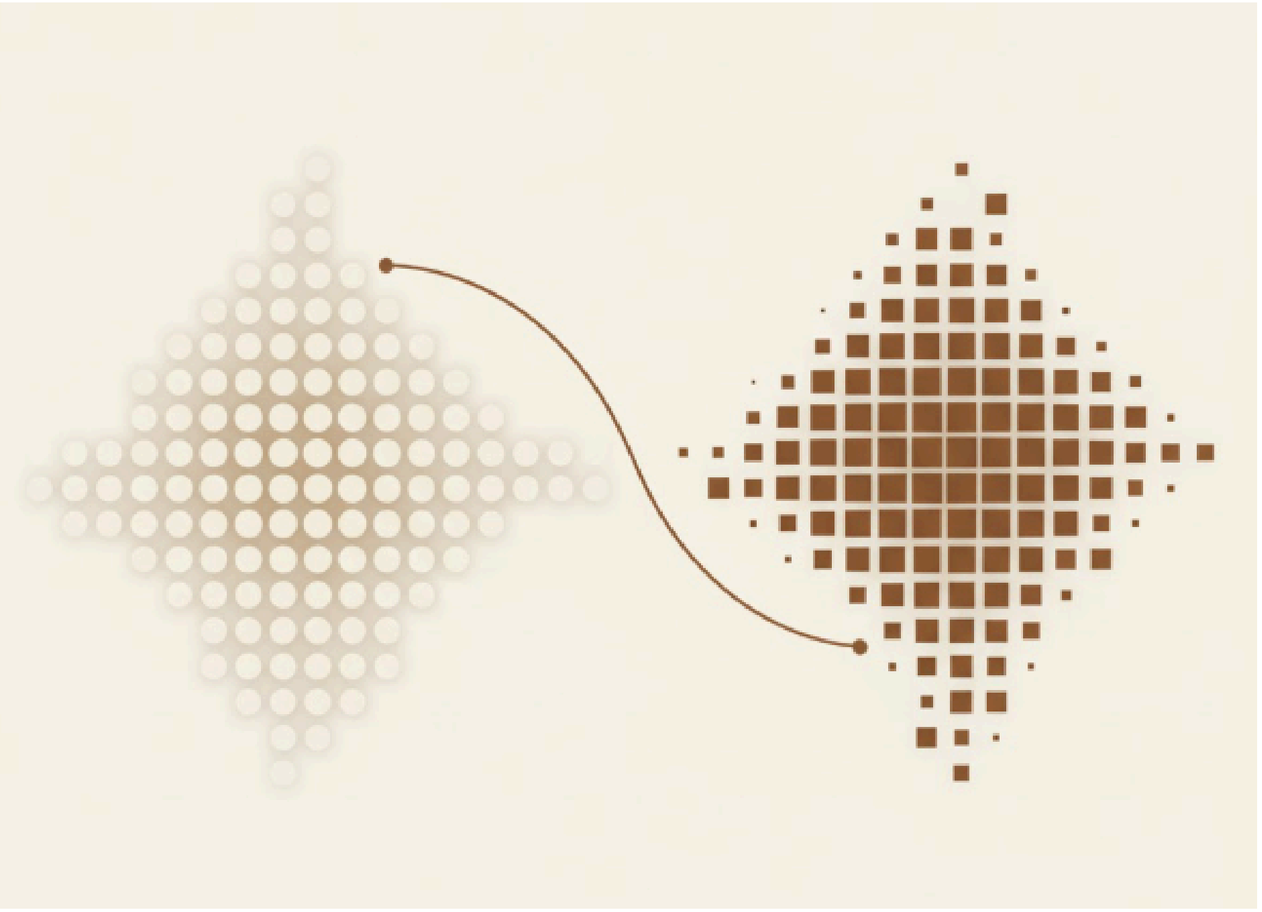
PROJECT REPORT

PRESENTATION

Introduction of Topic

Project tile: “Rice Variety Classifier”

Rice is one of the most widely consumed staple foods worldwide, with numerous varieties differing in grain size, shape, and color. Identifying rice types accurately is important for agriculture, food quality, and supply chain management. Traditional manual inspection is time-consuming and prone to error, especially for large datasets.
In this project, we use machine learning to classify rice varieties automatically. Instead of using complex deep learning models, we extract simple, interpretable features from rice images — such as area, perimeter, aspect ratio, and average color — and use algorithms like Logistic Regression, KNN, Decision Tree, and SVM for classification.
This approach demonstrates how feature engineering and classical ML can solve real-world problems efficiently, while keeping the model transparent and explainable. Our system can accurately distinguish between rice types like Arborio and Jasmine, and can be extended to other varieties in the future.

