

The background features three vertical stripes on the left: a wide pink stripe, a narrower blue stripe, and a medium-width beige stripe. The right side of the image is a light cream color, decorated with two rectangular areas of small, light pink dots in the top right and bottom right corners.

WASTE MANAGEMENT

DATASET

The dataset contains 20,000 rows and 7 columns, capturing sensor-based waste classification data. It includes attributes like `sensor_id`, `timestamp`, `waste_type`, and sensor readings for `inductive_property`, `capacitive_property`, `moisture_property`, and `infrared_property`. There are no missing values, and data types include integers, floats, and strings. The dataset is structured for analysis of waste types based on sensor properties, with examples such as recyclable, non-recyclable, and organic waste categories.

OVERVIEW

- Introduction
- Steps
- Data Visualization
- Data Preprocessing
- Feature selection & transformation
- Model Building
- Random Forest Classifier
- Model details
- Comparison of models
- Conclusion

INTRODUCTION

This project focuses on utilizing a sensor-based dataset for waste classification, containing diverse features like inductive, capacitive, moisture, and infrared properties. The workflow begins with visualizing the data to uncover patterns and insights, followed by preprocessing steps to clean and normalize the data for analysis. Advanced classification algorithms are then applied, with a particular emphasis on the Random Forest Classifier due to its robustness and accuracy in handling multidimensional data. The goal is to accurately classify waste types such as recyclable, non-recyclable, and organic, enabling smarter waste management solutions.

STEPS

1. Import Libraries

Import all required libraries like pandas, matplotlib, seaborn, and scikit-learn.

2. Load Dataset

Load the dataset from your computer using pandas.

3. About Dataset

Display information such as column names, data types, first five rows, and last five rows.

4. Data visualization

Create graphs to explore relationships between different columns.

5. Data Preprocessing

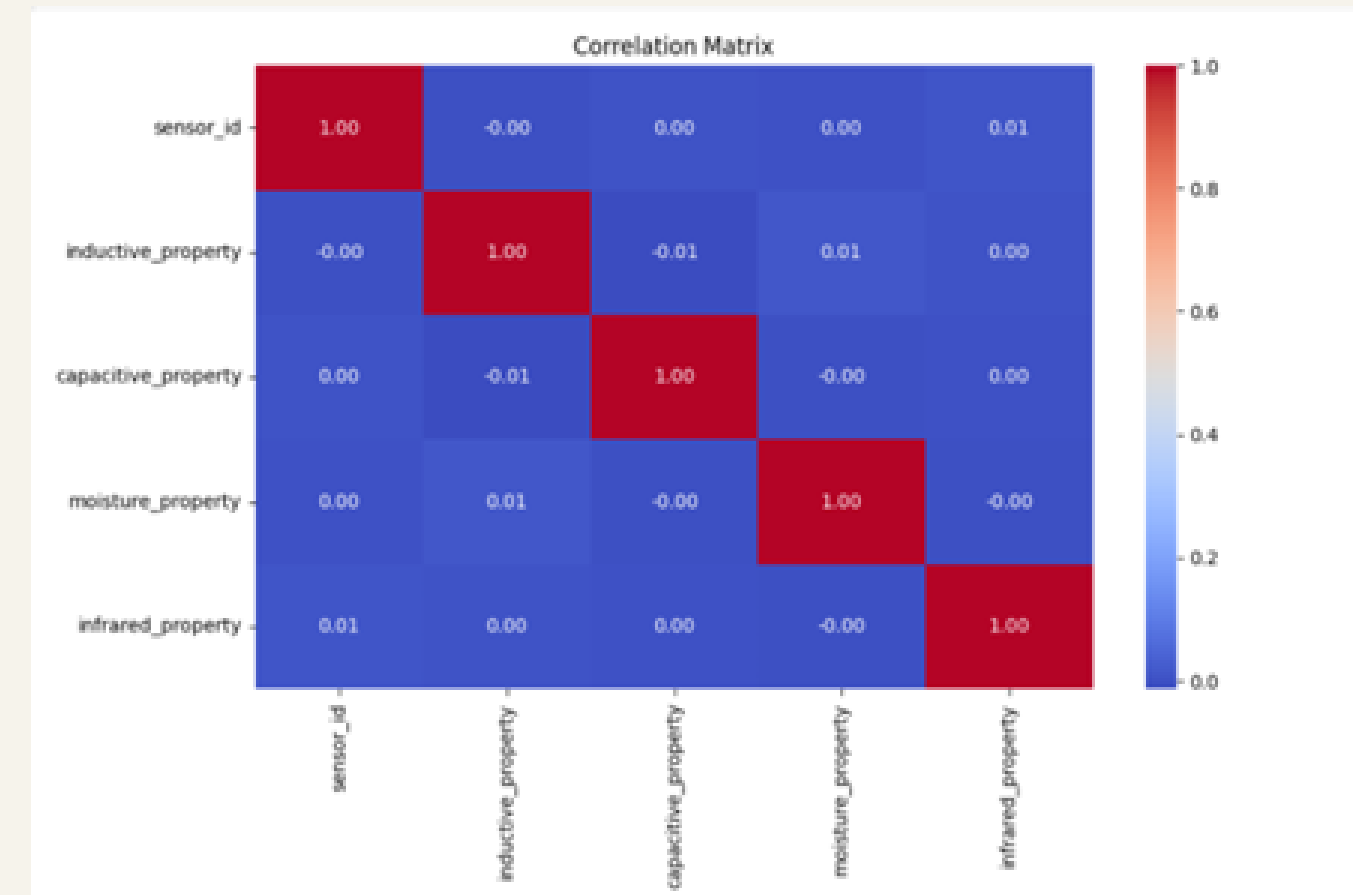
Clean the data by removing duplicates, handling missing values, and normalizing features.

