International Journal of Grid and Distributed Computing
Vol. 14, No. 1, (2021), pp. 1727–1733

A Performance Comparison of Kernel SVM with Tuned Hyperparameter GridSearchCV Algorithm using Machine Learning Techniques for Child Immunization

Mahadev Bag*, Dr. Abhisek Badholia* , Dr. Vishwaprakash Roy*

MATS School of Engineering and IT, MATS School of Science, MATS University, Raipur (C.G.)

(*Corresponding author's e-mail: bagmahadev1010@gmail.com)

Abstract:

To design a model for the diagnosis of Child Immunization that will be used to predict, analyse, monitor, and forecast the performance measure in the field of Child Immunization in an optimal period, identify the severity of the child growth. In this research, we have used the Child Immunization dataset and using classification model kernel SVM and GridSearchCV Hyperparameter algorithms or applying confusion matrix for accuracy. In this research based on our dataset, in Kernel SVM, we have got 75% accuracy and our Hyperparameter algorithm GridSearchCV, we have received 77% accuracy. In this research, we have concluded that with two algorithms Kernel SVM and GridSearchCV, our approach of hyper tuning parameters of GridSearchCV algorithm produced results better than Kernel SVM with our child immunization dataset.

Keywords: SVM, GridSearchCV, Poly, RBF, Sigmoid, Immunization

I. INTRODUCTION

Immunization of children against six preventable diseases (diphtheria, tuberculosis, pertussis, tetanus, polio, and measles) is vital to reduce childhood mortality and morbidity across the Universe. So, it is one of the pointers of development in most developing countries. The Expanded Program on Immunization (EPI) was launched in 1974 as a global program for controlling and reducing death from Vaccine-Preventable Diseases (VPDs). Thus, vaccine coverage is estimated as by resolution with Diphtheria, Pertussis, and Tetanus-3 (DPT-3) coverage achieved among children aged 12 to 23 months [3][10]. Machine learning (ML) is an efficient tool for performing artificial intelligence (AI) tasks, through the implementation of robust and computationally feasible algorithms [1][6]. In the last decade, machine-learning techniques have been used extensively for a wide range of child immunization applications [2][8]. The usefulness of machine learning into child immunization data to improve the performance of result in child immunization to achieving various tasks without requiring software re-programming. This study is based on the idea of developing a robust model for predicting child immunization. It involves the database of child immunization. Kernel SVM classification technique is used to predict child immunization successfully.[1] Also, in the study, we are using a hyperparameter model to predicting the best result for child immunization.[1]

II. PROPOSED METHODOLOGY

Dataset: Data is a raw fact, when we give a meaning full information of data than its produce valuable result for us, in our project we are using child immunization dataset[4][5][6], which collected from Govt. Block hospital Aarang, Raipur(C.G.), in this dataset there are 51 columns and 2618 rows are

ISSN: 2005-4262IJGDC

available, but in our project, we are using 10 input variables and 1 output variable, which are important for our project.

Table 2.1 Child Immunization Dataset

Weight	BCG	OPV0	OPV1	OPV2	OPV4	HEP1	PENTA1	PENTA2	PENTA3	Immunizetion
3.5	1	1	1	1	1	1	1	1	1	1
3.5	1	1	1	1	1	1	1	1	1	1
2	1	1	0	0	0	1	0	0	0	1
2.5	1	1	1	1	1	0	1	1	1	1
2.5	1	1	1	1	1	1	1	1	1	1
2.5	1	1	0	0	0	1	0	0	0	1
2.5	1	1	0	0	0	1	0	0	0	0
2.5	1	1	1	1	1	0	1	1	1	1

Data prepossessing: One of the most important mechanisms in a machine learning algorithms is to train our algorithms on a training dataset that is separate from the test set which we will get its accuracy[4]. In our raw dataset out of 51 columns and 2618 rows, we have taken only 10 important columns and filled the binary values 0 in empty cells, Here we are using 75/25 to split the dataset for training, Here we are using train_test_split class to sklearnLibray and,We have passed four parameters in train_test_split constructor.