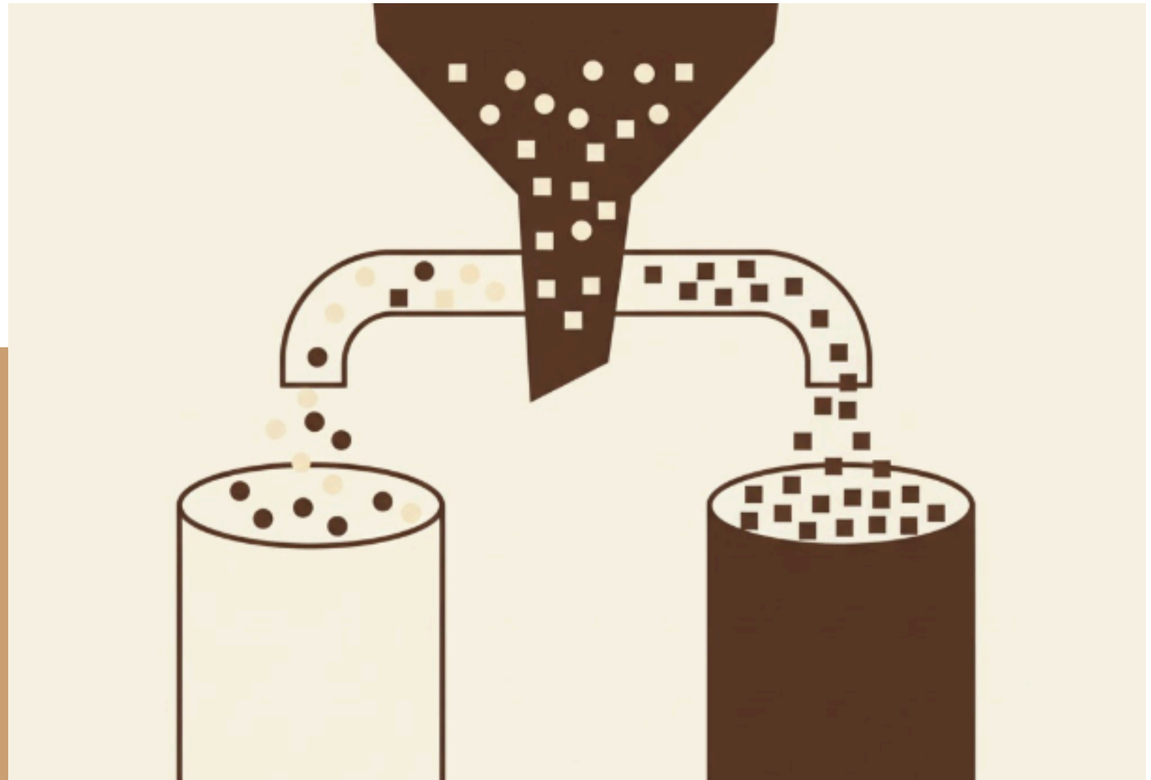


Abstract

This project focuses on classifying rice varieties automatically using images. Instead of complex deep learning, we extract simple features like area, perimeter, aspect ratio, and color from rice grains. These features are fed into Logistic Regression, KNN, Decision Tree, and SVM models to accurately identify types like Arborio and Jasmine.
The approach is fast, interpretable, and effective, showing how classic machine learning and smart feature extraction can solve real-world problems in agriculture and food quality.

