutated into more than one sorts, has advanced into a worldwide epidemic that requires immediate diagnosis. With the advancement of technology, the quantity of records to be had on COVID-19 grows each day, and facts mining can be used to extract applicable facts from huge quantities of statistics.

WEKA is a Java-based totally set of gadgets mastering algorithms created on the university of Waikato in New Zealand. WEKA's most important feature is to do facts-mining duties, and it changed into first utilised in schools as a getting to know resource. The product is now blanketed in the Pentaho business intelligence suite, which makes use of WEKA for enterprise intelligence. it can be used for a spread of purposes.  
  
-guidelines of the affiliation  
-deciding on an characteristic  
-Clustering  
-Pre-processing of records  
-classification of information  
-Visualization of information  
-evaluation of Regression  
-examine the workflow

**Features of WEKA:-**

1. Preprocessors:-

Preprocessor - A preprocessor is a software program application programmed that cleans up noisy records. It can undertake any additional evaluation if the statistics is noisy. Having said that, It can be that determine what issues are found in statistics; ABT (analyze Base table) in the information may be considered from an excel report, and information records may be found the use of Excel clear out menus. records cleansing, facts integration, records discount, and facts transformation are the four degrees of statistics preprocessing.

1. Classifiers: -

Classifiers- Following the techniques, users could also use the data once it has been preprocessed to obtain some knowledge. a deliberate grouping or categorization of items based on predetermined criteria. There are multiple types of classification techniques.

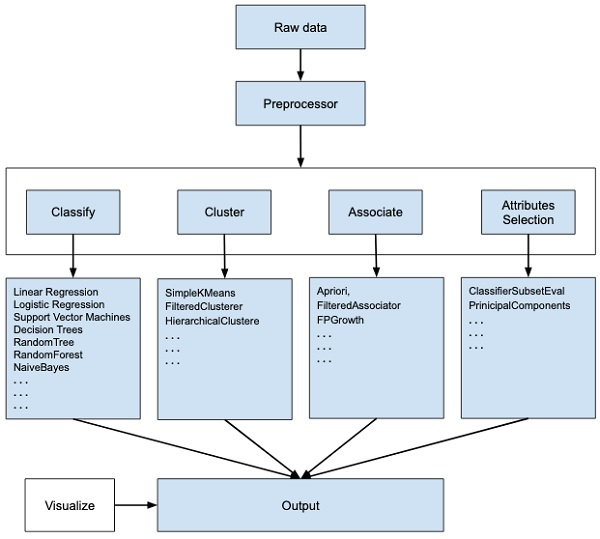


Figure 1 Pre-processor’s classifications.

Source- Mystudylab.com

# Literature review: -

Cardiovascular ailment is a blanket word that consists of a massive range of coronary heart-associated health troubles. The pathological troubles of the coronary heart, further to extraordinary factors of it, are expressly described via such scientific troubles. coronary heart disorder is a serious health problem. The quantity of males and females suffering from coronary heart ailment has risen over time . several research on the therapy of coronary heart illness had been completed. various diagnostic device getting to know algorithms were used, yielding distinctive chances. several research are being carried out to assess the inefficiency of the ok-famous character, J48, SMO, Nave Bayes, MLP, Random forest, Bayes internet, and REPTREE algorithms. simplest in rare times does Bayes net beat random forests, steady with the checks so far. (Srivastava, 2014)

The mathematical processes of Nave Bayes elegance are executed. The possibility precept of the Bayes precept is used to calculate the participation elegance. The Nave Bayes algorithm is based mostly on implicit independence. It refers back to the independence of feature values for a given beauty from one in all a kind function values. For the primary schooling database, the posterior possibility of the response parameter is calculated. the opposite parameters are similarly subjected to structured opportunity computations.

selection timber (J48): The middle of the logistic regression is just like that of a tree, with all inner nodes serving as an exam at the traits and all roots reflecting the available test effects. a couple of function desire metrics beneficial resource in deciding on the feature for information beauty branch. In data mining, various versions of desire timber are utilised. The author has estimated the accuracy of j48 is better then Naive bays.

