

A Performance Comparison Of Kernel Svm And Hyperparameter Algorithm Using Machine Learning Techniques For Pregnancy Women's

Mahadev Bag¹, Dr. Abhisek Badholia²

^{1,2}MATS School of Engineering and IT, MATS University, Gullu, Arang, Raipur (C.G.), India

Corresponding Author: Mahadev Bag (bagmahadev1010@gmail.com)

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract

Objective: To design a framework for the diagnosis of the ANC which will be used to predict, monitor, analyze and forecast the performance measure in the field of ANC in an optimal period, identify the severity of the disease and promote institutional delivery to pregnant women.

Method: In this study, we have used the RCH dataset and using classification model kernel SVM and GridSearchCV Hyperparameter algorithms or applying confusion matrix for accuracy.

Result: In this study based on the dataset in Kernel SVM we have got 85.75% accuracy and our Hyperparameter algorithm GridSearchCV we have got 86% accuracy.

Conclusion: In this study, we have concluded with two algorithms Kernel SVM and GridSearchCV, both the algorithms produced results based on parameters.

Keywords: ANC, SVM, GridSearchCV, sklearn, StandarScaler.

1. INTRODUCTION

Transformation of developing countries like India majorly attentions in the advancement of health care services for the enhanced diagnostic system. Health care service needs regular changes in quality of services at affordable cost and in an optimal time for diagnosing the patients [1-4] [6]. The quality of services includes diagnosing disease truly and providing good treatment at an earlier stage of the disease [5] [7].

In this technology age availability of data in the Health Care Information society leads to the need for a valid tool for planning, monitoring, analyzing, and decision making. In woman's health care data available as manual as well file format in the medical field may lead to delay the procedure of monitoring health care results, planning with health data gets late and analyzing and taking right decision may get late [5].

Medical tests, machine learning, and pattern classification methods have been widely used for the early diagnosis of diseases in medicine for decision making by specialists [8]. In the field of diagnostic systems, the possible errors are reduced in the diagnosis stage, and the detailed medical data can be analyzed in a shorter time [4]. Accurate prediction using health data and proper diagnosis of disease at a time helps to save many patients. This can be achieved by machine learning techniques for diagnosing disease in a large dataset and to predict the severity of the disease [8].

Machine learning is the process of selecting, exploring, modelling, predicting, and analyzing a large database to discover the unknown model [10]. It is the whole process of data extraction and analysis aimed at the production of decision rules for specified health care issues.

This study aims to design a framework for pregnancy women's healthcare management, to predict the accuracy in diagnosing the normal delivery in an optimal period using machine learning classification techniques.

2. MOTIVATION

Proper diagnosis and accurate prediction of the disease on time can save many patients, but in the current scenario, data are scattered with different archive systems and are not linked with one another in the health care environment [5]. Additionally, healthcare data comprise thousands of records that may contain valuable patterns hidden deep among them.

Lack of valid tool leads may lead to delay in monitoring, improper planning, defocus the analysis, and misleading decision making for decision-makers. Data mining is one of the optimal tools to diagnose disease in patients in a large database and predict the rigorousness of disease.

It is the biggest challenge for many hospitals and public health care services in diagnosing hepatitis at the early stage because the symptoms of hepatitis are identified only in the later stage [10]. Further, it is the general tendency of many certified medical professionals to make decisions based on their heuristic experiences rather than knowledge extracted from rich data repositories. Delay in quality of services contributed by scattered patient's information leading to defocus in quality of decision making by the physicians in the health care system [5].

The need is to develop a valid tool for predicting, monitoring, analyzing, and decision making for identifying disease and categorizing their types in hepatitis disease. It also requires proper synchronizing of health care services and futuristic prediction of risk management in hepatitis disease.

3. METHODOLOGY

In this research have been given 6 phases to predicting the different result between two classification methods based on default parameters and hyper tuning parameters their steps are as follows:

Importing Library:

In this research work, we have applied a machine learning classification model using python programming and anaconda tools for predicting the result. Importing a library is very important for pandas, NumPy array to work with dataset.

Importing dataset:

Without data you cannot perform any single task in the machine learning technique, in such a case first import the dataset from the database, here in our database table there are 22 columns and 1267 data is available, the dataset has been collected from District Health Department, Raipur, Chhattisgarh.

Splitting the Dataset:

Before going to training and testing, dataset splitting is a must for the best prediction result, here we have split 75% data for training and 25% data for testing.

