STATEMENT OF WORK

|  |  |
| --- | --- |
| Topics | Page |
|  |  |
| * Indexing | 1 |
|  |  |
| * PROJECT Title & MEMBERS | 2 |
|  |  |
| * Objectives | 3 |
|  |  |
| * Data Information | 3 |
|  |  |
| * Solution Design | 4 |
|  |  |
| * Technology Stack | 4 |
|  |  |
| * Working Approach   + Data Analysis & EDA   + Model Building   + Model Evaluation & Selection | 5 |
|  |  |
| * project timeline | 9 |
|  |  |
| * deployment and repository | 10 |

PROJECT TOPIC

**Wheat Seed Classification**

PROJECT MEMBERS

|  |  |
| --- | --- |
| ***MEMBER NAME*** | ***STUDENT ID*** |
| Smriti Raina | 100830255 |
| Vaibhavi Thakkar | 100833169 |
| Shrutika Raut | 100844617 |

OBJECTIVES

The objective of the project is to recognize the varieties of wheat based on the morphology. Multiple classification algorithms will be used to determine the type of wheat (namely, Kama, Rosa and Canadian) for best accuracy.

DATA INFORMATION

<https://www.kaggle.com/jmcaro/wheat-seedsuci>

This data was acquired from the 'UCI Center for Machine Learning' repository. It contains seven variables for three distinct types of wheat kernels: (*Kama*, *Rosa*, *Canadian*) designated as numerical variables 1, 2 & 3 respectively. The seven seed variables are:

1. Area
2. Perimeter
3. Compactness
4. Kernel Length
5. Kernel Width
6. Asymmetry Coefficient
7. Kernel Groove Length

All these parameters are continuous and real valued.

The last column is reserved for the Kernel type. This dataset has 199 entries. Some of these variables are explicitly dependent. For example, *compactness*: C = 4*pi*Area/(Perimeter)^2 has a linear proportional relationship with *area*, and a square proportionality with *kernel width.*

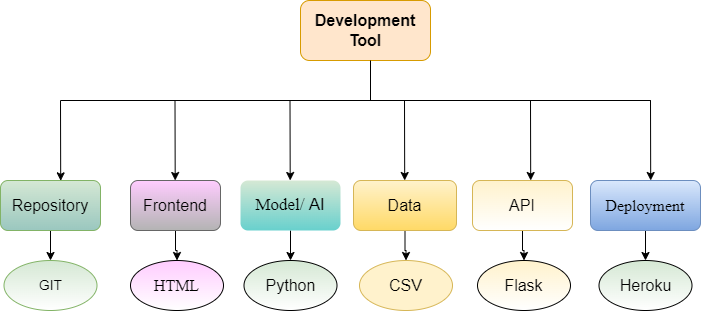
SOLUTION DESIGN

Diagram

Description automatically generated

***System Architecture***

TECHNOLOGY STACK

****

***Software tools***

WORKING APPROACH

Data Analysis

Below table shows detailed information of collected data for wheat seed classification where Type is the target variable.

Table

Description automatically generated

**Data Distribution:**

The Type feature contain three types of values namely, 1, 2 and 3 (*Kama*, *Rosa*, *Canadian respectively*). It is observed that data is not unbalanced.

Graphical user interface

Description automatically generated with low confidence

**Data Pre-processing:**

* **Normalization**

The data can be seen having different ranges. So, it was required to normalize using common scale to avoid distortion. Target variable was removed before normalization.

Chart, box and whisker chart

Description automatically generated

***Data before Normalization***

**Chart, box and whisker chart

Description automatically generated**

***Data after Normalization***

* **Detecting and Removing Outliers**

No outliers are detected

* **Data Correlation**

The heatmap shows that all the entities contribute to seed type prediction. Although Area and Perimeter has equal weightage, dropping either of them doesn’t have significant impact.

Chart

Description automatically generated

Model Building

Considering it’s a multiclass classification problem and given that there are limited features and data, our best approach was to consider

* Decision Tree Classifier
* Support Vector Classifier

Model Evaluation & Selection

Based on confusion matrix and accuracy, Support Vector Classifier proved to be best.

**Decision Tree Classifier** **Support Vector Classifier**A picture containing calendar

Description automatically generated

A picture containing chart

Description automatically generated

PROJECT TIMELINE

Graphical user interface, application, table, timeline

Description automatically generated

***Agile Gantt Chart***

PROJECT DEPLOYMENT & REPOSITORY

Project deployment was done on Heroku platform:

<https://wheatclassifier.herokuapp.com/>

Repository:

<https://github.com/TVaibhavi/Wheat-Detection-AI-Enterprise-Project>