I J R B A T, Issue (IX), Vol. II, May 2021: 67-73

A Double-Blind Peer Reviewed & Refereed Journal

OPEN ACCES

e-ISSN 2347 – 517X

Original Article

INTERNATIONAL JOURNAL OF RESEARCHES IN BIOSCIENCES, AGRICULTURE AND TECHNOLOGY

© www.ijrbat.in

FLOWER RECOGNITION USING MACHINE LEARNING

Sunil Bhutada, K.Tejaswi and S.Vineela

Department of Information Technology, Sreenidhi Institute of Science and Technology, Ghatkesar,Telangana-501301

sunilb@sreenidhi.edu.in, svineela1234@gmail.com

Communicated: 15.03.21 | Revision : 21.04.2021 Accepted: 11.05.2021 | Published: 30.05.2021

ABSTRACT:

Classification and Recognition systems are being used to use the existing data effectively and getting the feature extraction from the data and displays them. This project we want to flower classification using text recognition. Actually the dimensions of flowers are considered and the flower is mentioned in the dataset including the dimensions. The goal of this paper is to extract the features of flowers and display the advantages and disadvantages of the flower. According to the performance of few algorithms we considered KNN and Random Forest as it gave the maximum accuracy.

Keywords: Text Recognition, Flower Classification, Feature Extraction, Algorithm and Accuracy.

INTRODUCTION:

Flower Classification is the first part of this project inorder to classify all the flowers based on the sepal length,sepal width,petal length and petal width of the flower.In today's world there are many flora and fauna around us.The problem is people not recognizing the importance of them in our daily life for many uses.There are many ayurvedic plants,medicinal flowers and healthy fruits are around .This project is helpful for many by considering flowers for not only decorations but also for the treatments.Not only medicinal purposes,even the fragrance is also used in many therauptic usages.

We developed the project in such a way that when we want to find out all the details of a flower those are all displayed with the help of existing dataset.The advanatages and disadvantages are also displayed so that it would be more effective.We have performed experiment using Iris dataset.It consists of three different species with almost 150 tuples useful for flower classification and finding the accuracy.

Machine learning is being used in here which plays a vital role in classification.We got considerable accuracy and results using two different algorithms Knn algorithm and random forest. Machine learning is being used in here which plays a vital role in classification.

LITERATURE REVIEW

Sentiment depicts an emotion of a human which can be positive, negative, or neutral and the reviews are a collection of data that is sentiment rich will aid the product manufacturer to make satisfactory decisions for the growth of the company. Websites like Amazon and Trivago analyze the reviews based on Sentiment Analysis to build a large customer base. This result or outcome is used by the customer to make the right decisions while purchasing the product based on different feedback.Most of the data on online websites is unstructured and it is further cleaned and analyzed to develop or identify the opinion based on polarity. Author Zeenia Singhla has stated in the paper that algorithms like Support Vector Machine are used for processing. A combination of Product reviews is collected based on volume, variety, and velocity for

the understanding of People's perspective to detect the mindset and behavior of them on a particular product.The author Meena Belwal proposes the idea of collaborative data for the same product available on different websites and can be utilized by the manufacturer for the enrichment of Business Intelligence and proper decision making. The results are visualized using graphs for better understanding and get a quick idea of the entire feedback.

I. CONFIGURATION

Flow Chart: a) Data Assortment: Unstructured data in the form of reviews have been collected from Nykaa.com.

b) Data Preprocessing: Preprocessing steps such as cleaning, Integrating, Transformation, Reduction are performed.

c)Descriptive Statistics: Feature, Selection, Feature extraction, and Visualization of the outcome are done.

d)Lexical Based Opinion Mining: It is performed to generate the polarity of different brands of lipsticks and then visualize.

***Descriptive Statistics*:** Descriptive statistics focus on the gathering of unorganized data to decipher knowledge. This is the process of describing the previous data and analyzing the important features. The numeric data from the analysis is pictured in the form of word clouds and charts. The preprocessing steps play a vital role in descriptive Analysis which are as follows. All the raw data is integrated into a text file. From this consolidated data, the delimiters are identified to avoid inconsistency to clean the data. Then, separators are taken into consideration to split the sentences into tokens. The frequency of each token is calculated to transform the data. The generated frequency is sorted for reduction and selection. The selected tokens and their respective frequencies are used to visualize.

