Evaluation of Cost benefit Analysis using One-R Supervised Machine Learning Algorithm for Healthcare

Dr.Kavitha S[1] ,Dr. Prasad N.H.[2], Dr.Chandra Kanta Samal[3],Dr. Hanumanthappa M[4]
Assistant Professor, Department of Masters of Computer Applications[1]
Prof & HOD, Department of Masters of Computer Applications[2]
Associate Professor, Department of Computer Science, Acharya Narendra Dev College, Delhi University, Delhi[3]
Professor, Department of Masters of Computer Applications, Bangalore University, Bangalore[4]
[1,2]Nitte Meenakshi Institute of Technology Bengaluru, India
kavitha.s@nmit.ac.in[1] ,hod-mca@nmit.ac.in[2],chandrakantsamal@du.ac.in[3],hanu6572@gmail.com[4]

Abstract:

Pandemic was present for the entire world from 2019 to 20. Due to this reason the workload for doctors and other healthcare professionals were increased. This workload will be eased by machine learning and the development of computer-aided analytical systems. The goal of the proposed methodology is towards the prevalence of COVID-19 to cost/benefit predictions on real-life dataset. Our proposed methodology is given for weka classification for the accuracy measurement ratios by applying 1R machine learning classifiers Considering the development of clustering with positive and negative occurrences ratios in terms of cost-benefit analysis's initial care projections. In this study 1R Supervised Machine Learning Algorithm have been applied to Covid 19 dataset provided by healthcare organization. The best classification accuracy is obtained from the algorithm of 1R with 75.54%. In this paper visualization Cost/Benefit Analysis and also analysed.

Keywords: *Visualization,Classification, healthcare, Machine Learning, cost data analysis; decision making; predictions*

DOI Number: 10.14704/nq.2022.20.9.NQ44773 Neuro Quantology 2022; 20(9):6610-6615

1. INTRODUCTION

As per the study, the virus-infected individuals will experience a mild to severe respiratory disease. By keeping a distance of at least one metre between people, wearing a mask and often washing your hands or using an alcohol-based rub, you can prevent infection in both yourself and other people [10]. Wuhanhas been the site of the initial discovery of a novel coronavirus known as 2019-nCoV. There is evidence that the virus can spread from person to person. Some individuals with this disease, which affects the respiratory system, can eventually recover without the need for special care, especially if they have a robust immune system [10]. It may be difficult to provide the proper care for patients who are suspected of having COVID-19 because of the limited access to COVID-19 diagnostic kits. As a result, it's crucial to have an autonomous prediction system that checks for COVID-19 in a person. The required tools for creating a COVID-19 prediction model are machine learning classification methods, datasets, and machine learning software. There are three types of machine learning: reinforcement learning, unsupervised learning, and supervised learning [11].

2. LITERATURE SURVEY

The researcher Scandizzo, P.L presented a methodology to integrate cost–benefit analysis and SAM-CGE-based impact evaluation[1].In this paper researcher highlighted the evolution of techniques from simple features/learners to more complex learners and feature engineering or sampling techniques [2].Xiong X et., all used hybrid attribute discrimination method for psychometric data with fuzziness for the classification and achieved 77.5% accuracy[3] . Lakshmi K et al., applied the methodology and achieved 78.2% accuracy[4]. Patil BM et. al., applied hybrid prediction model to the diabetic

patients for KNN classification algorithm and achieved 94.2% accuracy [5]. Chen P et. al proposed a model for diabetes and applied Logistic method to get 85.35% accuracy [6]. Shah A et al. carried out research of adversarial images by deep learning based retinal image analysis algorithms[7]. Noman Sohail et. al.,discussed the methodology of classification and cost benefit Analysis of Diabetes mellitus Dominance and achieved 99.6441 accuracy[8].

3. DATA SET DESCRIPTION

For the current study,the dataset used is obtained from the Kaggle site's data set named worldometer_data which is provided by WHO. In the figure Fig. 1 visualizing the features and descriptions of the dataset COVID-19. Totally, this dataset has 209 instances and 16 attributes.

