10/8/2022

**Seed Classification Report**

AIDI- 2004- AI in Enterprise Systems

**Member**

Saurav Bisht - 100850875

Submitted to: Prof. Noopa Jagadeesh

**Table of Contents**

[**Introduction** 2](#_Toc111056419)

[**Problem Statement** 2](#_Toc111056420)

[**Dataset Exploration** 2](#_Toc111056421)

[**Algorithm Implementation** 7](#_Toc111056422)

[**User Interface** 9](#_Toc111056423)

[**Flask** 10](#_Toc111056424)

[**Heroku Deployement** 10](#_Toc111056425)

# **Introduction**

This project deals with the most general tasks of machine learning known as classification. Classification is a task that necessitates the application of machine learning algorithms to learn how to assign a class label to problem domain instances. The most basic example of this task is classifying emails “spam” or “not spam”.

The most general example of classification task is classification predictive modeling. This type of classification is referred to a predictive modeling problem where a label is predicted for a given features in terms of input data.

There are total four main types of classification tasks:

1. Binary Classification
2. Multi-Class Classification
3. Multi-Label Classification
4. Imbalanced Classification

The classification being performed on this project is Multi-Class Classification.

# **Problem Statement**

Tom Mitchell, a well-regarded machine learning researcher, proposed precise definition in 1998: **Well posed Learning Problem:** A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

The goal of the project is to determine the category of seed. To our problem there are three type of wheat seeds data (as listed in UCI machine learning repository). The implementation involves use of two types of algorithms that is K-means and Random Forest Classifier. We have 7 attributes from three different types of wheat samples in our dataset. The goal of the investigation is to discover a distinct relationship between the physical characteristics of wheat seeds and the type of wheat.

# **Dataset Exploration**

According to UCI, the dataset comprises of the tested group included kernels from three different wheat varieties: Kama, Rosa, and Canadian, each with 70 elements chosen at random for the experiment. A soft X-ray approach was used to identify high-quality visualization of the interior kernel structure. It is non-destructive and significantly less expensive than more sophisticated imaging techniques like as scanning microscopy or laser technology. KODAK 13x18 cm X-ray plates were used to capture the photographs. Combination harvested wheat grain from experimental fields was studied at the Institute of Agrophysics of the Polish Academy of Sciences in Lublin.

Attribute Information:

To construct the data, seven geometric parameters of wheat kernels were measured:  
1. area A,  
2. perimeter P,  
3. compactness C = 4\*pi\*A/P^2,  
4. length of kernel,  
5. width of kernel,  
6. asymmetry coefficient  
7. length of kernel groove.

The Dataset provided from UCI machine learning repository <http://archive.ics.uci.edu/ml/datasets/seeds>

A picture containing table

Description automatically generated

The dataset comprises of seven features with one target variable.

The type indicates the category of the seed.

A picture containing text

Description automatically generated4

Descriptive Statistics of Data

Table

Description automatically generated

As mentioned above, the dataset comprises of seven feature which have float datatype while the category column that is our target variable is of int-type.

Text, table

Description automatically generated

Since the problem rationale is of classification type thus, checking for the class imbalance and Null values is dataset is most crucial before Training and testing data preparation.

Text

Description automatically generated

Chart, bar chart

Description automatically generated

There is no class imbalance that can be seen from the above plot.

Correlation plot of the dataset

Chart, histogram

Description automatically generated

Plotting the most correlated columns with each other to the type of correlation columns are representing.

Chart, scatter chart

Description automatically generated

Train-Test Preparation

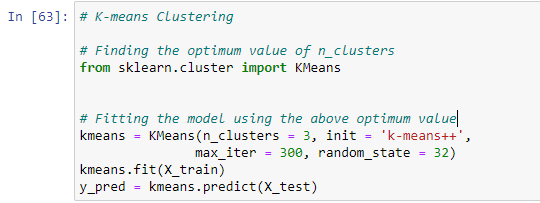
Graphical user interface, text, application, email

Description automatically generated

# **Algorithm Implementation**

K-means Clustering

The fundamental goal of a cluster analysis is to distinguish things that are distinct by placing them in different clusters and grouping objects that are similar in one cluster. The K-Means clustering algorithm is one of the most often used clustering techniques. It divides each object into a predetermined number of clusters that the user specifies (assume K clusters). One random cluster center should be selected for every cluster. It is desirable that these centers be located as widely apart as feasible. According on how close an observation is to the cluster mean, K-Means clustering assigns n observations to one of K clusters. The initialization of the clustering process involves selecting random data points from the dataset. The sum of squares within the cluster yields the closest distance.



Predictions were made using new datapoints

Text

Description automatically generated

The algorithm predicted the category of seed is of type 2. The validity of the prediction was checked using scatter plot with new data point centered around it using circular dimension.

Chart, scatter chart

Description automatically generated

Random Forest Classifier

Graphical user interface, text, application

Description automatically generated

The model achieved the accuracy of 96%. This model will be further used for the deployment.

Text

Description automatically generated

# **User Interface**

The front-end of the model was created using html while adding css styling sheet to it.

Graphical user interface, application

Description automatically generated

# **Flask**

Flask was used to produce the code interfacing with the front-end.

Text

Description automatically generated with medium confidence

# **Heroku Deployment**

Link to deployed model:

wheat-demo.herokuapp.com