The package being for performing Descriptive Analytics is Syuzhet. The methods used from this package for Analysis are get_text_as_string( )and get tokens( ). The get_text_as_string() method leads the input file which takes the path of the file as the argument. Then the get_tokens( ) method is used to split the sentences into words. The frequency of the generated tokens is calculated by as.integers() method. The tokens with the lowest frequency are eliminated. The organized data is represented as a bar graph and word cloud. This can be done theoretically by using Chi-Square Test to categorize the data based on the frequency distribution. In this method, the number of features and the number of reviews is compared with the expected values in each category to know the association. The Chi-Square Test is considered significant when there is a large difference between the observed and expected values. The difference varies depending on the size of the sample data.

Algorithm for calculating word frequency:

Step-1: importing syuzhet package

```
library(syuzhet)
```

Step-2: to get the words from text

```
Word file <- get_text_as_string("path ")
```

Step-3: Count the repetition of words

```
wordcount <- get_tokens(wordfile, pattern = "\\W")
```

Step-4: Get the frequency of each repeated word

```
syuzhet_vector <- get_sentiment(wordcount, method="syuzhet")
```

Step-5: Extract the most frequently repeated words using sentiments

```
words.freq<-table(unlist(wordcount))
s_v <- cbind(names(words.freq),as.integer(words.freq))
```

The process of Descriptive Statistics deals with the behavior of historic data. Hence it helps us acknowledge the crucial features. For this process, the assorted reviews of lipsticks from Nykaa.com of different brands have been used. The resulting graph depicts the features that highly influence the opinion of the consumer.

Lexicon based Opinion Mining

i. Natural Language Processing:

NLP deals with the intelligence of how a computer interprets human emotions. It is used to analyze the text and extract the subjective data from the reviews. One of the major approaches of NLP is POS tagging.

ii. POS Tagging:

To perform the POS tagging, tokenization of sentence is to be carried out which means dividing the text into smaller parts called tokens. Then, the parts of speech tagger are associated with each token depending on the context and sequence. Among different POS Tagging techniques, this research concentrates on the lexical based approach of opinion mining which is an unsupervised learning method.

The view of the consumer is of two types. They can be subjective and Objective. Subjective refers to opinion in form of a sentence, whereas objective refers to the rating of the products in terms of numbers. Lexicon-based opinion Mining concentrates on the subjective part of the view. This approach is the idea of generating the sentiment based on semantics, which indicates the context of the word in the sentence. The need for the semantic orientation is to remove the ambiguity in polarizing a word. Word polarity depicts the orientation of words in a sentence whether they are positive, negative, and neutral. This results in polarity values of the sentence once all the sentences are polarized the outcome is used to analyze the product polarity.

Proposed Algorithm:

Two main packages used for this process are pandas and nltk(Natural language processing tool kit). The preprocessed file is loaded using pandas. The polarity of the sentences is generated using Sentiment Intensity Analyzer.

Step1: Importing the required nltk package and pandas package.

Step2: Create a class with the function Sentiment_Intensity_Analysis to calculate polarity.

Step2.1: Initialize all three polarity variables and count the variable to zero.

Step2.2: Excel file of a dataset is loaded and parsed into a data frame.

Step2.3: Parsed data frame is converted to list for iteration.

Step2.4: Initialize Sentiment Intensity Analyzer object.

Step2.5: Create a loop that finds polarity scores using the polarity_scores method.

Step2.6: Repeat the loop for every sentence in the dataset.

Step2.7: For each iteration the add the polar_scores to the previously initialized polar variable.

Step3: Create an instance of a class and call the function to print polarity values of each s Flower categorization is detailed in here using deep convolutional neural networks.

KNN algorithm is used in order to classify the given dataset,obtained more than 80% accuracy.Text recognition is done in this paper and displayed the text by automatically reading from images.Feature extraction is detailed in here which helped for image recognition process.Approach of video text recognition focussing on category dependent method based on top down approach.Artificial neural network, Naive Bayes and CNN are included for character recognition.Text detection with algorithm detecting the frame having text from live video streaming.

