May-June 2020
ISSN: 0193-4120 Page No. 4091-4095

Analysis of Iris Flower Event Logs using Machine Learning Techniques

Saravanan. M.S.[1], Velu. C.M.[2], S. MuthuVijaya Pandian[3], Vinoth Kumar. P[4]

[1, 2]Department of Data Science and Computational Intelligence, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
[3]Department of EEE, SNS College of Engineering, Coimbatore, India
[4]Research Scholar, Department of IT
Veltech Hightech Dr.Rangarajan Dr.Sagunthala Engineering College, Chennai, India
[1]saranenadu@gmail.com, [2]cmvelu41@gmail.com, [3]ajay_vijay@rediffmail.com, [4]vinothkumarme@gmail.com

Article Info
Volume 83
Page Number: 4091-4095
Publication Issue:
May - June 2020

Abstract

The iris flower is the type of show flower, which is normally named after the scientific name. It is used in horticulture. Iris is widely grown as decorative plant in house and gardens. Iris grow well in most any garden soil types providing they are well-drained. First the British statistician and Biologist Ronald Fisher introduced the Iris flower dataset and defined it has multivariate feature in the year 1936 in his research article. At the same time the great scientist Edger Anderson composed the data to quantify the morphologic variation of Iris flowers from three different species, so it is called as Anderson's Iris dataset. The iris flower has the multiple metrics with taxonomic problems. The Iris dataset collected from three species such as Iris setosa, Iris virginica and Iris versicolor by each having fifty samples. The Iris flower measured by four parameters, they are length and width of the sepals and petals in centimeters. The Iris dataset used in this research proposal as a test case for many machine learning classification techniques such as k-nearest neighbor, logistic regression and support vector machine, etc. But this paper identifies the various species among 150 samples of the dataset to find the accuracy of classification using k-nearest neighbor and logistic regression machine learning techniques with different methods of parameter substitution.

Keywords**:** *Iris flower, multivariate, species, classification, k-nearest neighbor, logistic regression.*

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

1. Introduction

The iris flower classification is one of the major research in the recent studies, according to the biological research study it has major focus among various plant species [1]. The first scientific research on Iris flower started during the year 1936 by the biologist Ronald Fisher, also his research introduced the feature multivariate analysis on Iris dataset. The analysis of linear discriminant is an example for multivariate feature. While preparing the Iris dataset, it was prepared from two pasture called Gaspe Penisula for collecting three species on the same day at the same time by the same person with same apparatuss [2].

The collected dataset consists of 50 samples from the three species, first one called Iris setosa, second one Iris virginica and the last one Iris versicolor. From the each species four features were extracted, they are sepel length, sepal width, petal length and petal width. All these

May-June 2020
ISSN: 0193-4120 Page No. 4091-4095

four parameters are measured in centimeters [3]. This research proposal find the type of the species by having these four features of sepal length, sepal width, petal length and petal width. To predict accurate result, the dataset applied on k-nearest neighbor and logistic regression techniques by changing the parameters of algorithms.

2. Related Work

In the year 1959, the Arthur Samuel said that the computers can learn itself without human intervention using machine learning techniques. The artificial intelligence is the seed for machine learning techniques. In the artificial intelligence study the pattern recognition and computational intelligence are the branch, which explores the nature of predictions and providing the model for building the data driven approaches with various dataset inputs. The machine learning has range of computing processes with multiple program and visualization methods, applications such as web application mining, email filtering, computer vision and image processing etc.

Programs learn itself using machine learning algorithms will change the nature of the data and predict the result using various types of techniques. In most of the research case studies the predictive modelling and statistical modelling are used [4]. In machine learning methods, there are two types of processing is possible. First one called supervised learning, which always work with test set and training set to predict the target values. But on the other hand the second one unsupervised learning had only the test labels to predict the target values [5].