**Case studies: -**

Health data scoring analysis using WEKA:-

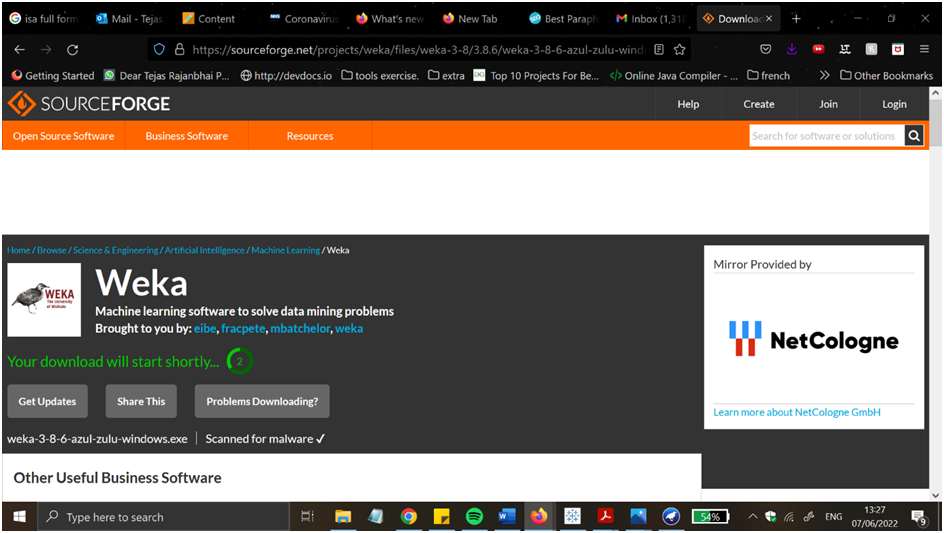
As consistent with past studies, the Government healthcare departments has visible numerous useful improvements inside the twenty-first century. The health organisation location makes use of present day-day techniques which includes facts mining to maintain up with new traits and traits in digital technology. the usage of WEKA software for credit score scoring changed into tested on this work the use of a case observe of a German health credit score dataset. consistent with the hypothetical results, there are a greater percent of patient classified as "fit," or those who pay their health tracking rating on time. clients with a bigger balance on their cutting-edge account and who've worked for a prolonged span of time are certainly higher risk, regular with the findings.

# 

# Methodology of WEKA :-

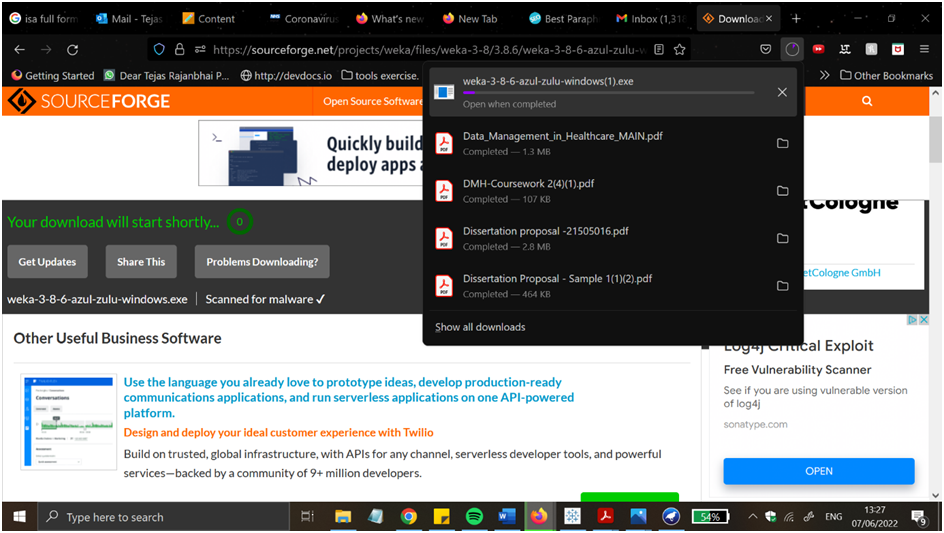
Dataset: - The datasets selected is related with diabetes patients, The dataset contains the closure of numbers of patients who are currently diabetes positive, with inclusive analysis the patients status on diabetes can be easily predicted by data mining.

In this chant here figure out the data driven of the diabetes patients in the WEKA software process. The chart is excised with serial number, naiveBays cross, J48 cross, and ZeroR cross and INK cross and at the last PART cross.



**Figure2: Downloading wizard from the WEKA**

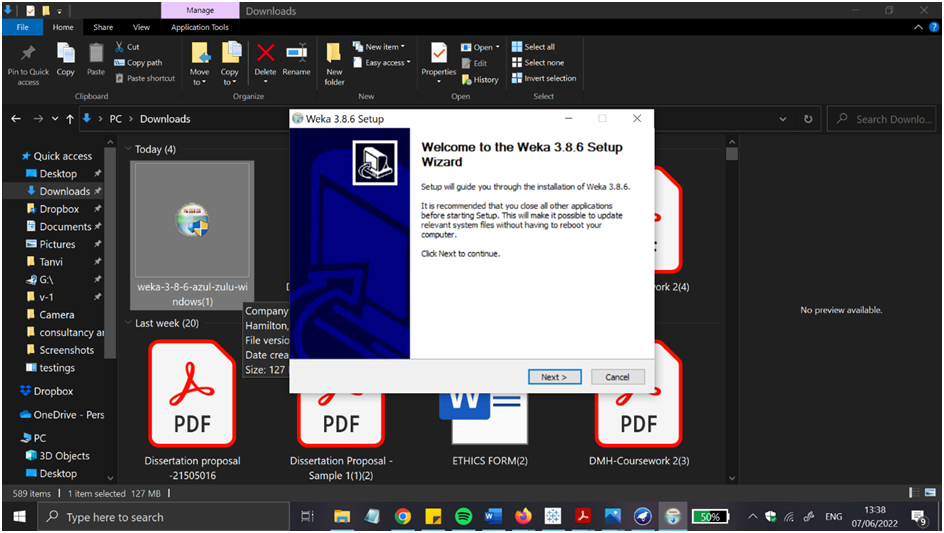
(Source: WEKA)

The first part of using the software WEKA must to do that part, which is downloading the wizard from the WEKA. In this part users know about the whole process of the installation of the software and then go to download it to use the software in their system. Finding the wizard and correctly downloading the software is a very important process (Saleh *et al* 2020) . 

