Iris Classification with Supervised ML using Algorithm of KNN in JavaScript

Danish Ul Hassan[1], Abid Qadir[1], Basit Hassan[1]

1Sindh Madressatul Islam University/Computer Science, Karachi, 75800, Pakistan

Email: danishulhassan7@gmail.com, abidumrani247@gmail.com, basithq@smiu.edu.pk

Corresponding Author: Danish Ul Hassan, danishulhassan7@gmail.com

Abstract: The Machine learning is what we have to predict the data of unseen nature and gives us predicted results. There are number different machine learning algorithms for prediction and here we will use the supervised machine learning algorithm k-nearest neighbors (KNN) which is used for both the classification and regression analysis problems. In this paper we are predicting the Iris. Iris species are of three types and here we are going to predict that using k-nearest neighbors (KNN) algorithm model with the help of JavaScript.

Key Words: Iris, KNN, Classification, Supervised, Data, Machine Learning, Algorithm

1. Introduction and Literature Review

Taneja, Shweta, Charu Gupta, Kratika Goyal, and Dharna Gureja [1] have applied KNN algorithm. KNN is one best way to recognize the pattern of cancer diagnosis text classification and help out in other categories. KNN algorithm quick learning base way. For testing of KNN algorithm taking the datasets from the UCI machine learning repository this dataset is condense of 150 instances. Each column is characterized by attributes sepal length sepal width and petal length and petal width and classified as either "Iris-Setosa" or "Iris-versicolor" or "Iris-virginica" classes. We select 141 instances randomly as the training instances to calculate the attribute weight sets and then classify the rest instances in the set.

Thirunavukkarasu, K., Ajay S. Singh, Prakhar Rai, and Sachin Gupta [2] has discussed the device getting to know is set prediction on unseen data or testing facts. In system getting to know a computer first learn to carry out a challenge with the aid of training dataset. Then the laptop performs the identical challenge with the testing statistics. The dataset includes one hundred fifty samples of information that has 3 Classes every include 50 samples. To train the system we split the dataset into two components schooling and testing dataset, then the system will train with the aid of training dataset after which it's going to check on trying out dataset. Now we can examine the model weather it apprehends the iris species as it should be or not we have taken the iris dataset from Scikit learn (machine learning library) in which iris dataset is already inbuilt. We used Anaconda software (Jupiter Notebook) to build the model. Initially we load the iris dataset from Scikit learn library. On this paper we tried to build a version that is able to apprehend the iris species as it should be on the basis of 3 classes, but some sample provide the misclassified result. Prediction for class0 and calss2 is 100% correct but prediction for class1 is 4% wrong.

Su, Benyu, Reza Malekian, Jingcun Yu, Xihui Feng, and Zhixin Liu [3] have experimentally tested our set of rules in NetBeans ide, using a trendy uci dataset-iris. Experimental results have proved that our proposed algorithm performs higher than traditional KNN algorithm. In future, we will put into effect our proposed set of rules on diverse different trendy uci datasets and additionally include tender computing techniques like fuzzy common sense to healthy the real global scenarios.

Prihandi, Ifan [5] in hi paper use Dataset is the embodiment of statistics in reminiscence that offers a regular relational software model no matter the beginning of the statistics supply. Used to set the query itself to be run by way of using data adapter in the usage of parameters in report generation. The KNN algorithm works with the sample dataset from the dataset the usage of iris. That is to be had through default from sklearn. This dataset carries 3 varieties of flower species together with petal and sepal sizes. This record is csv, which is used as a desk, every row indicates distinct varieties of flower species, whilst the columns display records features, particularly: sepal length, sepal width, petal period, and petal width respectively. There are three types of plant life, namely Setosa, versicolor and Virginia. There are 50 records samples for every sort of hobby. So, if in total there are 50x3 statistics samples = 150 information samples. Correct class of data type.

Iris Classification with Supervised ML using Algorithm of KNN in JavaScript

KNN has numerous parameters that determine the level of type accuracy, particularly the quantity of ok in KNN. Parameters with excessive accuracy, tuning parameters or finding hyperparameter.

Guo, Ying, Zeng-yuan Li, Er-xue Chen, and Xu Zhang [7] Forest type identification is one among most important contents of woodland inventory. The crucial role of faraway sensing within the wooded area mapping in widespread and woodland type identification especially is expressed by using numerous scientific sports Spot records are of excessive hobby for woodland type identification specifically for reasons in their high and medium decision. The data is covered wangqing forestry bureau in jilin province. The area belongs mountain place and the mainly woodland vegetation is conifer-broadleaf wooded area. The experiments had been carried out at the spot5 facts sets in 2007. For the era of schooling and validation facts units, full-size ground truth campaigns have been performed in summer 2007. Regarding the dominant plant life with this location, it is assumed that no crucial modifications are present all through the length of photo acquisition. To certain the proper parameter robotically, the have a look at used go away-one-out pass-validation method to test the precision as ok and distance measured approach were specific and decided on the top-of-the-line version based totally at the accuracy. The result showed the high-quality result become done as okay became 5 and the distance measured approach was Euclidean distance.

2. Dataset

For this research we have taken the dataset from the online website of University of California in CSV format and later convert it into JSON format so it can be used with JavaScript. The dataset we got it contains 150 samples of Iris Data classified in three main classes which are namely Iris-Versicolor, Iris-Setosa and Iris-Vergenica.