The supervised learning method useful to knowledge discovery to identify the better decision making and also mostly used in many areas of research. The computer science based research studies are used to predict optimum decision making according to the learning model of the machine learning algorithms [6]. The recent technologies predict and analyze the new types of information such smart agriculture systems and spatial data analytics and so on.

But in this research study, we are identifying and predicting the learning model and for the given iris flower dataset. The special types of algorithms find the better results with less error rate. Supervised learning is mainly used to reduce the error rate. For predicting the better pattern recognitions the supervised learning helpful to find the frequent items to predict the efficient behavior [7]. Therefore the iris dataset have better classification using pattern mining algorithms. The iris dataset has many classifier recognitions for each set of samples. Each classifier having its own learning experiences, with neural network learning algorithm and a decision tree.

Unsupervised learning algorithm has many disadvantages due to unstructured data classification. On the other hand, it will aim to classify the unlabeled iris datasets. The automated learning of computer will be achieved by supervised learning. The computer will train itself, no need to train anyone. The analysis can be done through samples. The computer will not able to find the best results using unsupervised learning [8]. The samples used to apply the classification in advance. The computer will learn itself using many classification algorithms using various algorithms [9] [10]

The figure 1, shown the various machine learning methods and algorithms [11].

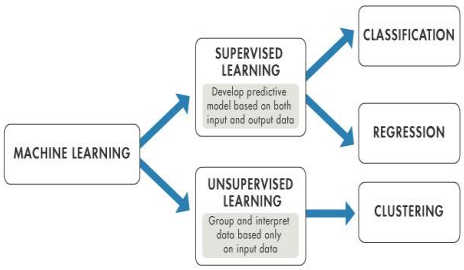

Figure 1: Block Diagram of Machine Learning Methods and Algorithms

3. Iris Flower Data Sets

The first Iris dataset was introduced in the research paper "The use of multiple measurement in taxonomic problems' during the year 1936 by R.A. Fisher. Then after the introduction of machine learning methods the Patrick S. Hoey analysed the Iris dataset via two distinct statistical methods [12]. First plotted using statistical method using scatter plot visualization technique to predict the accurate patterns to find the type of species on Iris dataset. He first tested using Java program, then later in recent days everyone using python with value added features. The concrete prediction gives the different statistical information from the Iris dataset. The Iris dataset scatter plot is done with 150 instances and yield the result as shown in figure 2. This figure has the various features in x-axis and target variable in y-axis.

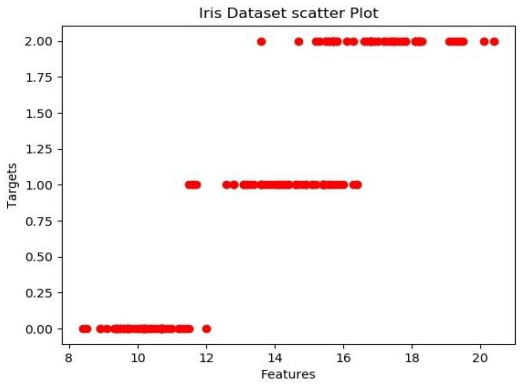

Figure 2: Iris Dataset scatter plot with various features and target variables.

May-June 2020
ISSN: 0193-4120 Page No. 4091-4095

4. Illustration of K-Nearest Neighbour Algorithm

The machine learning algorithms are used to learn the dataset and predict the result for the given operations to attain the objective. In this research paper, k-Nearest neighbor algorithm and Logistic regression taken to learn from the Iris dataset to predict the knowledge for obtaining the species of the Iris flower.

4.1. K-Nearest Neighbors Algorithm

This algorithms is in short, we called k-Nearest neighbor algorithm [13]. The classification will be done through this algorithms using Iris flower. The dataset used for this training method having the k-most similar even logs. The similar logs or instances is predicted using training set. The similarity between the data will be identified using various real values data. The Euclidean distance used in this kNN algorithm. The categorical labels or binary data using the hamming distance. The regression algorithm used the average value of predicted attribute.