**Figure3: Installing wizard of WEKA**

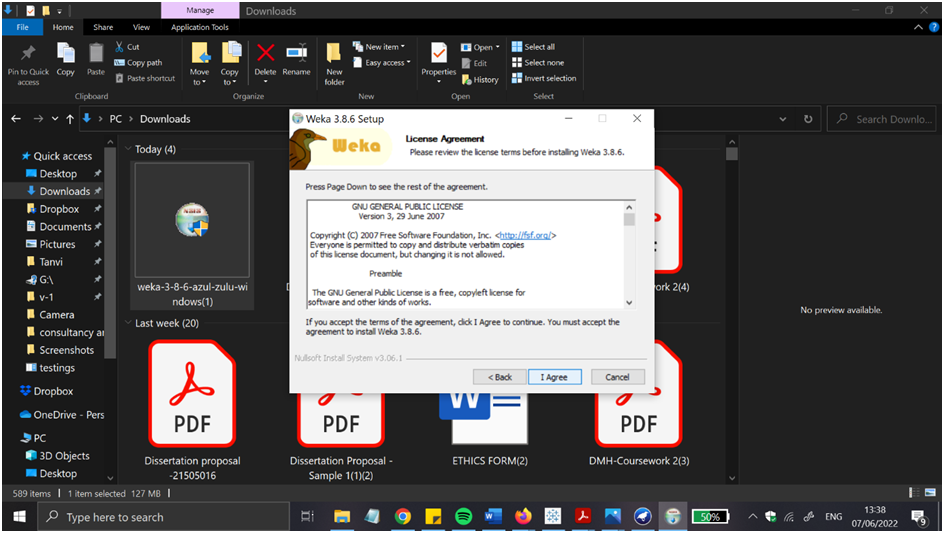
(Source: WEKA)

Here after downloading the software from the WEKA then completing the downloading process users go to the installing process. Here in this case of installation users are able to click the install option and easily install the WEKA in their system (Vasundhara, S., 2021).



**Figure4: Set up guide of wizard**

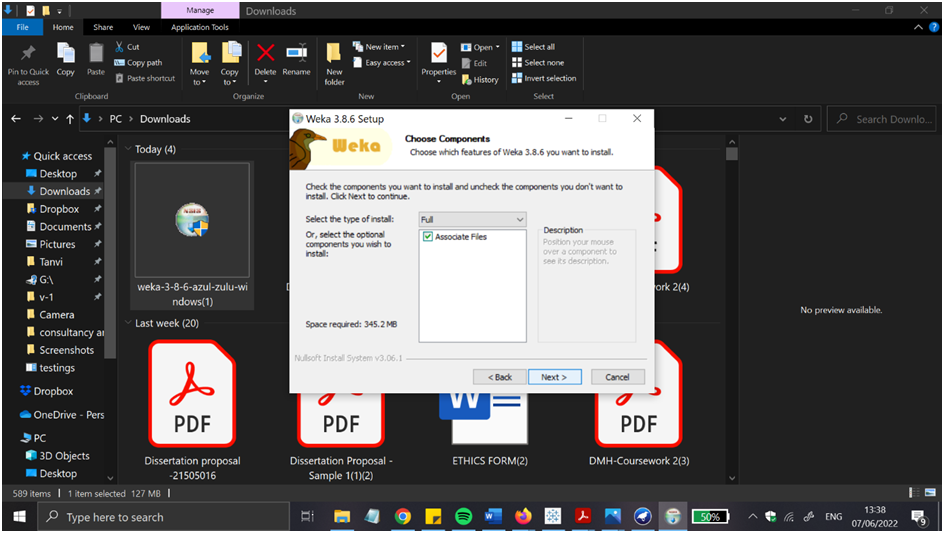
(Source: WEKA)

In this case normally the setup guide of the wizard is seen (Sarangam Kodati, D.R.V., 2018). In this process basically told the setup guidance and told about the download process of the software and also told about the software terms and conditions to the users. 