ARCHITECTURE

II. SYSTEM OVERVIEW

The overview of the proposed system for flower recognition system using text recognition is detailed in figure1.The first step is the data classification.The input data field is text data which is a dataset which is stored in the system memory.K-Nearest neighbour is the algorithm used for flower classification.The output is detailed with the flower details and advantages.

FEATURE EXTRACTION

Text feature extraction that extracts text information is an extraction to represent a text message, it is the basis of a large number of text processing. The basic unit of the feature is called text features. Selecting a set of features from some effective ways to reduce the dimension of feature space the purpose of this process is called feature extraction. During feature extraction, uncorrelated or superfluous features will be deleted.

IV. HOW THE SYSTEM WORKS

The flower classification and recognition based on text processing by taking existing flowers and their attributes in order to find their characteristics. The user can enter the flower attributes or just the name and get that the advantages and disadvantages of the flower as soon as we enter the flower details. The usage of the flowers are to be known in order to know the advantage by medicinal and other therauptic uses. The existing flower datasets are all considered and classified based on their attributes using KNN Algorithm and Random forest algorithm. The features are extracted from the text data and based on the feature extraction the flower recognition is done. The Gui used is user-friendly and provide better results. This system can be further improved to yield more accuracy by combining other features, such as numbers of petals and flower texture. More flower species can be added to our system.

KNN ALGORITHM :

In statistics, the k-nearest neighbors algorithm is a non- parametric classification method first developed by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover. It is used for classification and regression. In both cases, the input consists of the k closest training examples in data set.

RANDOM FOREST ALGORITHM:

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean/average prediction of the individual trees.

V. DATASETS

Iris flower data set is a multivariate dataset introduced by Ronald Fisher.

The data set contains 50 samples of each three species of Iris. There are four features: the length and the width of sepal and petal.

The aim is to classify iris flowers among three species (setosa, versicolor, or virginica) from measurements of sepals and petals' length and width. The iris data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant.

VI. EXPERIMENTAL SETUP

The dataset we collected is iris dataset which consists of 150 tuples with four attributes of flowers which includes sepal length, sepal width, petal length and petal width. The three different species of flowers included are setosa, versicolor and virginca.

Training dataset consists of 80% and testing data is 20% of the existing dataset. Accuracy is obtained using two different algorithms, the maximum accuracy obtained is 93.33%.

The first algorithm is Random forest and the second one is Knn Algorithm.

The input is the user interface which is a text field where the user enters the flower name,it then displays the advantages like medicinal values and other advantages of the flowers.

The output on the first page displays the histograms obtained after the classification. Then on entering the input flower name, output is displayed which consists of advantages of flowers.

VII. DESIGNS AND CONCLUSION

In this project, two algorithms are considered for classification and to obtain high accuracy. The two algorithms that are used are KNN algorithm and Random Forest algorithm.

The results obtained after classification is used to choose effective algorithm. The user interface has a text field to enter the flower name, it then displays the medicinal values and other uses of the flower.

The advantage of this project is that we can obtain the uses of flowers which includes therauptic uses and medicinal uses. The classification takes place when the user clicks classify button, so that every time when recognition is done, the classification is also done.

The disadvantage is that only few flowers are included in the dataset. The dataset used is iris dataset which consists only few rules which resulted in limitation of flower dataset.

CONCLUSION

The accuracy we obtained for Knn algorithm is more when compared to Random Forest Algorithm. In future, the scope of using flowers in ayurvedic medicines is more. These flowers are also used in beauty products .New techniques are being designed to use these flowers in Air purifiers also. There are many more advantages of using flowers in different fields, few more are yet to come.

The flower classification and recognition based on text processing by taking existing flowers and their attributes in order to find their characteristics.

The user can enter the flower attributes or just the name and get that the advantages and disadvantages of the flower as soon as we enter the flower details. The usage of the flowers are to be known in order to know the advantage by medicinal and other therauptic uses.

REFERENCES:

Project: Flower Species Segmentation and Classification Viraj Mavani Viraj Mavani Ayesha Gurnani Ayesha GurnaniJhanvi Shah

Flower Recognition System Based on Image Processing Tanakorn Tiay, Pipimphorn Benyaphaichit and Panomkhawn Riyamongkol

A Technical Review on Text Recognition from Images Pratik Madhukar Manwatkar K.