4.1.2 How does k-Nearest Neighbors work

The algorithm can be used for even log based approach and competitive learning and other algorithms for the kNN algorithm. The extreme cases of even log based methods will train the observations are the part of the kNN model [14].

The prediction algorithms such as lazy learning will be very much useful to the unseen data, called a localized model. This model also used for the costly computational methods and similar searching algorithms using training datasets. Therefore, we conclude that the kNN is the only best algorithm will give the best results for calculating the values between two event logs.

4.1.3 How do we choose the factor "K"

The "K" value will dominate this k-nearest algorithm, in the sample training set, we have six training samples are having constant values for the taken k value. So the boundaries or edges of each class with separate RC and GS.

The another boundary value of the dataset will contain the two cases of values, which contain the smooth and correct edges, also it will create the final values and will depend on the most frequent items. Also it will increase the "K" value. In this same method the "K" value will be increased for the given values and at the same time the boundary will decide the training errors. The error rate of the dataset depends on the validation parameters at the same time the "K" value may be used for the measuring idea of the two arguments. At the same time if K=1, it always bring zero to the training set. The curve value for this training set will help to find the validation error, if the value is similar to the choice of K with the value 1. If K-1, the optimum value of the training sample will give the better result and the predictions will yield the better curve plot.

4.2. Modeling

The python programming language has many packages to implement the learning algorithms, particularly the Scikit-learn package is useful for the four step Modeling Pattern.

Stage 1:

- ✓ First we have to import the required packages to implement the kNN algorithm.
- ✓ The K Neighbors Classifier used to find the distance from the predefined number of closest in event logs of the predicted the label from these type of dataset.
- ✓ The sklearn algorithm used to find the new point. The logistic regression method is imported used the sklearn packages.
- ✓ At the same time the lasso and Bayesian and coordinates used for the stochastic algorithms.

Stage 2:

- ✓ The estimator, is used to find the sklearn algorithm.
- ✓ This algorithm is used to find the Estimator, to find the fit the model using sklearn algorithms.
- ✓ At the same time the new data will be used to find the better predicted value from the dataset properties.
- ✓ KNeighbors Classifer class is best to predict the object used for the instantaneous values at the same time with the use of classifier of the regressor.
- ✓ The estimate will be calculated with the creation of many identified class attributes of the system predicted variables and this can be used for the future predicted values. At the same time the estimator will help the nest stage of the process.
- ✓ The predicted values are used for the parameters used for the willing the object used for the parameters such as KNeighborClassifier with the feature of auto and size of the leaf is thirty, and the metric used for this algorithm is minkowski, also the metric_parameters will be set none.
- ✓ Also fix the job equal to 1 and the neighbors value will be fixed as five and the weight of the algorithms is uniform.
- ✓ The class value will be fixed as 'ovr' and job number is one. The state of the warm_start will be fixed as 'false'. The k-Nearest algorithm can be taken to the next level with the verbose value is zero.

Stage 3:

- ✓ To find the fitness of the model, follow the rest of the process.
- ✓ This KNeighborsClassifier algorithm is used to find the x and y values at the same time the the logistic regression of the method with fitness of the data can be training of the fitness.
- ✓ The fitness of the value can be found with the training dataset, also the evaluation of the process can be fixed with evaluation data.
- ✓ This data can be analyzed with the evaluation data and monitoring the general identified properties of

May-June 2020
ISSN: 0193-4120 Page No. 4091-4095

feasible data. The training data will be very much analyzed with the unknown examples.

Stage 4:

- ✓ Finding the observation predictions.
- ✓ This is the last step to find the response value.
- ✓ The observation will help the algorithm and the k-Nearest neighbor algorithm can be used for the analyzed values at the same time the data structure used for this model will help the final analysis for objects, which is used for the dataset preparation at the same time the unknown values also identified using this supervised learning algorithm.
- ✓ This supervised algorithm will help the predicted values for the better analyzing the nearest matching values available in the training set.
- ✓ The similar values are grouped under this step for finding the response values.