4. TOOL DESCRIPTION

a. 1R classifier algorithm

Class for creating and utilising a 1-R classifier, which discretizes numerical properties and predicts using the minimum-error attribute. There are various options in the 1R algorithm such as numDecimalPlaces is used to find the number of decimal places to be used for the model's output of numbers is indicated by the parameter numDecimalPlaces. The next otion in 1R method batchSize[12] to find the ideal number of cases to process when using batch prediction. It is possible to offer more or less instances, and this allows implementations to choose their preferred batch size. One more option is Debug will return the output if true, the classifier may emit extra information to the console. The next option DoNotCheckCapabilities if set, does not verify classifier capabilities before building classifier.The minimum bucket size for discretizing numerical attributes is called minBucketSiz[9].

5. RESULTS AND DISCUSSIONS

Figure 2 to Figure 7 presents the cost/benefit analysis for COVID-19 dataset. Classification accuracy for five continents is show in the figure 2 to Figure 7 with cost and benefit analysis(Fig 2-7).

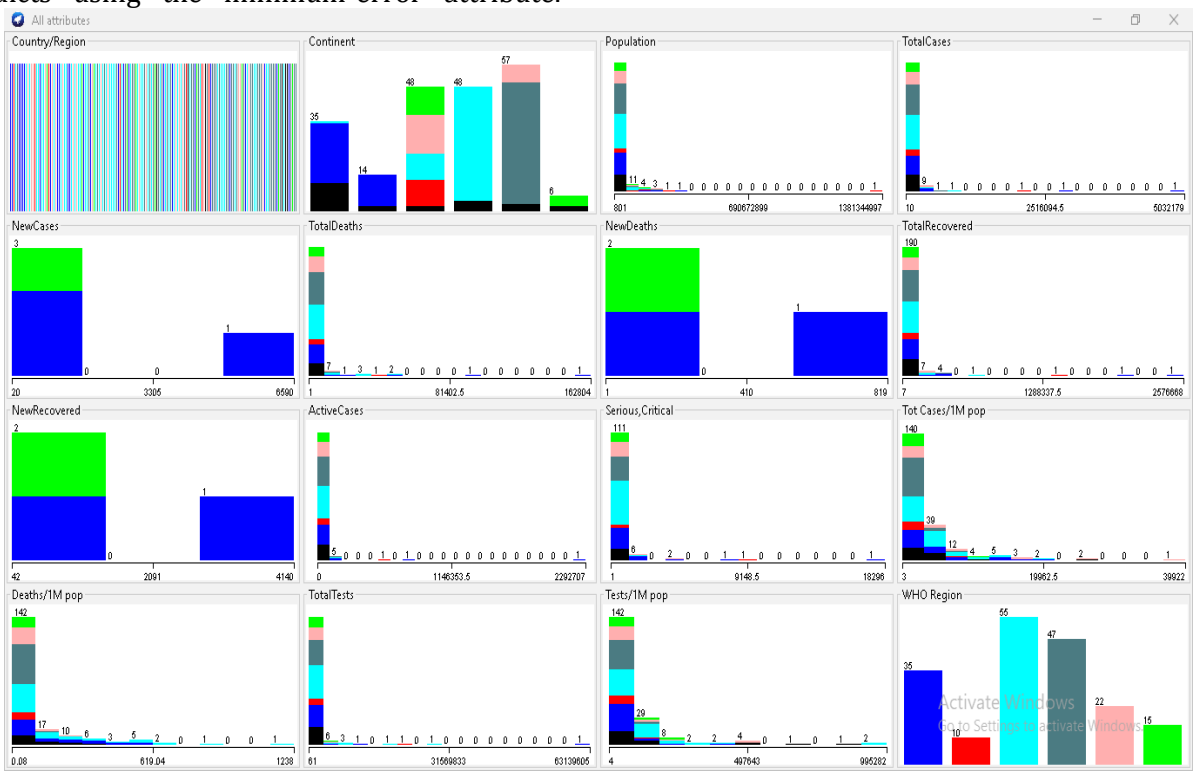


Fig. 1 Visualizing all the attributes

a. Cost and benefit analysis

Figure 2 shows the cost/benefit analysis for COVID-19 dataset forAmerica. Classification accuracy is 83..25% and threshold score is 1 in the cost and benefit analysis(Fig 2).