**Figure5: License agreement**

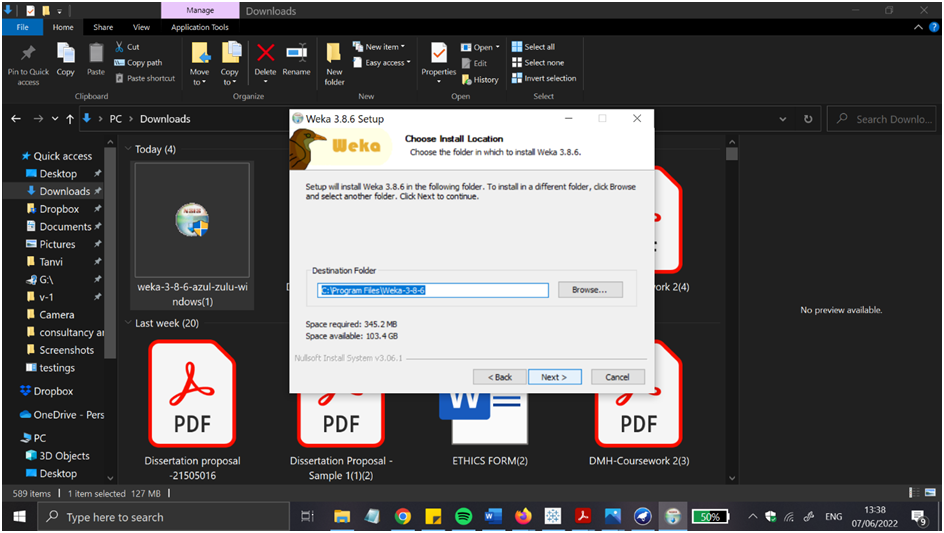
(Source: WEKA)

In this picture after knowing all the terms and conditions the used are go to the I agree button and agree with all the information and step up to get the main facing page of the WEKA software.



**Figure6: component section**

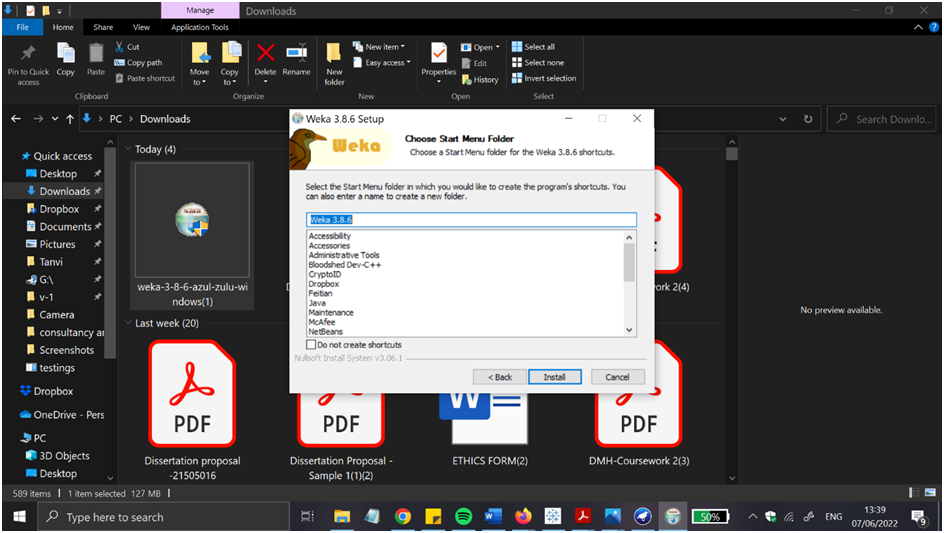
Source: WEKA

Here the users can see the components of the software and how to use the components to run the app and get the best result. 

**Figure7: set up of Path section**

(Source: WEKA)

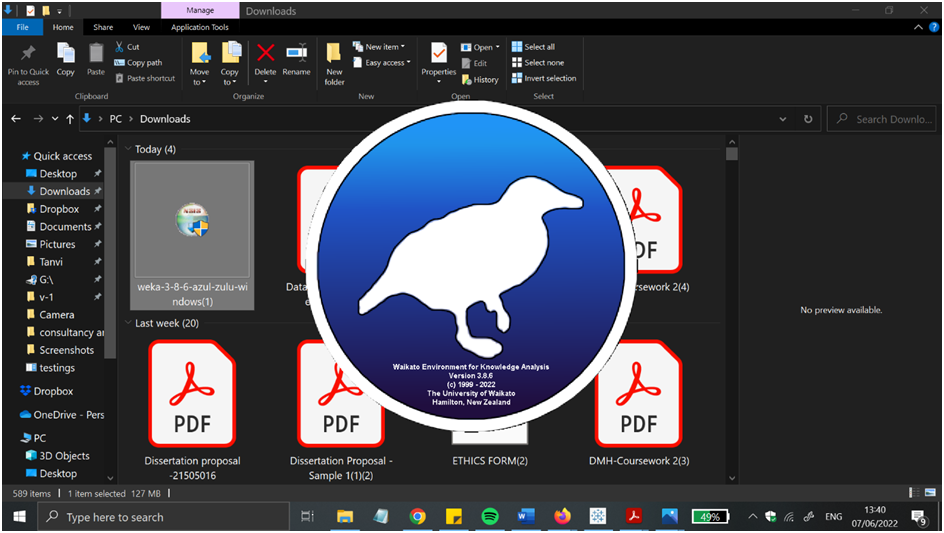
This picture indicates the set up of the path selection of the software WEKA and gets the result while using the software.

