# AIDI 2004-01 - AI in Enterprise Systems

Final Project Report

## Project Members

* Vineth Rajendran: 100835622 Product Owner
* Vaibhav Jain: 100837042 Scrum Master
* Joseph Mendez: 100324329 Product Analyst

## Project Title

Wheat Kernel Classification

## Objective

On a wheat seed dataset obtained from the [link](https://www.kaggle.com/jmcaro/wheat-seedsuci) having seven features for three different types of wheat kernels (Kama, Rosa, Canadian), we used EDA because it allows us to see the data's core structure, patterns, and relationships and to figure out which features to use to categorize the wheat kernels.

Then we build a classifier model to categorize the wheat kernels. We used plain machine learning classifiers and ensemble models to predict the type of wheat kernel based on the independent variables provide. Finally, developed a hyper tuned model to produce the best accuracy for the classification task at hand.

Benefits for the end user:

* Help in predicting the wheat kernel based on the values of the attributes provided.

## Data Set

This data was obtained from the [link](https://www.kaggle.com/jmcaro/wheat-seedsuci).

It has seven numerical variables for three different types of wheat kernels (Kama, Rosa, and Canadian), which are labelled 1, 2, and 3 accordingly.

The following are the seven seed variables:

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

## AI Workload pipeline

The AI Workload pipeline of our project is outlined below:

* + Data Collection, Preparation & EDA
    - Data is collected from the [link](https://www.kaggle.com/jmcaro/wheat-seedsuci).
    - EDA was performed to find the correlated variables if any or to find the data discrepancy and to find the insights and outliers present.
    - Null Value analysis
    - Outlier treatment using Tukey method
    - EDA, univariate, bivariate, correlation matrix
    - Feature selection using SelectFromModel
    - Balancing the imbalance data using SMOTE
    - Scaling the dataset
  + Understanding and using the right algorithms
    - Using the important algorithms that can be considered.
    - LogisticRegression
    - RandomForestClassifier
    - BaggingClassifier
    - DecisionTreeClassifier
    - AdaBoostClassifier
    - GradientBoostingClassifier
    - Understanding the above-mentioned algorithms and its working to tune the hyper-parameters in better way for higher accuracy.
  + Evaluate the model on Test-Train
    - Training the model and evaluating its accuracy on the training data.
  + Hyper-parameter Tuning
    - Understanding the accuracy and results of the model and fine-tuning the hyperparameters to boost the model accuracy if required and evaluate the accuracy on train data.
  + Model Finalization and deployment
    - Once done with the above-mentioned steps we finalized the model producing best accuracies and classify the wheat kernels with lesser errors. Model finalized was used to test on the unseen data and the final developed model was deployed.

## Project Milestones

The milestones achieved for this project are as follows:

* + Data Collection
  + Finding the right algorithms and its understanding
  + Preparing the model using the algorithms
  + Model Training and its evaluation
  + Hyperparameter Tuning
  + Model accuracy evaluation
  + Final model deployment

## Incorporate Agile Scrum philosophy

* + We used **GitHub** to follow the agile methodology and to keep proper track of code and to help the team to work simultaneously on the project.
  + We had **weekly sprint meetings** to keep track of the project milestones and build on top of the achieved milestones

## Model Deployed and predictions

Landing Page/ Home page

A screenshot of a computer

Description automatically generated

‘Kama’ – Type 1 kernal prediction

A screenshot of a computer

Description automatically generated

‘Rosa’ – Type 2 kernal prediction

A screenshot of a computer

Description automatically generated

‘Canadian’ – Type 3 Kernal prediction

A screenshot of a computer

Description automatically generated

## GitHub Url

**GitHub** url: <https://github.com/vaibhavjain12358/Wheat-Kernal-Classification>