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)
```

Feature Scaling: Feature selection offers a simple and effective way to overcome this challenge by eliminating redundant and irrelevant data[9]. Removing the irrelevant data progresses learning accuracy, reduces the computation time, and facilitates an enhanced understanding of the learning model or data. In our project, we are using StandardScaler class and Fit_transform method for the training set and transform for the testing set.

Model Selection: In this project, we are using the KernelSVM model, In this model passed parameters in SVC() method degree, gramma, and various kernel types like poly, RBF, sigmoid[9].

Accuracy: Machine learning algorithm accuracy is the measurement used to determine, which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training data [4]. In this research for accuracy, we have used the train_test_split() method for separate data into training and test set, fit() method for making predictions, predict() for evaluating our model, and classification_report() method for evaluations.

Tuning the hyper-parameters of an estimator

Hyper-parameters are parameters that are not directly learned within estimators. In the scikt-learn library, they are passed as arguments to the constructor of the estimator classes. Grid search is commonly used as an approach to hyper-parameter tuning that will methodically build and evaluate a model for each combination of algorithm parameters specified in a grid.

Create a dictionary called param_grid and fill out some parameters for kernels, c, and gamma.

```
param_grid = {'C': [0.1, 1, 10, 100, 1000],

        'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
```


ISSN: 2005-4262IJGDC

'kernel': ['rbf']}

Here are parameters grid are c, gramma, kernel type is RBF.

Fitting the model for grid search sing grid.fit() method. For accuracy prediction, we are applied predict() method and for classification report classification_report() method.

III. RESULTS AND DISCUSSION

To create a machine learning model, we required is a dataset as a machine learning model completely works on data. The collected data for a particular problem in a proper format is known as the dataset [4][7]. After the preparation of the dataset, we have received 11 columns out of 51 columns and 2618 rows of data for result prediction.

Table 3.1: Child Immunization Dataset

	Weight	BCG	OPV0	OPV1	OPV2	OPV4	HEP1	PENTA1	PENTA2	PENTA3	Immunization
0	3.5	1	1	1	1	1	1	1	1	1	1
1	3.5	1	1	1	1	1	1	1	1	1	1
2	2.0	1	1	0	0	0	1	0	0	0	1
3	2.5	1	1	1	1	1	0	1	1	1	1
4	2.5	1	1	1	1	1	1	1	1	1	1
...	...	...	...	...	...	...	...	...	...	...	...
2613	3.5	1	1	1	0	0	1	1	0	0	1
2614	3.2	1	1	1	1	0	1	1	1	0	1
2615	2.0	1	1	1	1	0	1	1	1	0	0
2616	2.5	1	1	1	1	0	1	1	1	0	1
2617	2.5	1	1	1	1	1	1	1	1	1	0

2618 rows × 11 columns

The below table has been divided into 75 percentage datasets and 25 percentage for the testing dataset, the below diagram showing the result of the training and testing dataset.

Feature selection is the process of using domain knowledge to extract features raw data. A feature is a property shared by independent units on which analysis or prediction is to be done. Features are used by predictive models and influence results. The feature selection table is as follows:

Table 3.2: Training and Testing Dataset

ISSN: 2005-4262IJGDC

```
[[3.  1.  1.  ... 0.  0.  0. ]
 [3.  1.  1.  ... 0.  0.  0. ]
 [3.  0.  0.  ... 0.  0.  0. ]
 ...
 [2.5 1.  1.  ... 1.  1.  1. ]
 [3.  1.  1.  ... 0.  0.  0. ]
 [2.6 1.  1.  ... 0.  0.  0. ]]
[1 0 1 ... 1 1 0]
[[2.5 1.  1.  ... 0.  0.  0. ]
 [3.5 1.  1.  ... 0.  0.  0. ]
 [3.  1.  1.  ... 0.  0.  0. ]
 ...
 [1.4 1.  1.  ... 1.  1.  1. ]
 [1.1 0.  0.  ... 0.  0.  0. ]
 [3.5 1.  1.  ... 1.  0.  0. ]]
[1 0 1 1 1 0 0 1 1 0 1 1 1 0 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 0 1
 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 1 0 0 0 1 1 1
 1 0 0 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1
```

Table 3.3: Feature's Selection Result