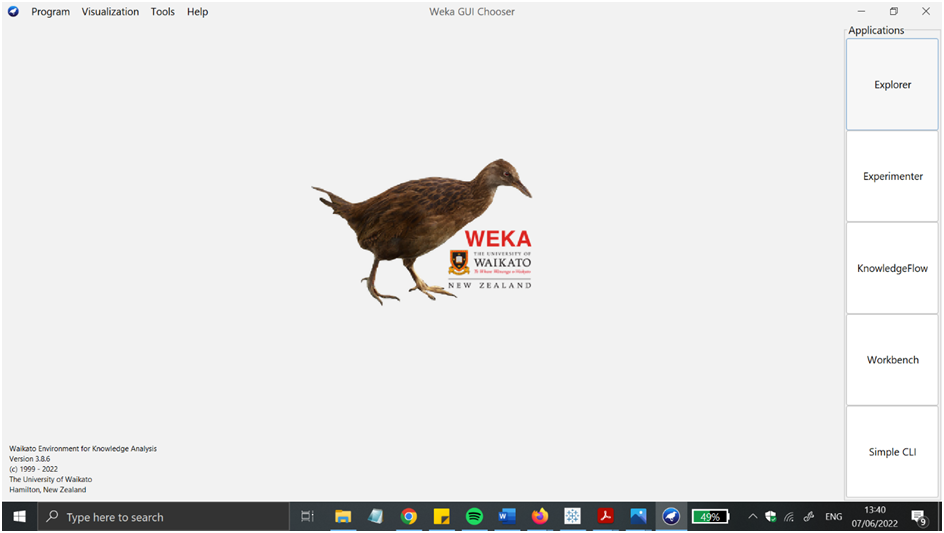


**Figure8: Starting of the menu folder**

(Source: WEKA)

After getting the path selection then the main menu folder is open and users can easily choose the option that the users want in the main menu of the software.

