



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

MACHINE LEARNING MODEL

Submitted by

SACHIN GUPTA

(1613105094)

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING WITH SPECIALIZATION OF

BUSINESS ANALYTICS OPTIMIZATION

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

Ms. VAJETNI MALA



SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report “MACHINE LEARNING MODEL” is the bonafide work of “SACHIN GUPTA (1613105094)” who carried out the project work under my supervision.

SIGNATURE OF HEAD

**Dr. MUNISH SHABARWAL,
PhD (Management), PhD (CS)
Professor & Dean,
School of Computing Science &
Engineering**

SIGNATURE OF SUPERVISOR

**Ms. VAJETNI MALA
School of Computing Science
& Engineering**

Table of Contents

1. Abstract
2. Introduction
 - (i) Overall description
 - (ii) Purpose
 - (iii) Motivations and scope
3. Problem Statement
4. Proposed Model
5. Preliminary / Existing System
6. Implementation
7. Output
8. Conclusion
9. References

Abstract:-Machine learning is about prediction on unseen data. There are three types of machine learning are called as Supervised, Unsupervised and Reinforcement learning. In this paper we have worked on supervised learning. We have taken the iris dataset and used K-nearest neighbor (KNN) classification algorithm. Our purpose is to build a model that is able to automatically recognize the iris species. Tools used for this in paper are Numpy, Pandas, Matplotlib and machine learning library scikit-learn.

Introduction:

Overall description- Machine learning is a category of algorithm that allow software applications to become more accurate in predicting outcomes without being explicitly programmed. In machine learning a computer first learn to perform a task by training dataset. Then the computer perform the same task with the testing data . Basically supervised learning is learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with correct answer. After that machine is provided with new set of data so that supervised learning algorithm analysis the training data and produces an correct outcome from labeled data.

The implementation of the model includes six basic steps of machine learning that are:

1. Collect data/prepare data
2. Choose algorithm
3. Creating object of the model

4. Train the model by training dataset
- 5 .Making prediction on unseen data or testing data
6. Evaluation of the model.

Purpose-: In this very model we are supposed to create the model that is able to classify the iris data set based on their feature.

Motivations and scope-: This model will work on the dataset and based on the features of the data set it will classify it into different category which will help in classification and also prediction of different classes.

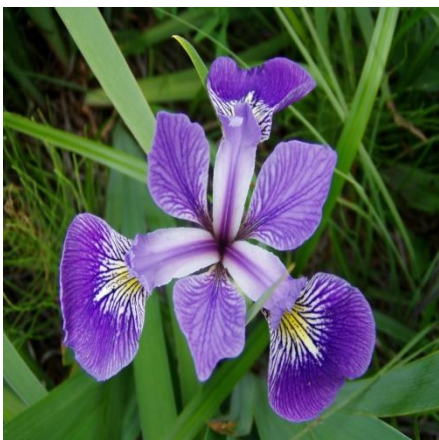
Problem Statement-: We have iris dataset where we have given 150 data set of three category. Every category has 50 data each. Now we have to build the model so that it can classify the data set based on its category or classes.

The three types of data are-

Setosa



Versicolour



Virginica



Proposed Model-:

KNN Classifier → K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition.

Preliminary / Existing System-:

As our intention is to design a model that is able to automatically recognize the iris species accurately. So for that we collected/prepared data which involve data preprocessing and splitting of data. Data preprocessing involve handling of missing data, handle of categorical data and handling of feature scaling. Categorical data involves nominal data and ordinal data which can be handle by pandas as well as machine learning and for handling missing data and feature scaling we use pandas and machine learning respectively. Splitting of dataset involves training data and testing data. We shuffle the data so that there is no any particular sequence in training as well as testing dataset. K- Nearest

Neighbors is the simplest supervised machine learning algorithm that classifies a data point based on how its neighbors are classified. But when it was done with decision tree the structure was too complex and it was difficult to understand how to do the further processing easily.

Implementation-:

I used Anaconda software (Jupyter Notebook) to build the model. Initially we load the iris dataset from Scikit learn library.

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
5	5.4	3.9	1.7	0.4
6	4.6	3.4	1.4	0.3
7	5.0	3.4	1.5	0.2
8	4.4	2.9	1.4	0.2
9	4.9	3.1	1.5	0.1
10	5.4	3.7	1.5	0.2
11	4.8	3.4	1.6	0.2
12	4.8	3.0	1.4	0.1
13	4.3	3.0	1.1	0.1
14	5.8	4.0	1.2	0.2
15	5.7	4.4	1.5	0.4
16	5.4	3.9	1.3	0.4

Above table shows the attributes of the dataset that is Sepal length, Sepal width, Petal length and Petal width. A dataset

contain value of all attribute. As the dataset is already preprocessed so we don't need to do data preprocessing. Now we decide target variable that is 0,1,2 .

Now we shuffle the input and output data and after shuffling the data we split the data into training and testing data. Where training data must be greater than testing data and training data must include all three class of data training data contains 120 rows and testing data contains 30 rows. We need four variable training input, training output testing input and testing output. Now we import model/algorithm K-Neighbors Classifier from SK learn Library and create an object for the KNN classifier. Now with the help of training data we train the model using fit method, after training the model we make prediction on unseen data that is testing data.

```
1 pred=Knn.predict(x_test)
```

```
pred
```

```
1 pred
```

```
array([1, 2, 0, 1, 0, 1, 2, 1, 0, 1, 1, 2, 1, 0, 0, 2, 1, 0, 0, 0, 2, 2,  
       2, 0, 1, 0, 1, 1, 1, 2])
```

```
1 pred[2]
```

```
0
```

```
1 y_test[2]
```

```
0
```

Output-:

```
1 from sklearn.metrics import accuracy_score
```

```
1 accuracy_score(y_test,pred)
```

```
0.9666666666666667
```

```
1 Knn.score(x_test,y_test)
```

```
0.9666666666666667
```

```
1 print('training accuracy :',Knn.score(x_train,y_train))
```

```
training accuracy : 0.975
```

```
1 print('testing accuracy:',Knn.score(x_test,y_test))
```

```
testing accuracy: 0.9666666666666667
```

Conclusion:-

In this paper we tried to build a model that is able to recognize the iris species accurately 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.

Reference:-

1.Louridas, P., & Ebert, C. (2016). Machine Learning. IEEE Software, 33(5), 110–115.

doi:10.1109/ms.2016.114

2. <https://www.geeksforgeeks.org/supervised-unsupervised-learning/>

3.<https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7>

4.<https://www.geeksforgeeks.org/k-nearest-neighbours/>

5. Kanu Patel¹, Jay Vala², Jaymit Pandya³ ¹Assist. Prof, I.T Department, BVM Engineering College, V.V.Nagar kanu.patel@bvmengineering.ac.in ²Assist. Prof., I.T Department, GCET Engineering College, V.V.Nagar, jayvala@gcet.ac.in ³Assist. Prof, I.T Department, GCET Engineering College, V.V.Nagar, jaymitpandya@gcet.ac.in