We represent each class with a numeric which based on the four attributes which are sepal length, sepal width, petal length and petal width, all the data of these four attributes are in centimeter(cm).

3. Implementation

We use VSCode software to write JavaScript code and with that we install node modules of ml library and csvtoJson library and other dependencies. The data we get is from the University of California in CSV format, we locally download the data and use in our training model.

A	B	C	D	E	F
4.5	2.3	1.3	0.3	Iris-setosa	
4.4	3.2	1.3	0.2	Iris-setosa	
5	3.5	1.6	0.6	Iris-setosa	
5.1	3.8	1.9	0.4	Iris-setosa	
4.8	3	1.4	0.3	Iris-setosa	
5.1	3.8	1.6	0.2	Iris-setosa	
4.6	3.2	1.4	0.2	Iris-setosa	
5.3	3.7	1.5	0.2	Iris-setosa	
5	3.3	1.4	0.2	Iris-setosa	
7	3.2	4.7	1.4	Iris-versicolor	
6.4	3.2	4.5	1.5	Iris-versicolor	
6.9	3.1	4.9	1.5	Iris-versicolor	
5.5	2.3	4	1.3	Iris-versicolor	
6.5	2.8	4.6	1.5	Iris-versicolor	
5.7	2.8	4.5	1.3	Iris-versicolor	
6.3	3.3	4.7	1.6	Iris-versicolor	
4.9	2.4	3.3	1	Iris-versicolor	
6.6	2.9	4.6	1.3	Iris-versicolor	
5.2	2.7	3.9	1.4	Iris-versicolor	
5	2	3.5	1	Iris-versicolor	
5.9	3	4.2	1.5	Iris-versicolor	
6	2.2	4	1	Iris-versicolor	

Figure1. Dataset Classes

Iris Classification with Supervised ML using Algorithm of KNN in JavaScript

The data is stored in a table form which contains the Sepal length, Sepal width, Petal length and Petal width. The total four attributes in numeric form along with a three classes pf Iris species. After analyzing and pre-processing our data we set the classes in numeric form as 0,1,2 for the prediction purposes.

Figure2. Images of Data

4. Iris Species

The above picture is from 'Datacamp' in which the left most picture is of Iris Versicolor, the center picture is Iris Setosa, and the last one is Iris Virginica. In left most picture the detailed sepal and petal are identified, and our data are in of centimeters of sepal, petal lengths and widths.

Table 1: Comparison of models and their word error rates(WER)

Model	Word Error Rate (WER)
Supervised	2.1
Semi-Supervised	2.0
Unsupervised (Our Research)	1.8

5. Evaluation

The choice metrics matters a lot when we are working with the Machine Learning algorithms as in our case, we are using KNN based algorithm and the performance is highly influenced due to these metrics. We check and verify our KNN based machine learning supervised model that is working properly or not as we expected.

Figure3: Type-2 results

The classification problems are simplest using Machine learning algorithms and, in our case, we try to predict the class of Iris-Flowers using JavaScript to check the accuracy of the KNN algorithm. The results of accuracy we get is 94.4% which is quite good in a quite smaller dataset.

Figure4: Type-0 results

6. Conclusion

From the dataset that we have we train the model using KNN algorithm which ultimately recognizes the species of Iris Flower the class 0 has an accuracy of 100% and class 2 also has an accuracy of 100% but class 1 has some wrong prediction which is 5.6%.

References

[1] Taneja, Shweta, Charu Gupta, Kratika Goyal, and Dharna Gureja. "An enhanced k-nearest neighbor algorithm using information gain and clustering." In 2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 325-329. IEEE, 2014.

[2] Thirunavukkarasu, K., Ajay S. Singh, Prakhar Rai, and Sachin Gupta. "Classification of IRIS dataset using classification based KNN algorithm in supervised learning." In 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 1-4. IEEE, 2018.

[3] Su, Benyu, Reza Malekian, Jingcun Yu, Xihui Feng, and Zhixin Liu. "Electrical anisotropic response of water conducted fractured zone in the mining goaf." IEEE Access 4 (2016): 6216-6224.

[4] Abushmmala, Faten F., and Heba A. Abughali. "Color harmony classification using machine learning algorithms: KNN and SVM." In 2020 International Conference on Promising Electronic Technologies (ICPET), pp. 150-154. IEEE, 2020.

[5] Prihandi, Ifan. "KNN on Iris Data with Python Programming."

[6] Riquelme, Javier A., Ricardo J. Barrientos, Ruber Hernández-García, and Cristóbal A. Navarro. "An exhaustive algorithm based on GPU to process a kNN query." In 2020 39th International Conference of the Chilean Computer Science Society (SCCC), pp. 1-8. IEEE, 2020.

[7] Guo, Ying, Zeng-yuan Li, Er-xue Chen, and Xu Zhang. "The study of parallel KNN in the identification of forest type based on multi-spectral data." In 2011 International Conference on Computer Science and Service System (CSSS), pp. 4113-4115. IEEE, 2011.

[8] SADEQ, SADEER, KHAMIS A. ZIDAN, and JANE J. STEPHAN. "A Review on Iris Recognition System Based Different Classification Techniques." Solid State Technology 64, no. 2 (2021): 5106-5120.