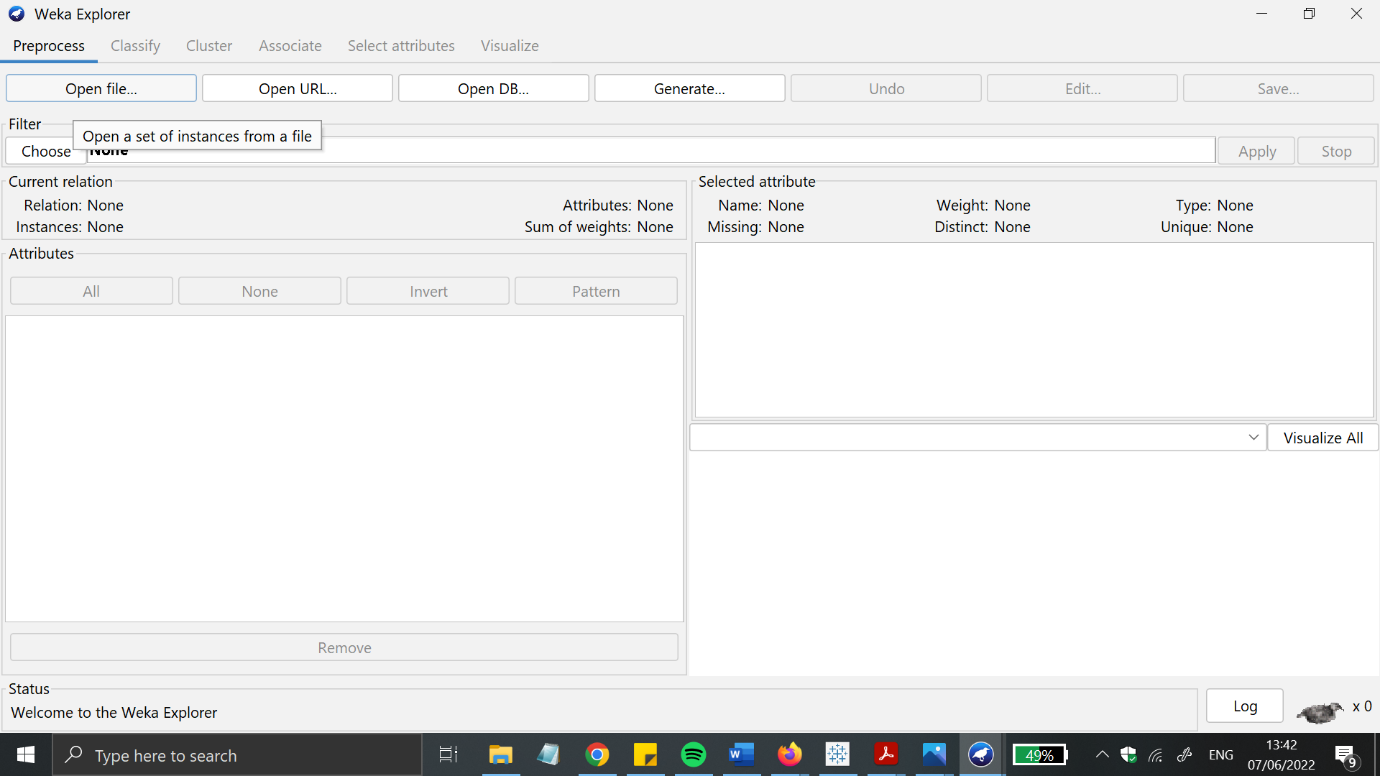
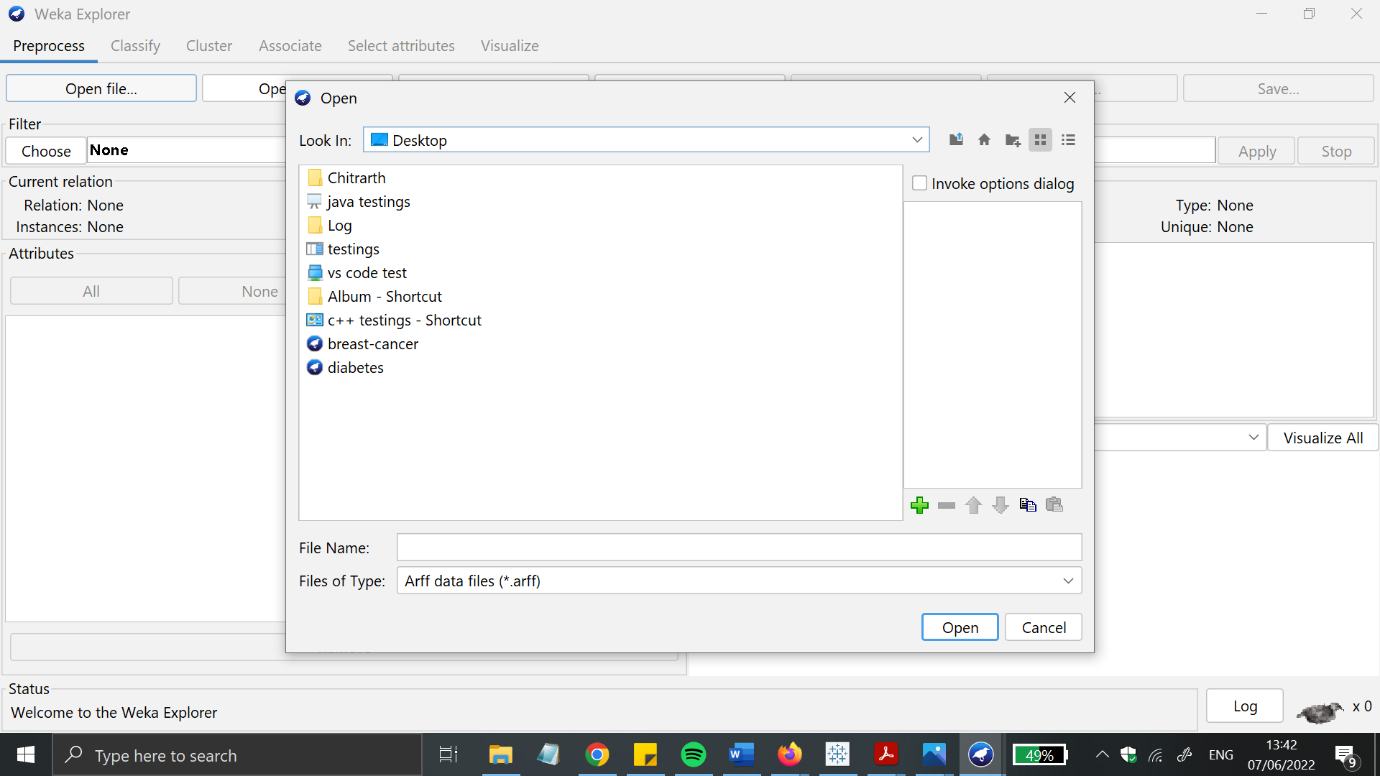