```
[[ 0.38625455  0.3714445   0.39159801 ... -0.66347979 -0.57205606
  -0.46905483]
 [ 0.38625455  0.3714445   0.39159801 ... -0.66347979 -0.57205606
  -0.46905483]
 [ 0.38625455 -2.69219226 -2.55363913 ... -0.66347979 -0.57205606
  -0.46905483]
 ...
 [-0.75357962  0.3714445   0.39159801 ...  1.50720492  1.74808041
   2.13194694]
 [ 0.38625455  0.3714445   0.39159801 ... -0.66347979 -0.57205606
  -0.46905483]
 [-0.52561278  0.3714445   0.39159801 ... -0.66347979 -0.57205606
  -0.46905483]]
[[-0.75357962  0.3714445   0.39159801 ... -0.66347979 -0.57205606
  -0.46905483]
 [ 1.52608873  0.3714445   0.39159801 ... -0.66347979 -0.57205606
  -0.46905483]
 [ 0.38625455  0.3714445   0.39159801 ... -0.66347979 -0.57205606
  -0.46905483]
 ...
 [-3.26121479  0.3714445   0.39159801 ...  1.50720492  1.74808041
   2.13194694]
 [-3.9451153  -2.69219226 -2.55363913 ... -0.66347979 -0.57205606
  -0.46905483]
 [ 1.52608873  0.3714445   0.39159801 ...  1.50720492 -0.57205606
  -0.46905483]]
```

Confusion matrix:

A Confusion matrix is an M x M matrix used for evaluating the performance of a classification model, where M is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model [4]. The accuracy of the kernel SVM model based on above mention kernel types and parameters are as follows:

Table 3.4: Accuracy result of Kernel SVM

ISSN: 2005-4262IJGDC

	precision	recall	f1-score	support
0	0.40	0.01	0.03	141
1	0.73	0.99	0.84	383
accuracy			0.73	524
macro avg	0.57	0.50	0.43	524
weighted avg	0.64	0.73	0.62	524

	precision	recall	f1-score	support
0	0.00	0.00	0.00	129
1	0.75	1.00	0.86	395
accuracy			0.75	524
macro avg	0.38	0.50	0.43	524
weighted avg	0.57	0.75	0.65	524

	precision	recall	f1-score	support
0	0.17	0.11	0.14	115
1	0.77	0.85	0.81	409
accuracy			0.69	524
macro avg	0.47	0.48	0.47	524
weighted avg	0.64	0.69	0.66	524

Table 3.5: Result of GridSearchCV Object and fit it to the Training Data

