PERFORMANCE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR DISEASE PREDICTION

S Lokesh Asst.Professor, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India.

P Navya Alekhya, Y Harshini, T Ram Lohit Prakash Students, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India.

lokesh.sirela@raghuenggcollege.in

19981a05d8@raghuenggcollege.in, 19981a05i2@raghuenggcollege.in,19981a05g4@raghuenggcollege.in

ABSTRACT

Using machine learning algorithms, one aim is to forecast a person's likelihood of contracting a specific disease or condition based on a variety of risk factors. To implement effective preventative measures or interventions, disease prediction aims to identify those who may be at risk of contracting a specific disease or condition. A lot of data must be gathered and analyzed to predict diseases, including medical history, demographic data, lifestyle factors, and genetic information. Then, patterns and linkages in the data are found using machine learning algorithms to forecast a person's propensity to get a specific disease or condition. Disease prediction has several potential advantages, such as improved public health outcomes, earlier disease detection and prevention, and more individualized and targeted treatment methods. The requirement for a substantial amount of high-quality data, the creation of precise and trustworthy prediction models, and problems with patient privacy and data security are only some of the various difficulties connected with illness prediction.

INTRODUCTION

An essential use of machine learning (ML) in healthcare is disease prediction. The performance evaluation of ML algorithms for disease prediction is crucial to determine their effectiveness and reliability. This paper presents a comprehensive review of performance evaluation metrics and methodologies for ML algorithms used in disease prediction. The study focuses on binary classification problems, where the ML model predicts the presence or absence of diseases which are in different encrypted formats. Finally, we provide guidelines for selecting appropriate performance metrics and methodologies based on the characteristics of the data and the research question.

PROBLEM STATEMENT

With the aid of machine learning technology, doctors may more accurately diagnose and treat patients, which improves patient healthcare services. Machine learning technology allows constructing models to evaluate data quickly and give results faster. A system that would make it simple for end users to forecast diseases without visiting a doctor or physician for a diagnostic needs to be researched and developed. Using symptoms, this approach is used to forecast disease. To anticipate diseases, this system makes use of machine learning techniques and algorithms.