Problem

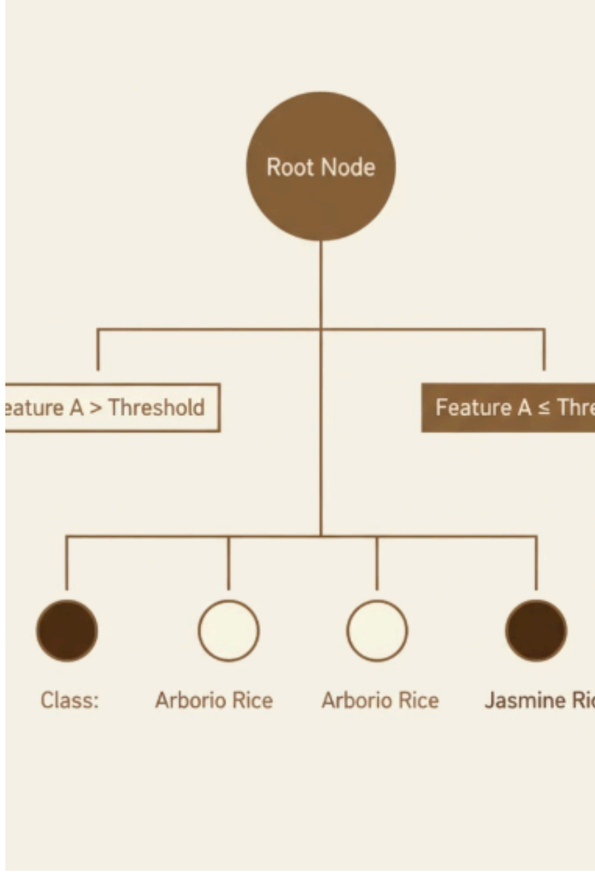
Rice is a staple food consumed worldwide, with many varieties that differ in size, shape, and color. Accurately identifying rice types is essential for agriculture, food quality control, and supply chain management. Traditional manual inspection is time-consuming, error-prone, and inefficient for large quantities of rice.



The problem addressed in this project is: How can we automatically and accurately classify rice varieties using images of rice grains?
To solve this, we extract key numeric features from rice images — such as area, perimeter, aspect ratio, and color — and use machine learning models like Logistic Regression, KNN, Decision Tree, and SVM to classify the rice type efficiently and reliably.

Objectives

1. Automate Rice Classification – Develop a system to classify rice varieties like Arborio and Jasmine using images, reducing reliance on manual inspection.
2. Feature Extraction – Extract meaningful numeric features from rice images, such as area, perimeter, aspect ratio, and color, to represent the grains.
3. Apply Machine Learning Models – Train and compare models like Logistic Regression, KNN, Decision Tree, and SVM to find the most accurate and interpretable solution.
4. Evaluate Model Performance – Analyze the accuracy, precision, recall, and confusion matrix of each model to ensure reliable classification.
5. Explainability – Understand which features contribute most to rice type classification, making the model transparent and interpretable.
6. Scalability – Create a workflow that can be extended to classify additional rice varieties in the future.



