Categorical classification of Fisher's Irish flower using Machine Learning

Pooja R Kotagi a, R H Goudar b, Geetabai S Hukkeri a, Ashwini S Patil c

abcCentre for PG Studies, Visvesvaraya Technological University, Belagavi, India

Abstract: There are many problems that are historically very easy for humans, and very difficult for computers. Machine learning (and deep learning in particular) is currently our best solution for many of those problems. This study investigates that by far Fisher's Iris data set has a dataset with multi varieties with three related species extensively used in pattern recognition literature. K-mean is the unsupervised learning algorithm by which the identification and classification problems will be solved easily. The simplest well known algorithm used here is the clustering or k-mean algorithm. A given dataset is taken with certain number of clusters (K clusters) defining k center for each of the cluster. The set classification is based on examining the patterns of petal and sepal length of Iris Flower.

Keywords: Classification, K-mean clustering algorithm, Iris flowers, Machine learning, Python.

1. INTRODUCTION

In today's tech globalization, the main developing focus areas are Data science, machine learning (ML) and artificial intelligence. Subdivision of computer science is ML . Author S. deduced in 1959 that “Computers have ability to learn without being explicitly programmed”. Machine learning is being used in a verities of computing tasks in which it is explicitly programmed using the well known algorithms to have better performance is hard and also unfeasible due to lack of sufficient data the accurate output of learning is difficult [1]. For the precision is claimed when the machine is prepared without being unequivocally customized or by overfitting. Consequently there is expanded development of learning methods which gives improved exactness model applications include to detect spam email, in detecting system gatecrashers, assuming out in what way to rank and PC visualization [2].

In writing audit, numerous analysts considered on Machine figuring out how to comprehend the learning techniques which is additionally called as calculations for a specific bunch of issue proclamations. Maybe a couple of analysts took a shot at the grouping of ML calculations to distinguish the condition for which a given calculation is suitable. Christopher J. M. portrayed and broke down the estimate of learning calculation with different procedures for progressively and statically choosing weighting. He presumed that the "best" normal cross approval precision is picked as the best classifier for all. issue explanation which are in same area [3].

Steven S. tested to discover the closest learning model for various issue proclamation, for example, Iris blossom groupings, expectation of bosom malignancy and survival times for heart assault patients. The weakness was the anticipated calculation was not totaly exact [4].

Thomas G.T.introduced in a word that why group strategies can outflank any single classifier inside the troupe with a mistake rate underneath 0.5 for developing gathering and audit the three crucial reasons. He likewise referenced that troupe strategy performs much superior to anything other learning models [5].

ML algorithms are structured into organization, depending on the required output of particular problems. General types of algorithms include: Depending on the required outcome of the algorithm, machine learning is structured into four main categories. They are supervised learning, unsupervised learning, semi supervised learning and reinforcement learning. The study of clustering is a branch of Machine Learning in which the grouped data are not been labelled as in supervised learning. Instead of responding to feedback, cluster analysis identifies the similarities in the data and reacts based on the presence or absence of such similarities in each new input of data. The mechanism or working model of unsupervised machine learning is categorized into two phases as shown in Figure 1.
In the underlying phase (Training Phase) the machine is being educated with named input information and in the expectation stage the machine predicts the yield contingent upon the info given. In this examination paper, Using directed learning the Iris bloom classifications are distinguished by giving marked info information to the machine.

2. METHODOLOGY

The Iris flower has three verities namely Iris, setosa, versicolor, and virginica as shown in Figure 2. The primary point of this exploration paper is that the PC ought to be able to total three unique orders of Iris bloom to three classes. The separation is finished utilizing solo learning model in which the unlabeled information is given and by utilization of k-mean grouping calculation the order is finished [7].

In the Clustering method the data items are set into of separate groups, the groups are named clusters. So that the items in a same group are similar compared to the objects in other group. Clustering techniques are applied in variety of application areas namely pattern recognition, data mining, machine learning etc. Clustering algorithms are categorized as Tough, Uncertain, Possibilistic, and Probabilistic [8]. The users do not need to tell the computer which class the Iris belongs to the computer will identify them by itself.