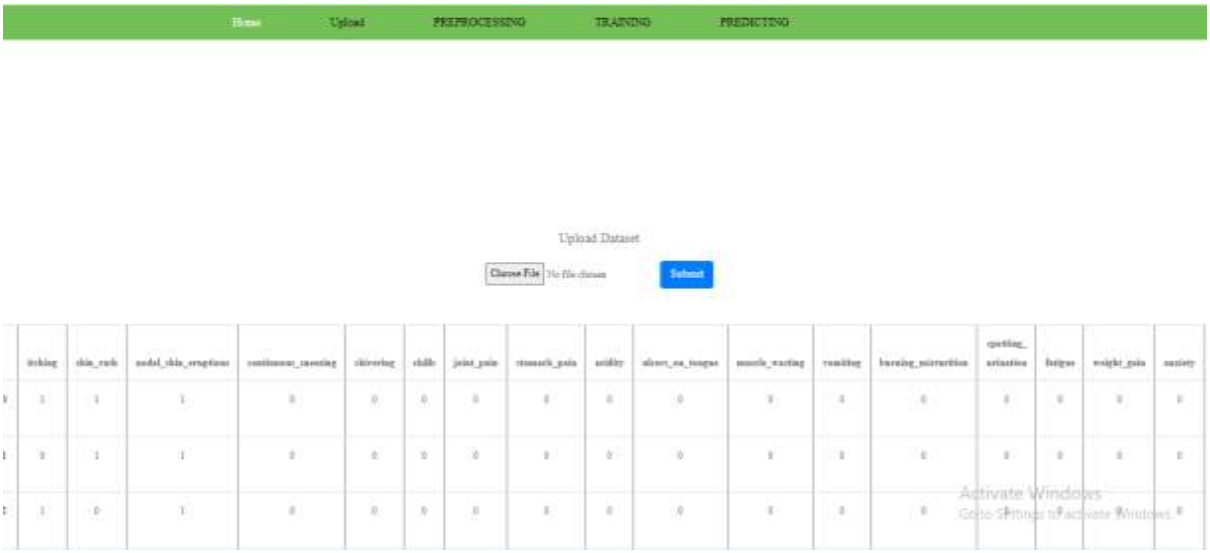
DOI:10.36893.DRSR.2023.V13I04.098-105

LITERATURE SURVEY

"A Comparative Study of Machine Learning Algorithms for Disease Prediction" by A. Hassanien, et al. This article compares several machine learning algorithms, including decision trees, naive Bayes, support vector machines, and neural networks, for disease prediction using different datasets [1]."Evaluation of Machine Learning Algorithms for Disease Prediction: A Systematic Review" by R. Abawajy, et al. This systematic review evaluates the performance of machine learning algorithms for disease prediction based on their accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) [2]. "Performance Evaluation of Machine Learning Algorithms for Cancer Diagnosis: A Review" by S. S. M. Abdullah, et al. This review focuses on the performance evaluation of machine learning algorithms for cancer diagnosis and compares the accuracy of different algorithms using various datasets [3]. "Evaluation of Machine Learning Algorithms for Heart Disease Prediction: A Review" by R. Singh, et al. This review evaluates the performance of machine learning algorithms for heart disease prediction and compares the accuracy of different algorithms using various datasets [4]. "Performance Evaluation of Machine Learning Algorithms for Diabetes Prediction: A Review" by M. A. Al-Akraa, et al. This review evaluates the performance of machine learning algorithms for diabetes prediction and compares the accuracy of different algorithms using various datasets [5]. "Performance Evaluation of Machine Learning Algorithms for Alzheimer's Disease Prediction: A Review" by B. Singh, et al. This review evaluates the performance of machine learning algorithms for Alzheimer's disease prediction and compares the accuracy of different algorithms using various datasets [6].