5. Iris Data Set Implementation

The identification of species applied on testing using various algorithms are shown the following table 1.

Table 1: Species Prediction from Iris Dataset by using various Machine Learning algorithms and Corresponding results with accuracy

Name of the Machine Learning Algorithm	**Parameters**	**Accuracy**
k-Nearest Neighbour	N =5	96.66%
k-Nearest Neighbour	N= 1	100%
k-Nearest Neighbour	Train and Split Method and N=1	96%
k-Nearest Neighbour	Train and Split Method and N=5	95%
Logistic Regression	Train and Split Method and sample=542	95%
Logistic Regression	Train and Split Method and sample = 945	96.66%

The figure 3 and 4 shows the Species Prediction from Iris Dataset by using k-Nearest Neighbor machine learning algorithm.

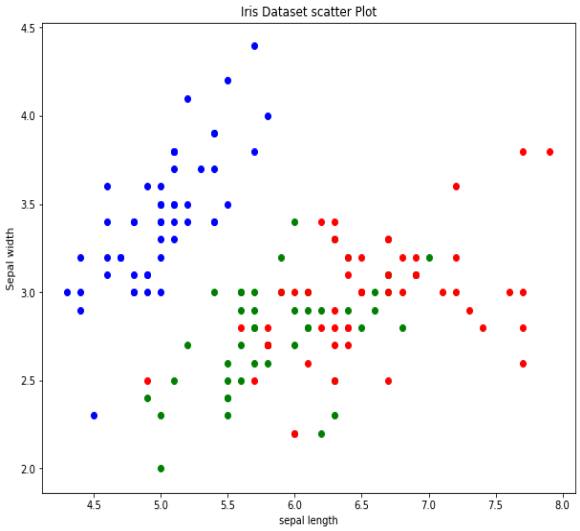

Figure 3: Species Prediction from Iris Dataset by using k-Nearest Neighbor machine learning algorithm, stage-1

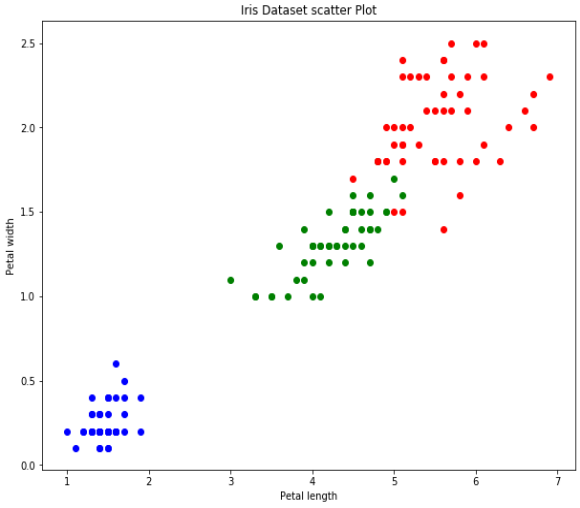

Figure 4: Species Prediction from Iris Dataset by using k-Nearest Neighbor machine learning algorithm, stage-2

6. Conclusion

In day-to-day the iris flower research says that the species identification against the various parameter of sepal and petal features. In this research article, the two machine learning algorithms logistic regression and k-nearest neighbor. The logistic regression requires some training and learns from linear classifier but on the other hand k-nearest neighbor can learn non-linear boundaries with predicted values. Hence the both the studies introduce different results to predict the species features. In addition to that the frequent pattern features of iris flower will give better understanding with learning to classify the flowers. The supervised method used in this study, in future unsupervised method also can be applied to predict unknown labels of the dataset.

May-June 2020
ISSN: 0193-4120 Page No. 4091-4095

References

[1] Khanaa V., & Thooyamani K.P. (2013). Using triangular shaped stepped impedance resonators design of compact microstrip quad-band, Middle - East Journal of Scientific Research, 18(12),1842-1844. 101.