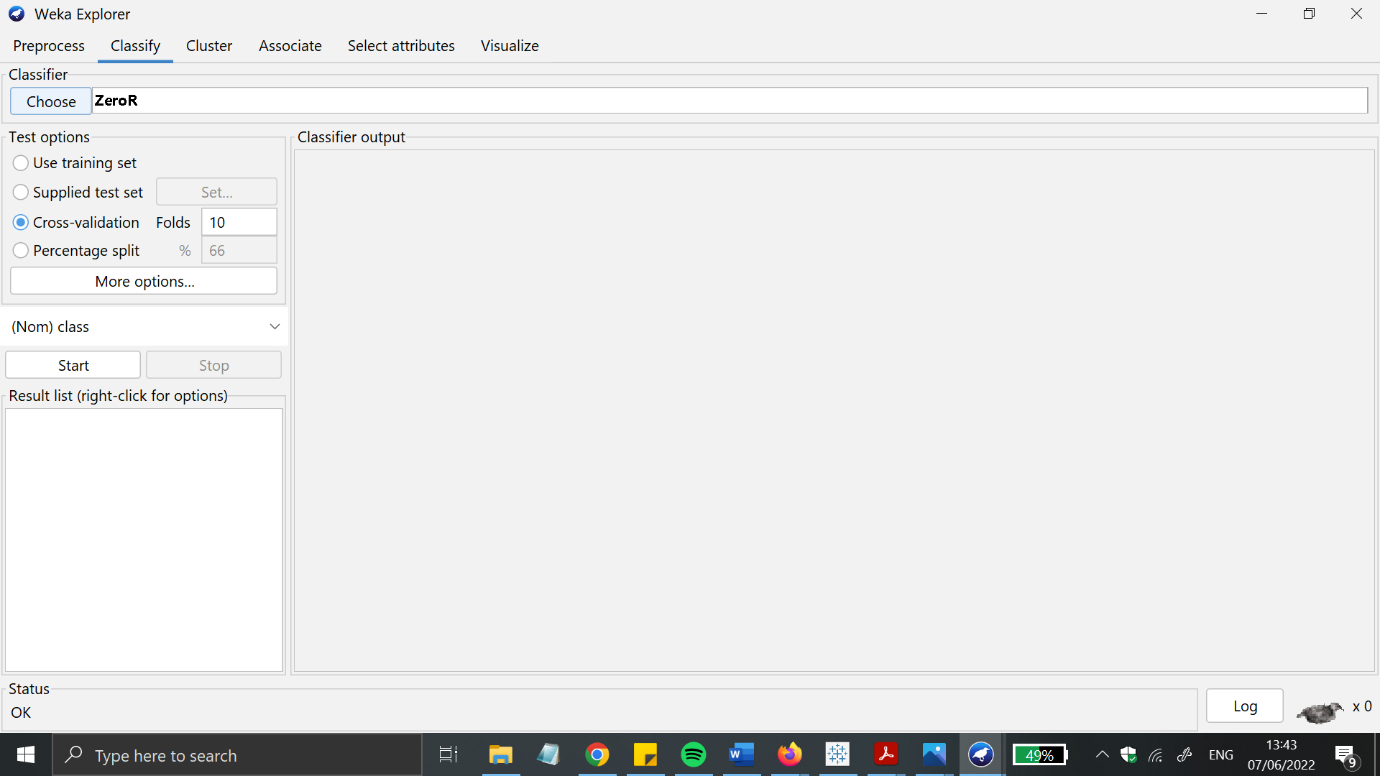
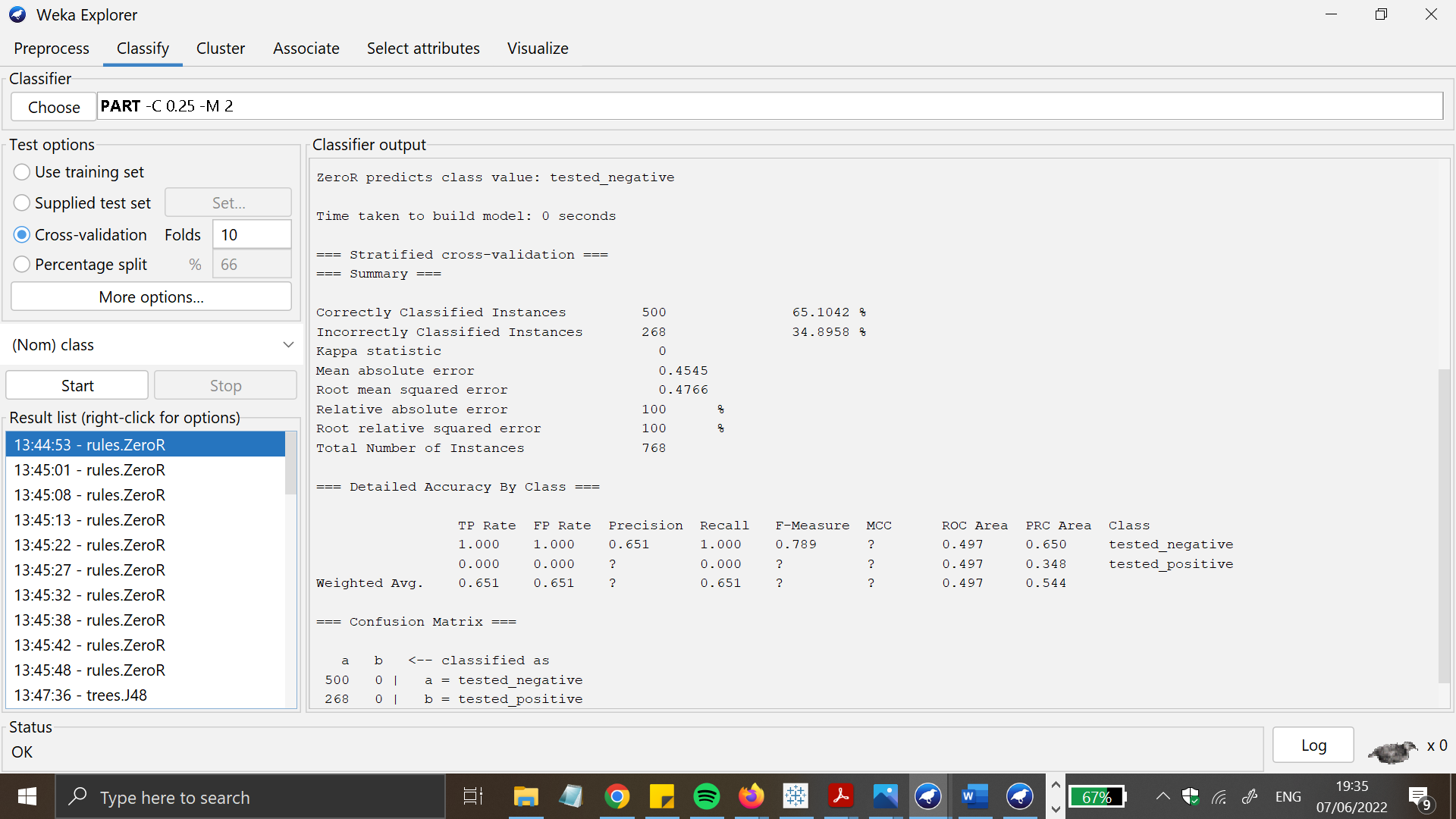