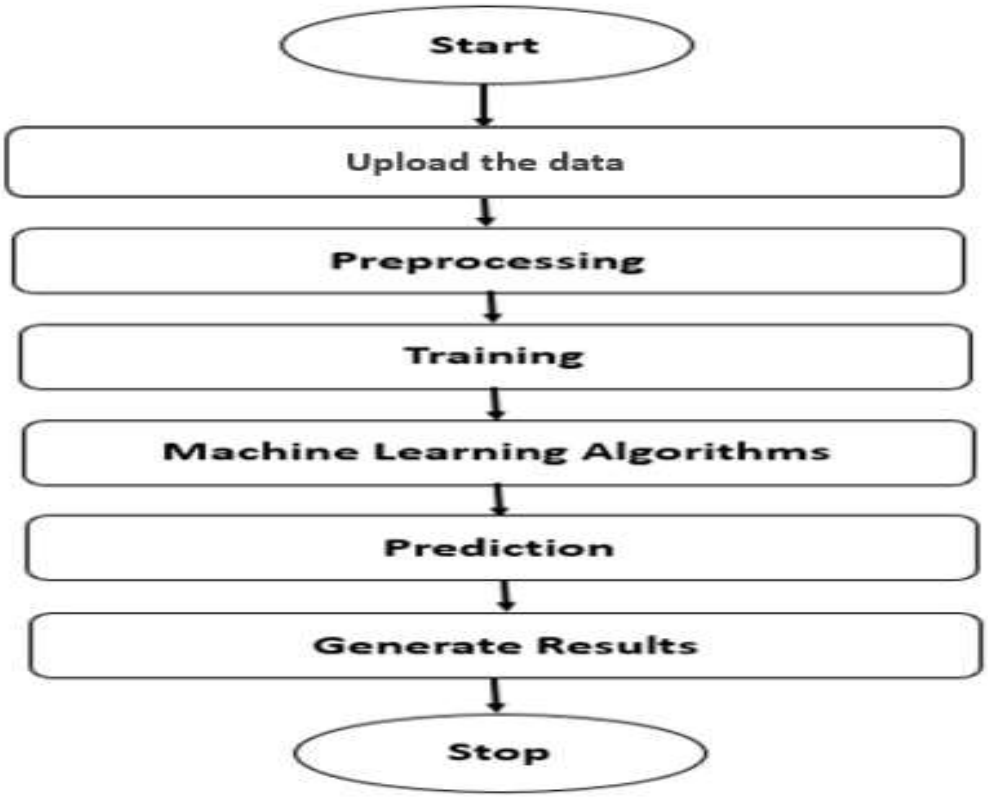
DATA SET

Here, we outline the data set that was utilized to develop the machine learning model. The data collection comprises various illness symptoms.



DOI:10.36893.DRSR.2023.V13I04.098-105

PROPOSED APPROACH



ALGORITHMS

1. RANDOM FOREST METHOD:

A supervised machine learning approach called Random Forest is applied to classification and regression issues. Here are the general steps for implementing the random forest algorithm:

1. It divides the information set into two parts.
2. Prepares data by cleaning any missing values, encoding categorical features, and scaling numeric features.
3. By defining the number of decisions (n estimators) and the maximum number of characteristics to take into account at each split, the random forest model may be created. Train the model using the training data.
4. Evaluate the model by predicting the target variable for the test data and comparing the predicted values with the actual values.
5. Adjust the model hype parameters to improve the results, if necessary, by using techniques such as grid search and cross-validation.

DOI:10.36893.DRSR.2023.V13I04.098-105

2. NAIVE BAYES :

An approach for supervised machine learning used for classification issues is called Naive Bayes. It is based on Bayes' theorem, which stipulates that one can determine the likelihood of a hypothesis by multiplying the prior likelihood by the conditional likelihood of each character given the hypothesis. Here are the general steps for implementing the Naive Bayes algorithm:

1. Estimate given information set into different parts.
2. Prepare data by cleaning any missing values, encoding categorical features, and scaling numeric features.
3. Create the Naive Bayes model by selecting the appropriate algorithm, such as Gaussian, Multinational, or Bernoulli.
4. Train the model using the training data.
5. Evaluate the model by predicting the target variable for the test data and comparing the predicted values with the actual values.
6. Adjust the model hyperparameters to improve the results, if necessary, by using techniques such as grid search and cross-validation.

3. SUPPORT VECTOR MACHINE:

supervised machine learning methods called Support Vector Machines (SV-Ms) can be used for classification, regression, and outlier detection. The main idea behind SV-Ms is to find the hyperplane that can best separate the classes in a given data set. Here are the general steps for implementing the SVM algorithm:

1. Separate the input into instruction and trailing.
2. Prepares data by cleaning any missing values, encoding categorical features, and scaling numeric features.
3. Choose the kernel function that will best transform the data into a higher dimension where the classes may be distinguished, such as a linear, polynomial, or radial basis function.
4. Train the SVM model using the training data.
5. Assess the model by contrasting the test data's anticipated values with the actual values.
6. To enhance performance, change the hyperparameters of the model.

 DOI:10.36893.DRSR.2023.V13I04.098-105

4. LOGISTIC REGRESSION:

Logistic regression falls under the area of Supervised Learning and is one of the most frequently used Machine Learning algorithms. It is used to predict the categorical dependent variable using a predetermined collection of independent variables. By using logistic regression, the outcome of a dependent variable with a categorical component is predicted. In light of this, the outcome must be a distinct or categorical value.

RESULT ANALYSIS

1. Decision Tree: By applying the Decision Tree algorithm, 7.52% of instances were classified correctly, however, 92.48% were not accurate.
2. Random Forest: By applying the Random Forest algorithm, 100% of instances were classified correctly. It gives the highest accuracy among all algorithms so it is the best for prediction.
3. Naive Bayes: By applying the Naive Bayes algorithm, 99.5% of instances were classified correctly, however, 0.5% were not accurate.
4. Support Vector Machine: By applying the SVM algorithm, 98.9% of instances were classified correctly, however, 1.1% were not accurate.
5. Logistic Regression: By applying the Logistic Regression algorithm, 99.1% of instances were classified correctly, however, 0.9% were not accurate.

