Quality Analysis and Classification of Rice Grain Based On K-Nearest Neighbor and Decision Tree

MUHAMMAD JUNAID ASIF

L1S22MSDS0012

MASTERS OF SCIENCES IN DATA SCIENCE

UNIVERSITY OF CENTRAL PUNJAB

A Project Proposal

Submitted to

Saad Azhar Saeed

On the topic

Quality Analysis and Classification of Rice Grain Based On K-Nearest Neighbor and Decision Tree

Submitted By:

Er. Muhammad Junaid

(L1S22MSDS0012)

For partial fulfillment of term project assigned for course

"Tools and Techniques in Data Science"

Offered in Spring 2022

Session: 2022 - 2024

FACULTY OF INFORMATION TECHNOLOGY

UNIVERSITY OF CENTRAL PUNJAB

Abstract:

This project proposal outlines the development of system for classification and quality analysis of different kind of rice grains. Six different type of rice including Jasmine, Gonen, Basmati, Saila, Arborio Rice will be taken for the data analysis and classification.

Data will be extracted by taking 1000 images of each type. Different image processing techniques will be applied for smoothing of images and detection of rice grains. Different kind of pre-processing techniques will be applied for data cleaning, smoothing and integration. Impurities involved in rice grains will be eliminated by applying different data cleaning techniques.

After pre-processing the complete data, K-NN and Decision tree algorithms will be used for classification and quality analysis of different kind of rice grains. Accuracy of both techniques will be determined and compared.

Research questions:

RQ1: How will you classify different types/grains of rice?

RQ2: How will you determine the quality of different grains?

RQ3: How will you handle the data of impurities contained by rice grain?

RQ4:

RQ5: What will be the accuracy of quality and classification determined by your proposed techniques.

Proposed Methodology:

To analyze the quality and classify different kind of rice grains, K-NN and Decision tree-based methodology is proposed. Focus of this system is to extract data of five different types of rice including Jasmine, Basmati, Arborio, Saila, Super Colonel and Old Awami Rice. For acquisition of data 1000 images of each type will be used and data will be formulated after applying different kind if image processing techniques.

Then this acquired data will subjected to different processes of data pre-processing (cleaning, integration, transformation, reduction and smoothing). On the basis of pre-processes data, quality analysis and classification will be performed.

About Data:

Data Name	Data Types	Attribute Type	No. of instances	No. of Attributes
Rice Grain Data	5 Classes	Integer / Real numbers	5000 (1000 each type)	10

Main Classes of Data: There will be five main classes of data as follows:

1. Jasmine
2. Gonen
3. Basmati
4. Ipsala
5. Arborio

Main Dimensions of Data: Data will consist of following dimensions for five different kind of rice grains:

1. ID
2. Area of grain
3. Major Axis length of single grain
4. Minor Axis length of single grain
5. Eccentricity of single grain
6. Convex Area of single grain
7. Extent (OR) Aspect Ratio. This will decide depending on availability of data.
8. Perimeter
9. Roundness
10. Class

Total Number of Data Value sets: As proposed earlier, 1000 images of each type will be extracted/process for data acquisition. There will be almost total 5000 number of instances.