```
Fitting 5 folds for each of 25 candidates, totalling 125 fits
[CV] C=0.1, gamma=1, kernel=rbf ....................................
[CV] .......... C=0.1, gamma=1, kernel=rbf, score=0.761, total=   0.1s
[CV] C=0.1, gamma=1, kernel=rbf ....................................
[CV] .......... C=0.1, gamma=1, kernel=rbf, score=0.761, total=   0.1s
[CV] C=0.1, gamma=1, kernel=rbf ....................................
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.0s remaining:    0.0s
[Parallel(n_jobs=1)]: Done   2 out of   2 | elapsed:    0.0s remaining:    0.0s
[CV] .......... C=0.1, gamma=1, kernel=rbf, score=0.761, total=   0.1s
[CV] C=0.1, gamma=1, kernel=rbf ....................................
[CV] .......... C=0.1, gamma=1, kernel=rbf, score=0.761, total=   0.1s
[CV] C=0.1, gamma=1, kernel=rbf ....................................
[CV] .......... C=0.1, gamma=1, kernel=rbf, score=0.763, total=   0.1s
[CV] C=0.1, gamma=0.1, kernel=rbf ..................................
[CV] ........ C=0.1, gamma=0.1, kernel=rbf, score=0.761, total=   0.1s
```

Result of Optimal parameters after tunning

{'C': 0.1, 'gamma': 1, 'kernel': 'rbf'}

SVC(C=0.1, gamma=1)

Found the best estimator using a grid search. Take this grid model to create some predictions using the test set and then create classification reports and confusion matrices

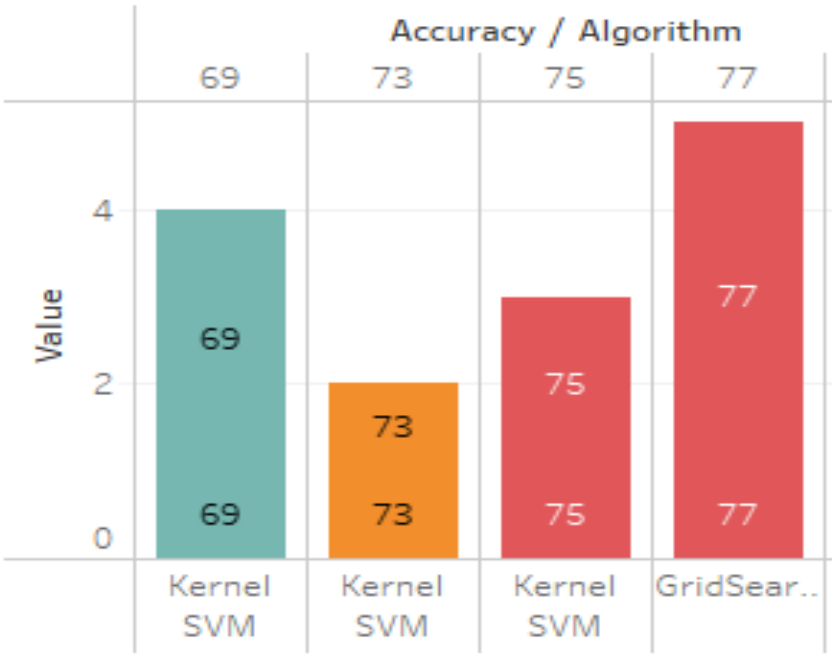
Table 3.6: Accuracy result of GridSearchCV Hyper-parameters Algorithm

ISSN: 2005-4262IJGDC

	precision	recall	f1-score	support
0	0.00	0.00	0.00	119
1	0.77	1.00	0.87	405
accuracy			0.77	524
macro avg	0.39	0.50	0.44	524
weighted avg	0.60	0.77	0.67	524

Table 3.7: Comparison's study of Kernel SVM and GridsearchCV algorithms

S.No.	Algorithm	Kernel Type	Accuracy %
1	Kernel SVM	Poly	73
2	Kernel SVM	Rbf	75
3	Kernel SVM	Sigmoid	69
4	GridSearchCV	Rbf	77



Graph 3.1: Accuracy based on Kernel SVM and

GridSeachCV: Based on Kernel types the GridSearchCV hyper-parameter accuracy level 2% high as compared to default kernel SVM[4][9].

IV. CONCLUSION

ISSN: 2005-4262IJGDC

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 received result, we have changed some parameters in Hyper-Parameter GridSearchCV algorithms, we found it 2% more accurate as compared to the existing kernel SVM algorithm, In the future, we will create our model to find the result more accuracy than this both algorithms.

V. REFERENCES

[1] ItaumaI song Itauma, Hasan Kivrak, and Hatice Kose, "GESTURE IMITATION USING MACHINE LEARNING TECHNIQUES", 978-1-4673-0056-8/12/$26.00 c 2012 IEEE.

[2] Sami S. Alwakee, Bassem Alhalabi,HadiAggoune, Mohammad Alwakeel, " A Machine Learning Based WSN System for Autism Activity Recognition", 2015 IEEE 14th International Conference on Machine Learning and Applications.,

[3] Elias Legesse* and Worku Dechasa,"An assessment of child immunization coverage and its determinants in Sinana District, Southeast Ethiopia",Legesse and Dechasa BMC Pediatrics (2015) 15:31 DOI 10.1186/s12887-015-0345-4.

[4] Mahadev Bag, Dr.Abhishek Badholia, "A Performance Comparison Of Kernel Svm And Hyperparameter Algorithm Using Machine Learning Techniques For Pregnancy Women's", Turkish Journal of Computer and Mathematics Education Vol.12 No.13 (2021), 1217-1221.

[5] Bárbara Badanta-Romero,, Giancarlo Lucchettib, Sergio Barrientos-Trigoa," Access to healthcare among Chinese immigrants living in Seville, Spain", 0213-9111/© 2019 SESPAS. Published by Elsevier Espana, ˜ S.L.U. This is an open-access article under the CC BY-NC-ND license.

[6] Itauma Isong Itauma, Hasan Kivrak, and Hatice Kose," GESTURE IMITATION USING MACHINE LEARNING TECHNIQUES", 978-1-4673-0056-8/12/$26.00 c 2012 IEEE

[7] Elaine Short, David Feil-Seifer, Maja Mataric´," A Comparison of Machine Learning Techniques for Modeling Human-Robot Interaction with Children with Autism", HRI'11, March 6–9, 2011, Lausanne, Switzerland. ACM 978-1-4503-0561-7/11/03

[8] Sami S. Alwakeel, Bassem Alhalabi, Hadi Aggoune, Mohammad Alwakeel, "A Machine Learning Based WSN System for Autism Activity Recognition", 2015 IEEE 14th International Conference on Machine Learning and Applications

[9] Jainy Sachdeva,Vinod Kumar, Indra Gupta, Niranjan Khandelwal, Chirag Kamal Ahuja," Multiclass Brain Tumor Classification using GA-SVM", 978-0-7695-4593-6/11 $26.00 © 2011 IEEE DOI 10.1109/DeSE.2011.31

[10] Shang-Ming Zhou, Ronan A. Lyons, Owen Bodger, Joanne C, "SVM with Entropy Regularization and Particle Swarm Optimization for Identifying Children's Health and Socioeconomic Determinants of Education Attainments Using Linked Datasets", 978-1-4244-8126-2/10/$26.00 ©2010

ISSN: 2005-4262IJGDC