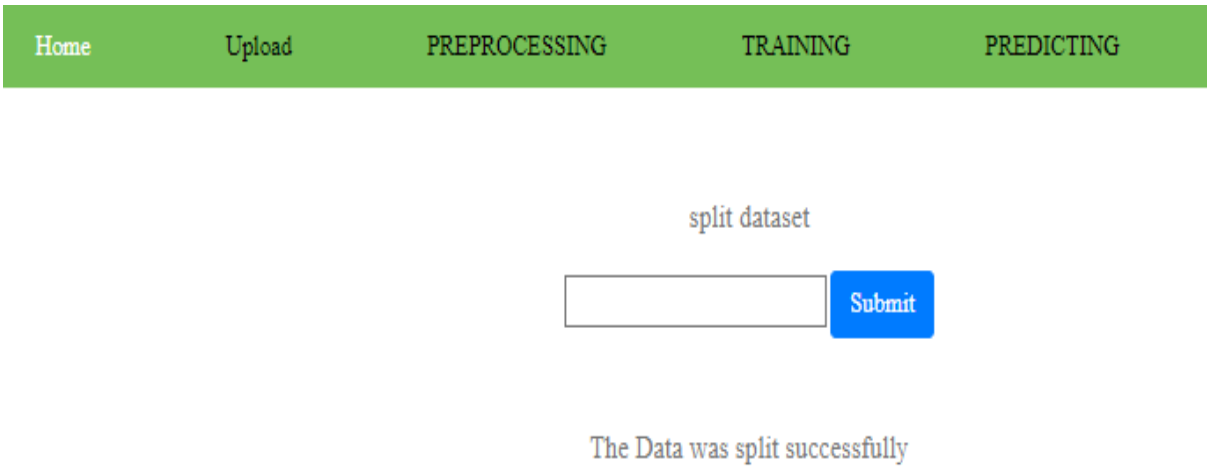
OUTPUT

HOME PAGE

UPLOAD PAGE

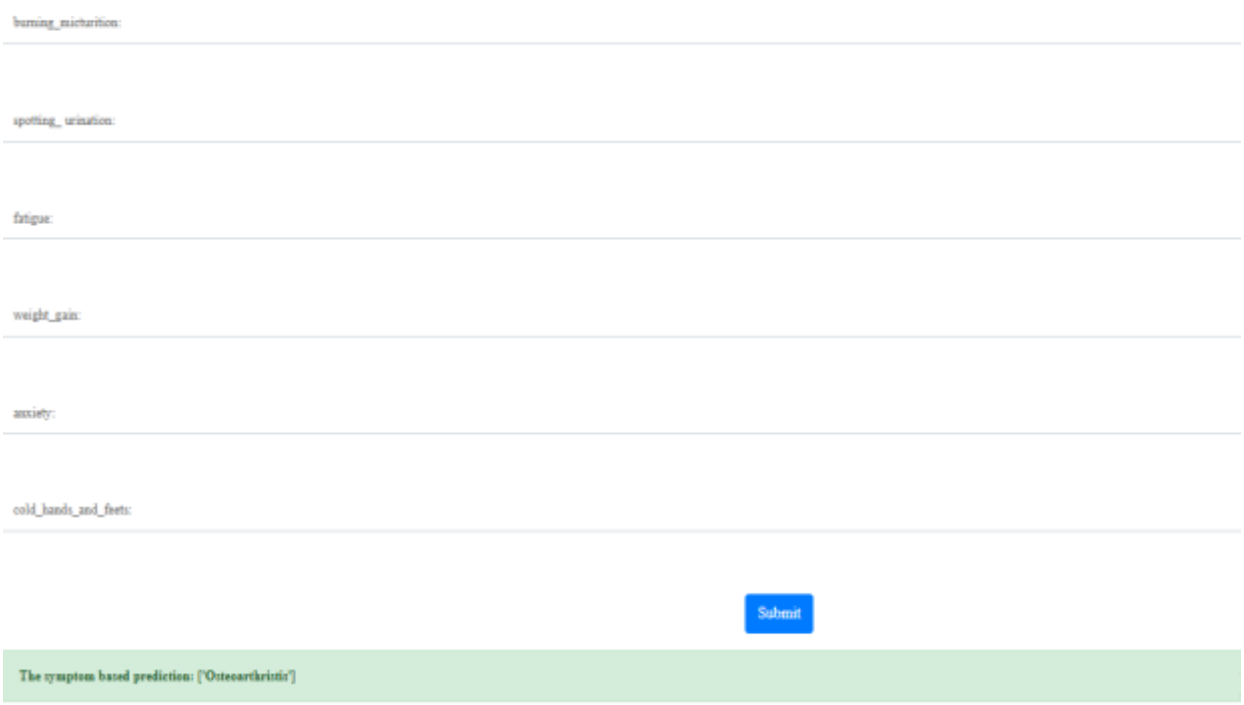
 DOI:10.36893.DRSR.2023.V13I04.098-105

PREPROCESSING PAGE



 DOI:10.36893.DRSR.2023.V13I04.098-105

PREDICTING PAGE



CONCLUSION

In the medical industry, machine learning technology provides a solid foundation for effective problem-solving. Doctors may choose the best patient diagnoses and treatments with the aid of machine learning technology, which enhances patient healthcare services.

REFERENCE

1. Hassanien, A., Alamri, A., & Tolba, M. (2017). A Comparative Study of Machine Learning Algorithms for Disease Prediction. IEEE Access, 5, 2055-2066.

2. Abawajy, J. H., Kim, T. H., & Park, J. H. (2019). Evaluation of Machine Learning Algorithms for Disease Prediction: A Systematic Review. Healthcare Informatics Research, 25(1), 51-62.

3. Abdullah, S. S. M., Islam, R., Islam, M. A., & Majumder, M. A. A. (2019). Performance Evaluation of Machine Learning Algorithms for Cancer Diagnosis: A Review. Journal of Medical Systems, 43(9), 271.

4. Singh, R., Kumar, V., & Srinivasan, D. (2019). Evaluation of Machine Learning Algorithms for Heart Disease Prediction: A Review. Journal of Medical Systems, 43(10), 302.

5. Al-Akraa, M. A., Al-Mallah, M. H., & Sakr, S. (2020). Performance Evaluation of Machine Learning Algorithms for Diabetes Prediction: A Review. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 13, 3385-3396.