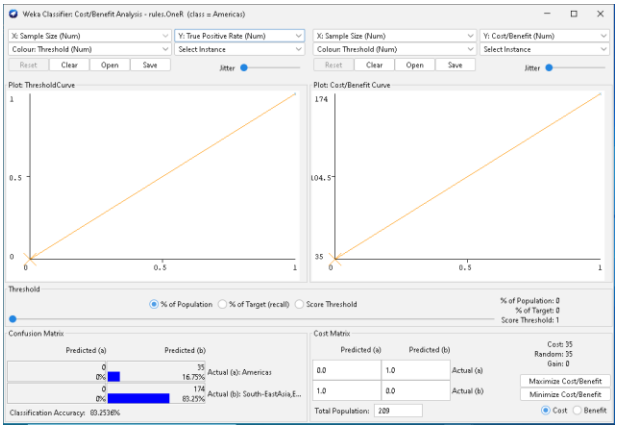


Fig. 2 Cost/Benefit analysis of America

Figure 3 shows the cost and benefit analysis for COVID-19 dataset for Southeast Asia. Classification accuracy is 95.215% , threshold score is 0, cost and random value is 10 in the cost and benefit analysis(Fig 3).

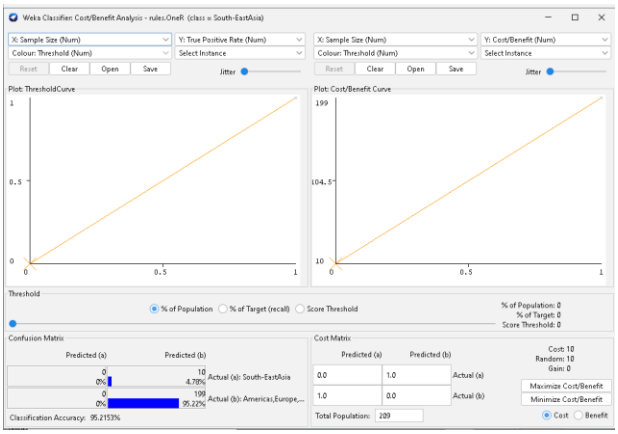


Fig. 3 Cost/Benefit analysis of SouthEastAsia

Figure 4 shows the cost and benefit analysis for COVID-19 dataset for Europe . Classification accuracy is 73.68% , threshold score is 0, cost and random value is 55 in the cost and benefit analysis(Fig 4).

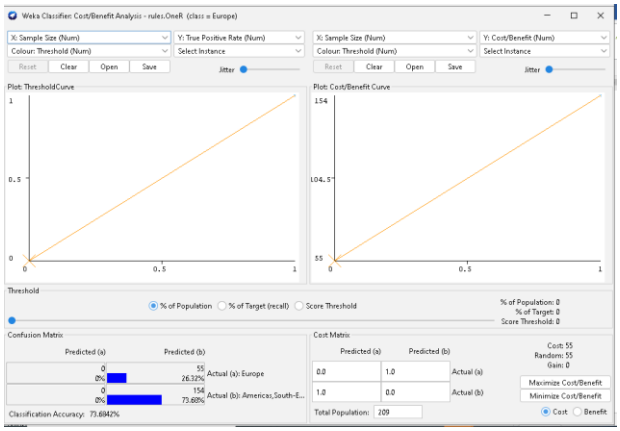


Fig. 4 Cost/Benefit analysis of Europe

Figure 5 shows the cost and benefit analysis for COVID-19 dataset for Africa. Classification accuracy is 77.51% , threshold score is 0, cost and random value is 47 in the cost and benefit analysis(Fig 5).

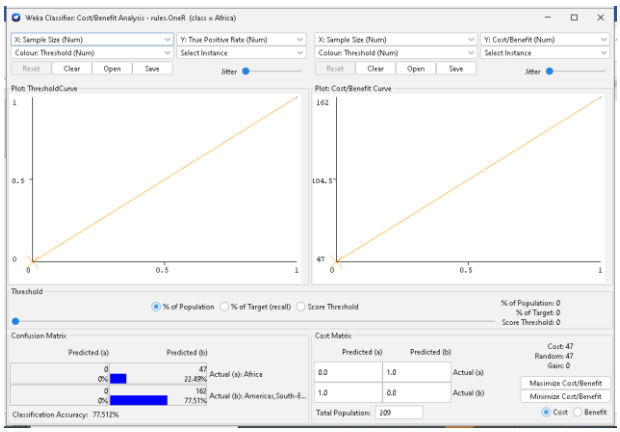


Fig.5. Cost/Benefit analysis of Africa

Figure 6 shows the cost and benefit analysis for COVID-19 dataset for Eastern Mediterranean. Classification accuracy is 89.47% , threshold score is 0, cost and random value is 22 in the cost and benefit analysis(Fig 6).