Technology Tools

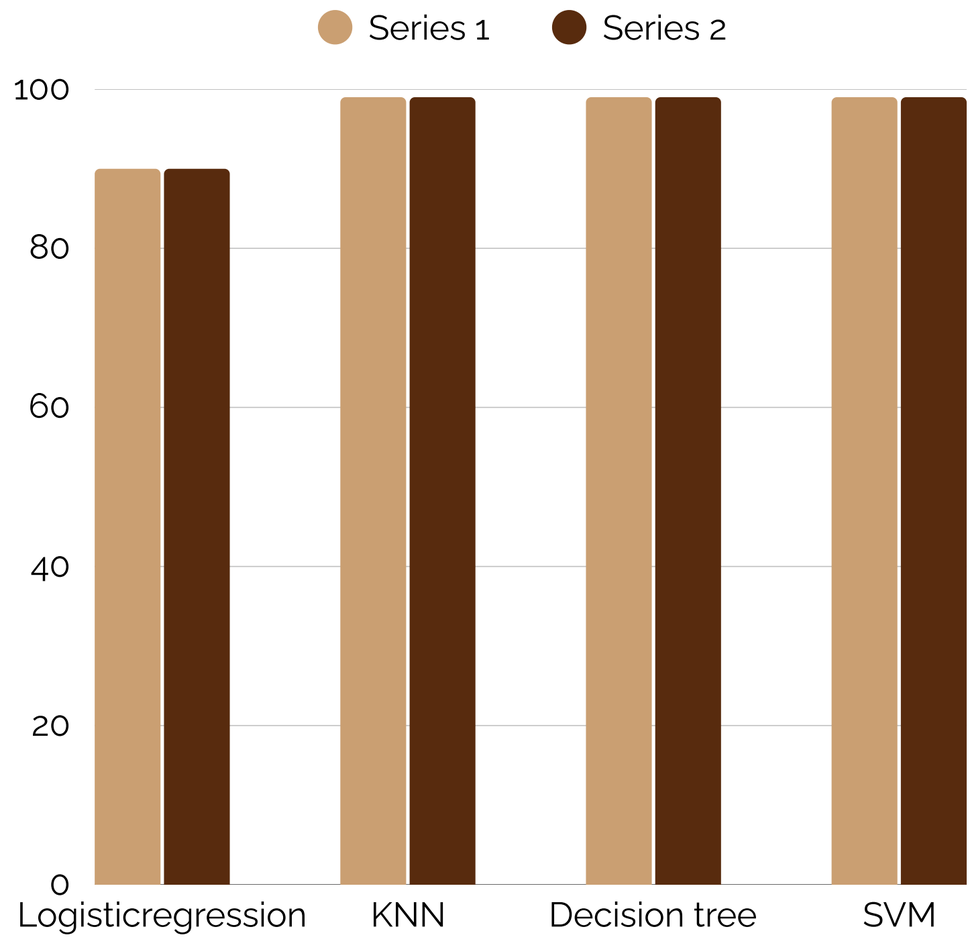
- Python – Primary programmir language for data processing extraction, and machine learning.
- Google Colab – Cloud-based environment for running Python code, accessing GPU, and handling datasets.
- OpenCV – Library for image processing, feature extraction, and manipulating rice grain images.
- NumPy & Pandas – Libraries for numerical computations and handling tabular data.
- Matplotlib & Seaborn – Visualization libraries for plotting EDA charts, distributions, and confusion matrices.
- Scikit-learn – Machine learning library used for training models like Logistic Regression, KNN, Decision Tree, and SVM.
- Jupyter Notebook / Colab Notebooks – Interactive coding and documentation environment for step-by-step analysis.
- CSV / Excel – For storing and managing extracted rice features in tabular form.

Research Methodology

- Data Collection: Rice grain images (Arborio & Jasmine) were used.
- Preprocessing: Images converted to grayscale and segmented.
- Feature Extraction: Extracted area, perimeter, aspect ratio, and mean color.
- Modeling: Trained Logistic Regression, KNN, Decision Tree, and SVM on extracted features.
- Evaluation: Checked accuracy, confusion matrix, and classification metrics.
- Prediction: New grain features are fed to the trained model for rice type prediction.

Data Analysis

- Logistic Regression is slightly lower due to linear boundaries but is interpretable.
- KNN, Decision Tree, and SVM achieve near-perfect accuracy because features are well-separated.
- Area and perimeter are the most important features for classification.
- Overall, the project shows that simple features + classical ML can classify rice types effectively.



Conclusion

This project demonstrates that rice varieties like Arborio and Jasmine can be accurately classified using simple image features such as area, perimeter, aspect ratio, and mean color. Models like KNN, Decision Tree, and SVM achieved near-perfect accuracy, while Logistic Regression was slightly lower but interpretable. The project highlights the importance of feature extraction, classical machine learning, and model explainability, showing that effective and reliable classification is possible without complex deep learning. This framework can be extended to classify more rice varieties or other agricultural products, making it practical and scalable.

References:

1. Murat Koklu. Rice Image Dataset. Kaggle. https://www.kaggle.com/datasets/muratkokludataset/rice-image-dataset
2. Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.
3. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
4. Raschka, S., & Mirjalili, V. (2019). Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing.
5. McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51–56.