[2] Saravanan. M.S, "Breast Cancer Classification Using Visual Data Mining Techniques", Published in the International Journal of Advanced Science and Technology by SERSC Publications, India, Vol. 28, No. 13, December 2019, pp. 188-197, ISSN: 2005-4238.

[3] Asiri S., Sertkol M., Güngüneş H., Amir M., Manikandan A., Ercan I., & Baykal A. (2018). The Temperature Effect on Magnetic Properties of NiFe 2 O 4 Nanoparticles. Journal of Inorganic and Organometallic Polymers and Materials, 28(4), 1587-1597.

[4] Thaya, R., Malaikozhundan, B., Vijayakumar, S., Sivakamavalli, J., Jeyasekar, R., Shanthi, S.,Vaseeharan B., Ramasamy P., & Sonawane, A. (2016). Chitosan coated Ag/Zn Onanocomposite and their antibiofilm, antifungal and cytotoxic effects on murine macrophages. Microbial pathogenesis, 100, 124-132.

[5] Saravanan. M.S, H. Selvi. "A Study of dyslexia using different machine learning algorithm with data mining techniques", Published in International Journal of Engineering & Technology, UAE, Volume 7, Issue 4, December 2018, pp. 3406-3411, ISSN 2227-524X, doi: 10.14419/ijet.v7i4.14545.

[6] Kolanthai, E., Ganesan, K., Epple, M., & Kalkura, S.N. (2016). Synthesis of nanosized hydroxyapatite/agarose powders for bone filler and drug delivery application. Materials Today Communications, 8, 31-40. 97.

[7] Saravanan. M.S, R. Rajpriya. "Automatic Classification of Cervical Cancer using Support Vector Machine", Published in Journal of Advanced Research in Dynamical & Control Systems, Volume 11, Special Issue 4, May 2019, pp.744-752, ISSN 1943-023X.

[8] Thilagavathi, P., Manikandan, A., Sujatha, S., Jaganathan, S.K., & Arul Antony, S. (2016). Sol–Gel Synthesis and Characterization Studies of NiMoO4 Nanostructures for Photocatalytic Degradation of Methylene Blue Dye. Nanoscience and Nanotechnology Letters, 8(5), 438-443.

[9] Thamotharan C., Prabhakar S., Vanangamudi S., & Anbazhagan R. (2014). Anti-lock braking system in two wheelers. Middle - East Journal of Scientific Research, 20(12), 2274-2278.93.

[10] Saravanan. M.S, R. Rajpriya. "Feature Extraction of Cervical Cancerous and Non-Cancerous Cells based on Edges, Regions and Color-Intensity using Fuzzy Logic", Published in Journal of Theoretical and Applied Information Technology, Pakistan, Volume 97, Number 4, February 2019, pp. 1379-1390, ISSN: 1992-8645.

[11] Saravanan. M.S, H. Selvi. "Using Machine Learning Techniques for Predicting and Diagnosing of Dyslexia in School Age Children", Published in Journal of Advanced Research in Dynamical & Control Systems, Volume 11, Special Issue 4, May 2019, pp.753-760, ISSN: 1943-023X.

[12] Daugman, J, How iris recognition works. Circuits Syst Video TechnolIEEETrans 14(1):21–30, 2014.

[13] X. Brédart, "Bankruptcy Prediction Model Using Neural Networks," Accounting and Finance Research, vol. 3, no. 2, pp. 124128, 2014.

[14] M. A. Doori and B. Beyrouti, "Credit Scoring Model Based on Back Propagation Neural Network Using Various Activation andError Function," IJCSNS International Journal of Computer Science and Network Security, vol. 14, no. 3, pp. 16-24, 2014.

[15] D. Dutta, A. Roy and K. Choudhury, "Training Artificial Neural Network using Particle Swarm Optimization Algorithm," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 3, pp. 430-434, 2013.