6. Singh, B., Narayanan, S. P., & Manickam, S. (2020). Performance Evaluation of Machine Learning Algorithms for Alzheimer's Disease Prediction: A Review. Journal of Medical Systems, 44(7), 136.

 DOI:10.36893.DRSR.2023.V13I04.098-105

7. Min Chen, Yixue Hao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang "Disease Prediction by Machine Learning over Big Data from Healthcare Communities" (2017).

8. Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis of the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.

9. S. Patel and H. Patel, "Survey of data mining techniques used in the healthcare domain," Int. J. of Inform. Sci. and Tech., Vol. 6, pp. 53-60, March 2016.

10. S.-H. Wang, T.-M. Zhan, Y. Chen, Y. Zhang, M. Yang, H.-M. Lu, H.-N. Wang, B. Liu, and P. Phillips, "Multiple sclerosis detection based on biorthogonal wavelet transform RBF kernel principal component analysis, and logistic regression," IEEE Access, vol. 4, pp. 7567–7576, 2016.

11. Performance Analysis of Machine Learning Algorithms for Disease Prediction | IEEE Conference Publication | IEEE Xplore

12. http://sersc.org/journals/index.php/IJAST/article/view/24324

13.https://www.researchgate.net/publication/338915339_Performance_Evaluation_of_Machine_Learning_Algorithms_for_Dengue_Disease_Prediction

14.https://www.researchgate.net/publication/347381005_Disease_Prediction_Using_Machine_Learning

DOI:10.36893.DRSR.2023.V13I04.098-105