R. SinghK. R. Singh

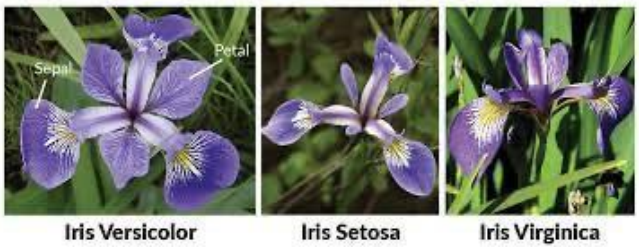
A Review Paper on Face RecognitionTechniques Vijay H Mankar Vijay H Mankar Sujata BheleSujata

R. Sharma, B. Kaushik and NGondhi, "Character Recognition using Machine learning and Deep lear. ning"

S. L. Wasankar, H. Mahajan, D. Deshmukh and Munot, "Machine learning with text recognition", 2010 IEEE International Conference on Computational Intelligence and Computing Research.

Video text recognition using feature compensation as category-dependent feature extraction

M. Mori.



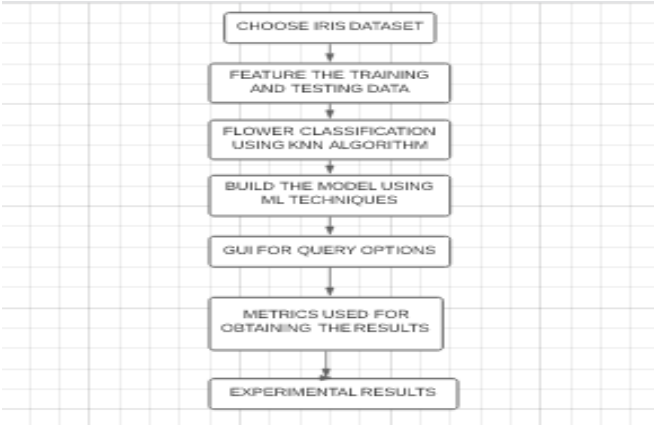
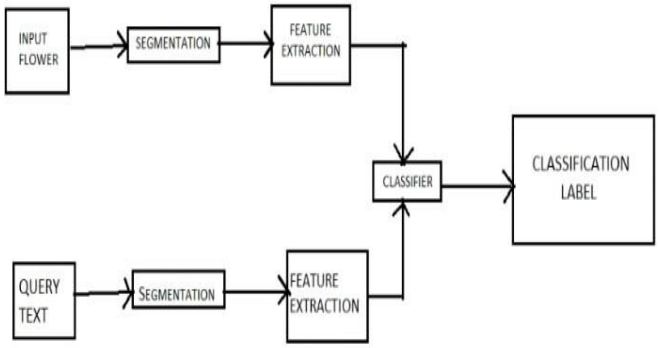


FIGURE: FLOW CHART



	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa

Iris Dataset

```
In [7]: #Import Random Forest Model
        from sklearn.ensemble import RandomForestClassifier

        #Create a Gaussian Classifier
        clf=RandomForestClassifier(n_estimators=100)

        #Train the model using the training sets y_pred=clf.predict(X_test)
        clf.fit(X_train,y_train)

        y_pred=clf.predict(X_test)

In [8]: #Import scikit-learn metrics module for accuracy calculation
        from sklearn import metrics
        # Model Accuracy, how often is the classifier correct?
        print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

        Accuracy: 0.9333333333333333
```

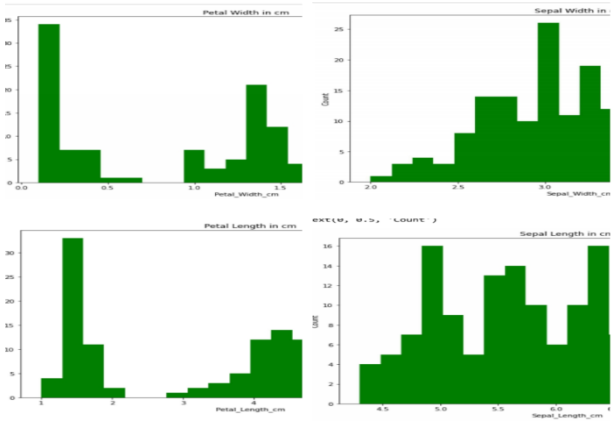


Figure : Histograms of flower attributes in classification