**Figure9: The starting phase of WEKA**

(Source: WEKA)

At the end of the software the main phase of the software comes out and users can easily use the software in the healthcare propose.

Figure 10 WEKA console page.Figure 11 importing datasets from sourceFigure 12 analysis performanceFigure 13 Analysis results from 10 cross validation 5 classifiers

**Results and analysis: -**

whilst encoder filters are as compared with regards to time it takes to assemble a format and the variety of accurate examples diagnosed, it is clean that Supervised studying promises tremendous performance. Here, in this below table I have mentioned the results from 10 cross validation results from 5 classifiers.

## WEKA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr number** | **NaiveBayes, cross validation 10 folds** | **J48, cross validation 10 folds** | **ZeroR cross validation, cross validation 10 folds** | **IBK cross validation, 10 folds** | **PART cross validation 10 folds** |
| **1)** | **70.2797%** | **72.3776%** | **70.2797%** | **69.9301%** | **65.035%** |
| **2)** | **70.979%** | **72.3776%** | **70.2797%** | **74.1259%** | **66.433%** |
| **3)** | **69.9301%** | **71.3287%** | **70.2797%** | **69.5804%** | **72.3776%** |
| **4)** | **74.1259%** | **72.7273%** | **70.2797%** | **73.0769%** | **70.979%** |
| **5)** | **73.7762%** | **71.6783%** | **70.2797%** | **74.1259%** | **69.9301%** |
| **6)** | **73.0769%** | **73.7762%** | **70.2797%** | **72.3776%** | **72.028%** |
| **7)** | **74.1259%** | **72.7273%** | **70.2797%** | **72.3776%** | **70.2797%** |
| **8)** | **75.5245%** | **72.0283%** | **70.2797%** | **72.3775%** | **71.3287%** |
| **9)** | **75.5245%** | **71.6783%** | **70.2797%** | **72.3773%** | **70.979%** |
| **10)** | **73.0769%** | **72.7273%** | **70.2797%** | **73.4266%** | **69.2308%** |

**Figure17: chart of diabetes patients in WEKA software.**

Source: WEKA

Here in this chart displays the patient ratio which is made from the WEKA software (Firdaus *et al* 2018).

The WEKA is running on the java programming so it can run any system and to work on the large data set and helps automating in the work so in the industry purpose using this the company get more profit in their production and achieve their success in the industrial field.

**Future works and conclusion:-**

operating with the WEKA tool is tested intensive in this paper. we are able to undergo information mining thoughts first, and then the numerous ranges of information mining in Weka. the steps involved in acting severa records mining standards are outlined. WEKA's API is extraordinarily useful because it may be used to tweak any set of policies for higher effects.

**References: -**

Journals

Billur, E.C.E.R. and AKTAŞ, A., 2019. Clustering of European Countries in terms of Healthcare Indicators. *International Journal of Computational and Experimental Science and Engineering (IJCESEN)*, *5*(1), pp.23-26.

Ratra, R. and Gulia, P., 2020. Experimental evaluation of open source data mining tools (WEKA and Orange). *Int. J. Eng. Trends Technol*, *68*(8), pp.30-35.

Seun, E., Adekunle, Y.A., Omotosho, O.J., Adebayo, A.O. and Omolara, T., 2019. Data Privacy Preserving Model for Health Information System. *International Journal of Engineering Research and Technology*, *12*(6), pp.745-752.

saleh, B. jumae (2018). APPLYING DATA MINING TECHNIQUES ON ACADAMIC INSTITUTIONAL SYSTEM USING WEKA. *International Journal of Recent Trends in Engineering and Research*, pp.86–88.

Priangga, P. and Suryani, E. (2019). Decision Analysis of Giving Credit Using Pairwise Comparisons and Scoring Methods (Case Study: Bank XYZ). *IPTEK The Journal for Technology and Science*, 30

Srivastava, S. (2014). Weka: A Tool for Data preprocessing, Classification, Ensemble, Clustering and Association Rule Mining. *International Journal of Computer Applications*, 88(10), pp.26–29. doi:10.5120/15389-3809.

‌

‌

‌