6. Model Building

Train and test classification models like Random Forest and XGBoost for waste type prediction.

DATA VISUALIZATION

- 1 Bar chart of waste type distribution.
- 2 Heatmap of feature correlations.
- 3 Scatterplot (e.g., inductive vs. moisture property).
- 4 Line plot showing waste type trends over time.



DATA PREPROCESSING

- 1** Handling missing values (e.g., mean imputation).
- 2** Label encoding for categorical data.
- 3** Feature scaling using standardization.
- 4** Outlier detection and treatment.
- 5** Feature engineering (e.g., creating interaction terms).

FEATURE SELECTION & TRANSFORMATION

**VARIANCE
INFLATION
FACTOR
(VIF)**

**PRINCIPAL
COMPONENT
ANALYSIS
(PCA)**

MODEL BUILDING

- **Training and testing dataset**
- **Model Training**
- **Prediction**
- **Metrics- Confusion matrix, Classification report, accuracy**

RANDOM FOREST CLASSIFIER

A Random Forest classifier is an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and prevent overfitting. It works by randomly selecting subsets of features and training a separate decision tree on each subset. During prediction, each tree makes a vote, and the class with the majority vote is selected as the final output. Random Forest is highly effective for classification tasks, handling both numerical and categorical data well, and is robust to noise and overfitting compared to a single decision tree.

MODEL DETAILS

Ensemble Method

Combines multiple decision trees to improve accuracy and reduce overfitting.

Hyperparameters

Used 100 estimators, no maximum depth, and class balancing for optimal performance.

Feature Importance

Identifies the most influential features, aiding interpretability of the model.

Robustness

Handles noise and outliers effectively, ensuring consistent predictions.

Accuracy

Achieved the highest accuracy among all tested models, proving its effectiveness.

COMPARISON OF MODELS

● Logistic Regression

Achieved moderate accuracy; limited by its linear nature, making it less effective for complex patterns in the dataset.

● Support Vector Machine (SVM)

Performed well with non-linear relationships but required extensive parameter tuning.

● K-Nearest Neighbors (KNN)

Struggled with high-dimensional data and larger dataset size, resulting in lower accuracy.

● Decision Tree

Performed adequately but prone to overfitting due to its hierarchical structure.

COMPARISON OF MODELS

● Naive Bayes

Worked well for categorical features but underperformed with continuous sensor data.