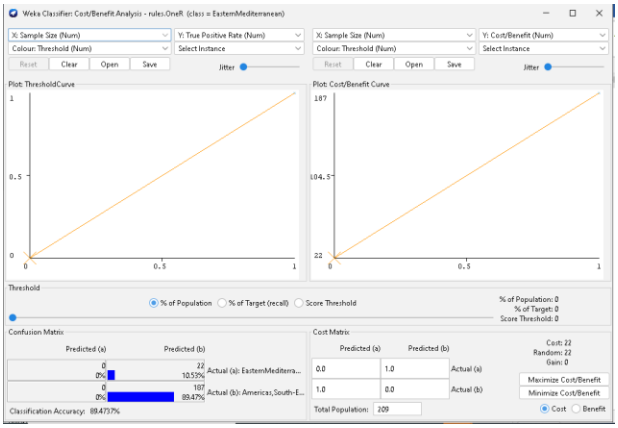


Fig.6 Cost/Benefit analysis of Eastern Mediterranean

Figure 7 shows the cost and benefit analysis for COVID-19 dataset for Eastern Mediterranean. Classification accuracy is 92.843% , threshold score is 0, cost and random value is 15 in the cost and benefit analysis(Fig 7).

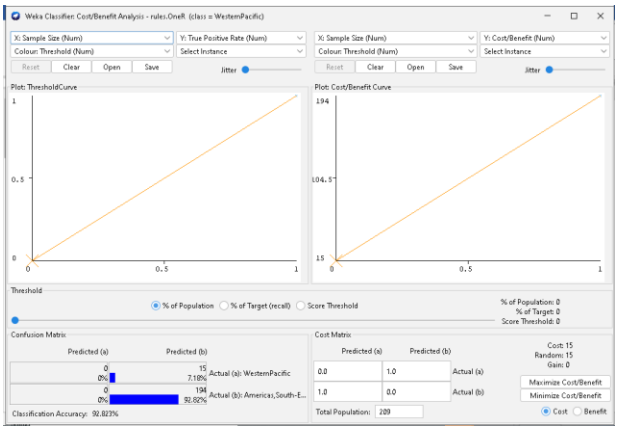


Fig. 7 Cost/Benefit analysis of WestPacific

Table 1. Cost and Benefit analysis for the COVID-19 dataset

Sl.No.	Classification Accuracy(%)	Area	Cost/ Benefit
1	83.25	Americas	35
2	95.215	South East Asia	10
3	73.68	Europe	55
4	89.47	Eastern Mediterranean	22
5	92.83	WesternPacific	15

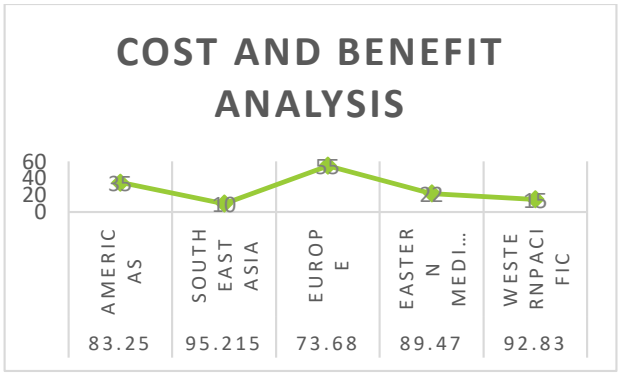
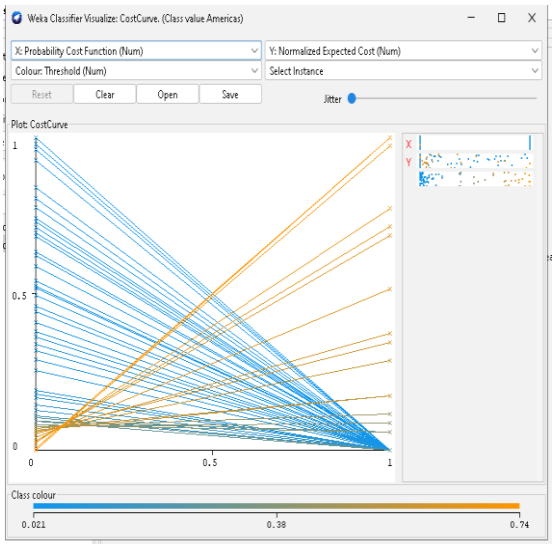