Feature Scaling:

Result transformation is very important for any model, in this scenario feature selection is a very important concept in machine learning, so that model result transform, for feature scaling in this study applied StandarScaler class in sklearn library for feature scaling.

Data Modelling:

Before predicting the result, model training is a must, Modelling is about defining the rules of the relationship by which the data will organize. Data modelling applies different data analysis approaches to build machine learning models to predict and forecast the future. in this research, we have trained the Kernel SVM model and GridSearchCV model with parameters tuned.

Confusion Matrix:

Finally, we have applied confusion matrix for result accuracy in both the model.

4. RESULT AND DISCUSSION

Preparation, as the name suggests, is all about preparing the data you have for analysis. Clean the data, highlight inconsistencies, deal with missing values, and convert it into a format suitable for your analysis. After the preparation of the dataset, we have got 22 columns out of 49 columns and 1267 rows for result prediction.

Tabel 1: Mother Health care data set

Table 3: Feature Scaling result

```

[[-2.98187752e-01  0.00000000e+00  0.00000000e+00 ... -5.77425618e-01
 -3.70960950e-01 -7.86039399e-01]
 [-9.84358430e-01  0.00000000e+00  0.00000000e+00 ...  3.90412639e-01
  2.83676021e-01  2.70209776e-01]
 [-9.84358430e-01  0.00000000e+00  0.00000000e+00 ...  3.90412639e-01
 -3.70960950e-01  1.32645895e+00]
 ...
 [ 8.45430046e-01  0.00000000e+00  0.00000000e+00 ... -9.35064898e-02
 -5.34620193e-01  1.32645895e+00]
 [ 1.59259367e-01  0.00000000e+00  0.00000000e+00 ...  3.87463036e+00
  1.20016778e-01  5.34272069e-01]
 [-5.26911311e-01  0.00000000e+00  0.00000000e+00 ...  3.27733590e-03
 -3.70960950e-01 -7.86039399e-01]]
    
```

In this paper, we have applied Kernel SVM and -Parameter model to train the dataset, the parameter of kernel SVM are (kernel='rbf' random_state=0). Now model produced prediction result as follows:

Table 4: Kernel SVM prediction result

```

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

The parameters of GridSearchCV model are (C :[0.1,1,10,100,1000], gamma :[1, 0.1, 0. 01, 0.001, 0.0001], kernel='sigmoid').

Finally applied confusion matrix to find the accuracy score in both the model in kernel SVM there is 85.75 percent accuracy score and after parameter tuning in GridSearchCV model we have got 86.00 percent accuracy score in such point of view GridSearchCV model is better than kernel SVM after parameter tuning.

5. CONCLUSION

Machine learning is the process of selecting, exploring, modeling, predicting, and analyzing a large database to discover the model that is an unknown dataset. In this research we have to find the result based on Kernel SVM, after the got result, we have changed some parameters in Hyper-Parameter GridSearchCV algorithms, we found the better result as compare to Kernel SVM, In the future, we will create our model to find the result more than this both algorithms.

REFERENCES

1. S. Bashir, U. Qamar, and F. H. Khan, "IntelliHealth : A medical decision support application using a novel weighted multi-layer classifier ensemble framework," *J. Biomed. Inform.*, vol. 59, pp. 185–200, 2016.
2. L. Bolondi et al., "Characterization of Small Nodules in Cirrhosis by," pp. 27–34, 2001.
3. D. Çalis, "Expert Systems with Applications Short communication A new intelligent hepatitis diagnosis system : PCA – LSSVM," vol. 38, pp. 10705–10708, 2011.
4. H. Chen, D. Liu, B. Yang, J. Liu, and G. Wang, "Expert Systems with Applications A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11796–11803, 2011.
5. G. Sathyadevi, "Application of CART algorithm in hepatitis disease diagnosis," *Int. Conf. Recent Trends Inf. Technol. ICRTIT 2011*, pp. 1283–1287, 2011.
6. R. Duangsoithong and T. Windeatt, "Relevant and Redundant Feature Analysis with Ensemble Classification," 2007.
7. R. Lin, "An intelligent model for liver disease diagnosis," 2009
8. K. Polat and G. Salih, "A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS," vol. 8, pp. 164– 174, 2007
9. S. Baki, "Expert Systems with Applications Diagnosis of liver disease by using CMAC neural network approach," vol. 37, pp. 6157–6164, 2010.
10. K. Leung et al., "Data Mining on DNA Sequences of Hepatitis B Virus," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 8, no. 2, pp. 428–440, 2011.