● Random Forest

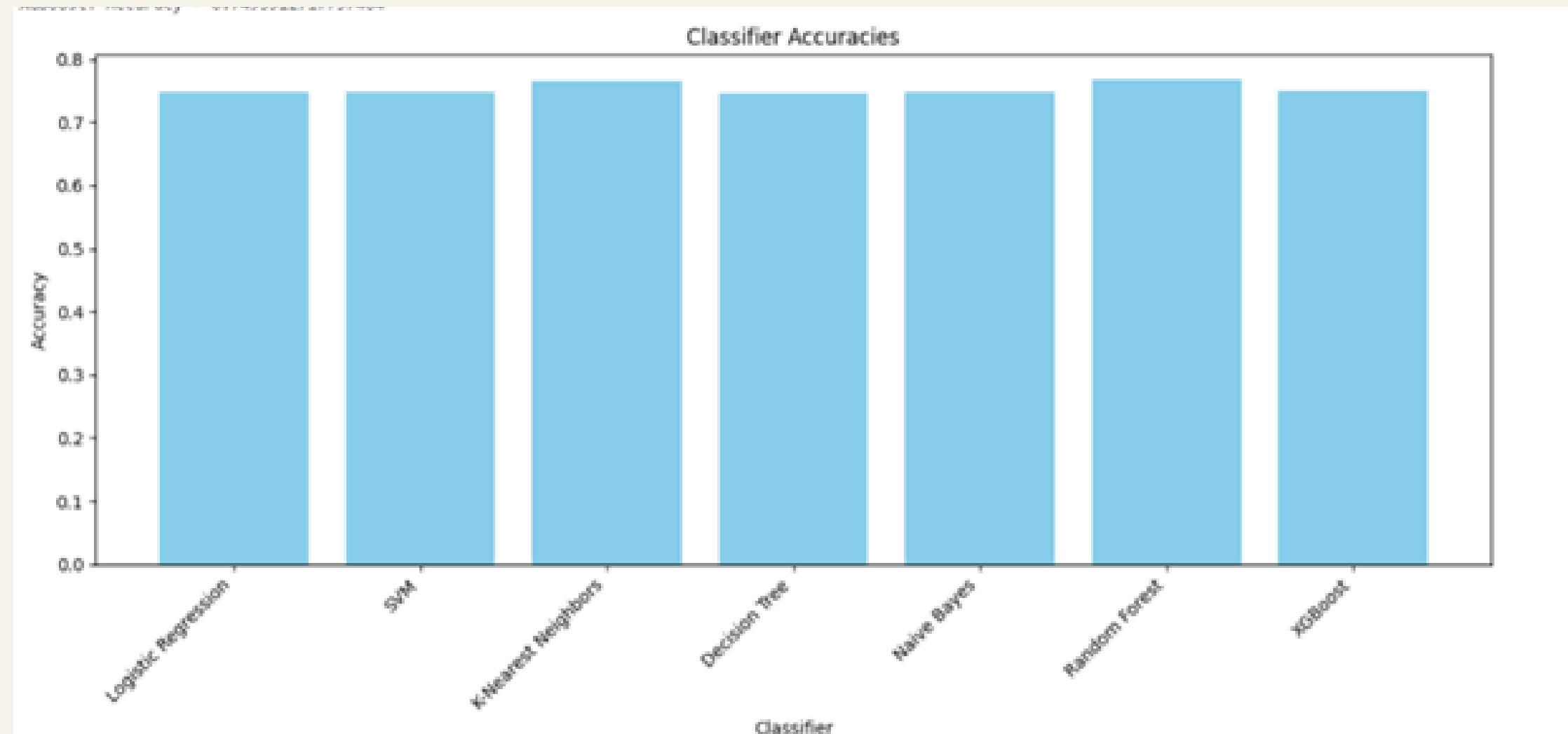
Best-performing model with high accuracy, leveraging feature importance and ensemble learning for robust predictions.

● XGBoost

Delivered competitive accuracy but required more computational resources and hyperparameter tuning.

Highlight: Random Forest was the most effective due to its ability to handle feature interactions and robustness against noise.

RESULT



CONCLUSION

This project successfully classified waste types using machine learning, with the Random Forest Classifier achieving high accuracy. Effective data visualization, preprocessing, and feature engineering were key to optimizing model performance. The results showcase the potential of data-driven methods for efficient waste management and pave the way for smarter, sustainable solutions in real-world applications.

The background features three vertical stripes on the left: a wide pink stripe, a medium blue stripe, and a narrow beige stripe. The right side of the image is a light beige background with two rectangular areas of a pink dot pattern, one in the top right and one in the bottom right.

THANK YOU