Fig. 8 Cost and Benefit analysis

The execution summary of 1R algorithm is (Table 1) ,Correctly Classified Instances rate is 19.0217 %, Incorrectly Classified Instances are 80.9783 %, Kappa statistic is 0, Mean absolute error is 0.2699, Root mean squared error is 0.5195, Relative absolute error is 0.5195, Root relative squared error is 143.5327 %,Total Number of Instances are 184, Ignored Class Unknown Instances are 25(Fig. 8).

b. Visualizing the cost curve

Cost curve generates points that show the trade-offs in probability cost that may be made by altering the threshold value between classes. For instance, the common threshold value of 0.5 indicates that for an occurrence to be projected as "positive," the anticipated probability of "positive" must be greater than 0.5(Fig. 9 -11).



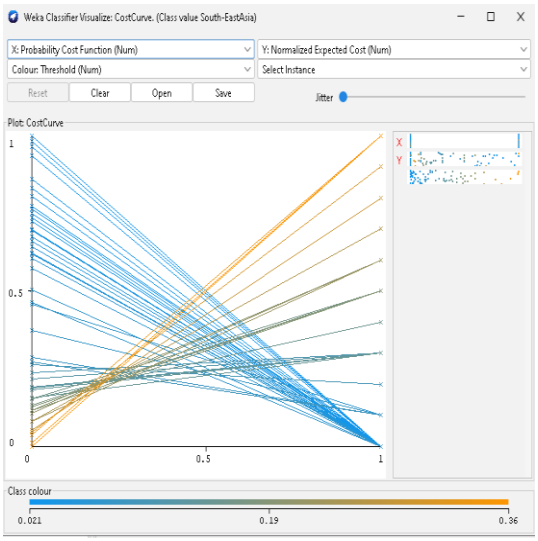
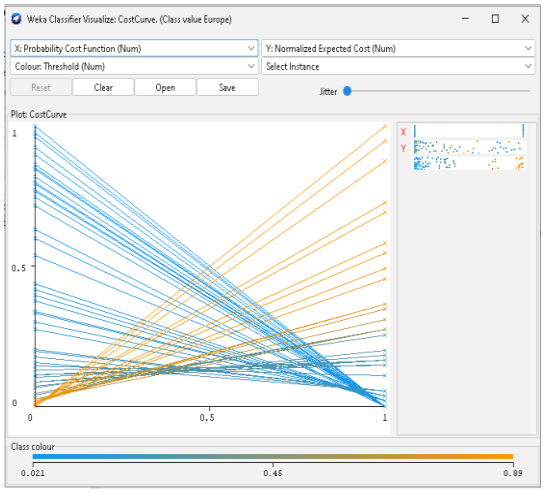


Fig. 9 Cost curve for Americas & South East Asia



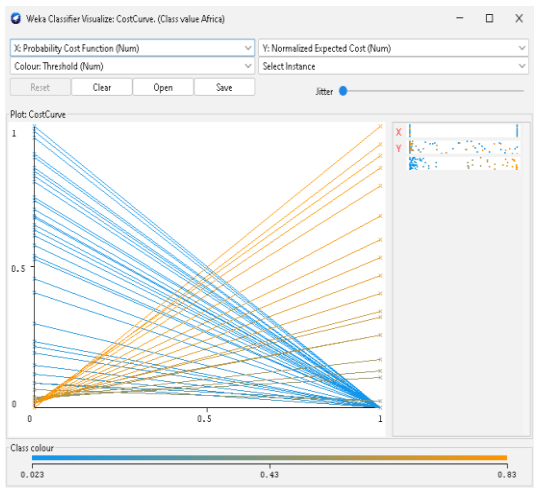
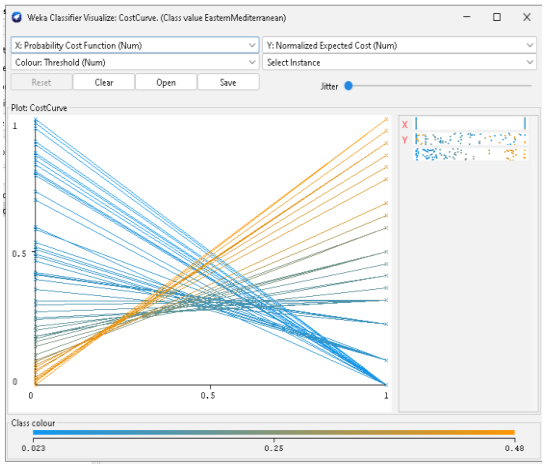