The data set contains three classes of data and with each class is of 50 instances, where each class refers to a type of iris plant. Each class is linearly independent from the other two classes. The character information will include the length and width of both sepal and petal, with same unit, centimeter. Unsupervised learning categorizes the input data into classes depending on integral relationship measure [9].

For pattern recognition, K-means is a standard clustering algorithm and here the implementation is done on Python programming language.

Initially the k mean working algorithm randomly selects k centers for k clusters which are called as cluster centroids, then for the nearest centroid data points are assigned. For each cluster the centers are calculated and are updated the position of centroid for single count. The process is repeated till the centroid change their position and become actual centers.

The python code and the result of k-mean clustering is shown in the below figure a and b respectively.

```python
from sklearn.cluster import KMeans

colors = ['blue', 'red', 'green', 'yellow', 'pink', 'orange', 'brown', 'gray', 'white', 'black']

k = 3
kmeans = KMeans(n_clusters=k, random_state=0)
kmeans.fit(X_train)

centroids = kmeans.cluster_centers_
colors = ['red', 'blue', 'green']

fig, ax = plt.subplots(figsize=(10, 6))
ax.scatter(X_train[:, 0], X_train[:, 1], c=kmeans.labels_, s=100, cmap='viridis')
ax.scatter(centroids[:, 0], centroids[:, 1], c=colors, s=200, alpha=0.5)
ax.set_xlim(-10, 10)
ax.set_ylim(-10, 10)
ax.set_xlabel('Feature 1')
ax.set_ylabel('Feature 2')
ax.legend(['Cluster 1', 'Cluster 2', 'Cluster 3'])
plt.show()
```

To find optimal number of cluster k the elbow method is used, it is technique to determine number of cluster and k. From the Figure 3, k value is termed as 3 at the Elbow point of plot the optimal number of cluster is determined. From which it is determined that there are 3 classes of species named versicolor, setosa, virginica.

3. RESULTS AND DISCUSSIONS

In this experiment we present a relative study of classification technique of dataset with an integration of clustering and classification technique of data set using Fisher’s Iris dataset. Fisher’s iris data set, which consists of...
three different species of iris flowers are taken and for evaluating the classification of Iris flower using k-mean clustering the real dataset is appropriately classified. Making use of 50 samples with each of having four features were considered and are separated using k-mean clustering algorithm.

The results of the present work shows that classification technique gives a encouraging result with maximum accuracy rate and toughness among the classification and clustering algorithms. The three different species Iris Setosa, Iris Versicolour, Iris Virginica classification is as shown in Figure 4 with each having its own centroids.

![Fig. 4: The clustered data with 3 centroids](image)

CONCLUSION

In this paper we apply machine learning algorithms to identify and classify different species of Iris flower. As the python is Interpreted, Modular, Dynamic, Object-oriented, Portable, High level, Extensible in C++ & C is used to write the code in this presented work.

The primary goal of unsupervised learning is to predict the output without providing the input as labelled data. Here in this project we make predictions on unseen data which is the data not used to train the machine hence the machine learning model built should accurately predicts the species of future flowers or input unlabelled dataset rather than exactly calculating the label of previously trained data.

REFERENCES


Authors Profile

**Pooja R Kotagi**, currently a M.Tech student at Centre for PG Studies, Visvesvaraya Technological University, Belagavi.

**Dr. R H Goudar**, currently working as an Associate Professor, Dept. of CNE, Visvesvaraya Technological University, Belagavi. He has 14 years of Teaching Experience at Professional Institutes across India. He worked as a faculty at International Institute of Information Technology, Pune for 4 years and at Indian National Satellite Master Control Facility , Hassan, India. He published over 130 papers in International Journals, Book Chapters and Conferences of High Repute. His Subjects of Interest include Semantic Web, Network Security and Wireless Sensor Networks.

**Geetabai S Hukkeri**, currently a Full Time Integrated Ph.D Research scholar at Centre for PG Studies, Visvesvaraya Technological University, Belagavi. She has completed, B. E. in Computer Science and Engineering, M.Tech. in Computer Network Engineering. Her area of Research Interests are Big Data, Machine Learning and Information Retrieval.

**Ashwini S Patil**, currently a M.Tech student at Centre for PG Studies, Visvesvaraya Technological University, Belagavi.