Fig. 10 Cost curve for Europe & Eastern Mediterranean



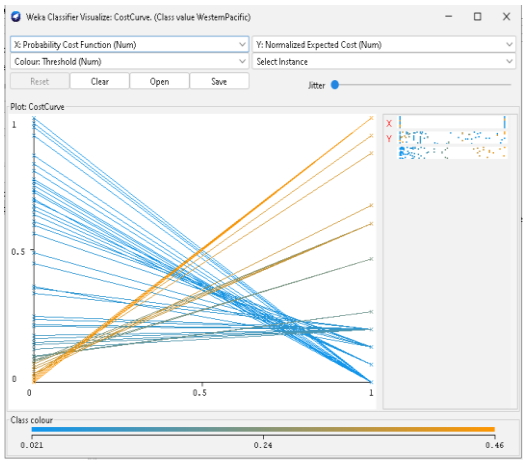


Fig. 11 . Cost curve for Western Pacific

CONCLUSIONS

This paper has aimed to establish a prediction model for COVID-19 occurrence. We have applied the 1R classification algorithm, analysed confusion matrix and proposed the cost benefit analysis ratio for five continents. We have used weka data mining tool for classification and the improvement in classification 1R algorithm to have fruitful results for cost benefit analysis. In the cost benefit analysis achieved 95.21% accuracy for South East Asia and the result of cost and benefit value is 10.Our future work has aimed to perform cost and benefit analysis using various clustering algorithms with different datasets.

References

1. Scandizzo, P.L. Impact and cost–benefit analysis: a unifying approach. *Economic Structures* **10**, 10 (2021). https://doi.org/10.1186/s40008-021-00240-w

2. Ahmed M, Afzal H, Majeed A, Khan B. A Survey of Evolution in Predictive Models and Impacting Factors in Customer Churn. Adv Data Sci Adapt Anal. 2017 Jul;09(03):1750007.
3. Xiong X, Qiao S, Li Y, Zhang H, Huang P, Han N, et al. ADPDF: A Hybrid Attribute Discrimination Method for Psychometric Data With Fuzziness. IEEE Trans Syst Man, Cybern Syst. 2018;1–14.
4. Lakshmi K, Ahmed DI, Siva Kumar G. IJSRSET32184139 | A Smart Clinical Decision Support System to Predict diabetes Disease Using Classification Techniques. 2018 IJSRSET |. 2018;4
5. Patil BM, Joshi RC, Toshniwal D. Hybrid prediction model for Type-2 diabetic patients. Expert Syst Appl. 2010 Dec;37(12):8102–8.
6. Chen P, Pan C. Diabetes classification model based on boosting algorithms. BMC Bioinformatics. 2018 Dec;19(1):109.
7. Shah A, Lynch S, Niemeijer M, Amelon R, Clarida W, Folk J, et al. Susceptibility to misdiagnosis of adversarial images by deep learning based retinal image analysis algorithms. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE; 2018. p. 1454–7.
8. Noman Sohail , Ren Jiadong , Muhammad Musa Uba , Muhammad Irshad and Ayesha Khan, Classification and cost benefit Analysis of Diabetes mellitus Dominance , IJCSNS International Journal of Computer Science and Network Security, VOL.18 No.10, October 2018 29 Manuscript received October 5, 2018 Manuscript revised October 20, 2018
9. https://waikato.github.io/weka-wiki/getting_help
10. World Health Organization (WHO). Coronavirus 2021. Available online: https://www.who.int/health-topics/coronavirus (accessed on 23 May 2021)
11. Supervised vs Unsupervised Learning: Key Differences. Available online: https://www.guru99.com/supervised-vsunsupervised-learning.html (accessed on 27 May 2021
12. https://www.cs.waikato.ac.nz/ml/weka/