A FIELD PROJECT REPORT

on

"ENHANCED PREDICTION OF FOOD WASTAGE USING ENSEMBLE LEARNING AND FEATURE SELECTION"

Submitted

by

221FA04399
P Venkata Kamya

221FA04027
K Akshay

221FA04653
M Vijayalakshmi

221FA04037
P Ruthvik

Under the guidance of

Mrs.M.Bhargavi

Assistant Professor Department of CSE,VFSTR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH

Vadlamudi, Guntur.

ANDHRA PRADESH, INDIA, PIN-522213.

CERTIFICATE

This is to certify that the Field Project entitled **"ENHANCED PREDICTION OF FOOD WASTAGE USING ENSEMBLE LEARNING AND FEATURE SELECTION"** that is being submitted by 221FA04399 (Venkata Kamya), 221FA04653 (Vijayalakshmi), 221FA04027 (Akshay), 221FA04037 (Ruthvik) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of Mrs. M.BHARGAVI, M.Tech., Assistant Professor, Department of CSE.

M .Bhargavi	Dr. S. V. Phani Kumar	Dr.K.V. Krishna Kishore
Assistant Professor, CSE	HOD,CSE	Dean, SoCI

DECLARATION

We hereby declare that the Field Project entitled **"ENHANCED PREDICTION OF FOOD WASTAGE USING ENSEMBLE LEARNING AND FEATURE SELECTION"** is being submitted by 221FA04399 (Venkata Kamya), 221FA04653 (Vijayalakshmi), 221FA04027 (Akshay), 221FA04037 (Ruthvik) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of Mrs.M.BHARGAVI, M.Tech., Assistant Professor, Department of CSE.

By

221FA04399 (Venkata Kamya),

221FA04653 (Vijayalakshmi),

221FA04027 (Akshay),

221FA04037 (Ruthvik)

Date:

ABSTRACT

This paper presents an ensemble learning-based framework for forecasting food wastage into this scenario, managing variability in data, dependency on certain environmental conditions, and socio-economic differences. For carrying out feature selection technique RFE is applied using a wide dataset that considered environmental conditions, socio-economic indicators, and historical records on such phenomena. For evaluation purposes, various ensemble models are selected: Random Forest, Gradient Boosting, LightGBM, and so forth and evaluated using MSE, RMSE, R^2 and MAE as metrics. The technique gave very accurate predictions and withstood dropouts, and the features importance analysis provided valuable information to the stake holders/policymakers on how to design better sustainable food distribution strategies.

Keywords - Food Wastage Prediction, Machine Learning,Bagging regressor,Random Forest,Gradient Boosting,LightGBM,Extra Trees,Stacking Regressor

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER-1

INTRODUCTION

1. INTRODUCTION

Probably one of the major concerns across the world today is food wastage, which traverses to the environment, the economy, and society at large. Food and Agriculture Organization of the United Nations estimated that one-third of all food produced globally for human consumption was lost or wasted each year. This not only deepens food poverty but also contributes significantly to environmental degradation through greenhouse gas emissions and the unnecessary use of resources.

Researchers have now increasingly used sophisticated models of machine learning in predicting and minimizing food waste across all possible situations. Projections were able to offer stakeholders perception of the patterns and trends associated with food waste, thereby improving strategies on minimizing wastes. This paper presents ensemble learning techniques, including Random Forest, Gradient Boosting, LightGBM, Extra Trees, and Bagging Regressor, using Recursive Feature Elimination (RFE) in feature selection.

The main objective of this study is to build up a robust pipeline of building prediction starting from preprocessing of data up to selection of feature and application of ensemble learning models for the prediction of food wastage based on event-specific data. We will apply these models on a broad dataset that considers all the environmental conditions and socio-economic factors, along with historical records, to come up with accurate and reliable forecasts of food wastage.

These models are evaluated and ranked against measures including mean squared error (MSE), root mean squared error (RMSE), R-squared (R^2), and mean absolute error (MAE) with the goal of generating implications of results toward further improvement of efforts toward increasing the sustainability of food supply chains.

In Section 2, review of the existing literature related to predictions of food wastage using machine learning techniques has been done in detail. Section 3 documents the proposed methodology. This comprises methodologies on data preprocessing, feature selection, and model implementation. Section 4 discusses the results of experiments. Section 5 concludes by analyzing the performance of the model, its implications, and future research areas.

CHAPTER-2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 *Literature review*

A literature survey is a systematic examination of existing research on a particular topic. It serves as the foundation for any scholarly investigation, offering insights into current knowledge, identifying research gaps, and providing context for new studies. By synthesizing and summarizing relevant literature, researchers can formulate precise research questions, build upon existing work, and avoid duplication. In essence, a literature survey is an essential tool for ensuring the validity and relevance of new research within the broader academic landscape.

A literature survey is a systematic examination of existing research on a particular topic. It serves as the foundation for any scholarly investigation, offering insights into current knowledge, identifying research gaps, and providing context for new studies. By synthesizing and summarizing relevant literature, researchers can formulate precise research questions, build upon existing work, and avoid duplication. In essence, a literature survey is an essential tool for ensuring the validity and relevance of new research within the broader academic landscape.

NO.	Author(s)	Model/Approach	Accuracy/Results	Limitations
1.	K.S.Chowdary	XGBoost (Regression), Feature Extraction (PCA, RFE)	73% accuracy	Data quality and availability
2	Md Masrur Masuk Shopnil	Random Forest, Logistic Regression, SVM, Label encoding	75% accuracy	Lack of awareness, uncertainty in demand, resource constraints
3	N.M Ifham	Decision Tree, Random Forest, One-hot encoding	86.11% accuracy	Data quality, feature selection, model complexity
4	Pavan	Logistic Regression, Random Forest,	70-90% accuracy (99.53% for food	Technology access, cost,

	Manjunath	SVM, CNNs (IoT-based dataset)	freshness)	data privacy, scaling ability
5	Farzana Shaikh	Inception v3 (image classification from TrashNet)	83.30% accuracy	Training images need expansion, real-time processing optimization
6	J. Relin Francis Raj	LSTM networks, Fuzzy logic controllers (IoT sensor data)	>97% accuracy	Anomaly detection, environmental factor tuning
7	Asha V	CNN, ResNet50, YOLO (waste segregation)	95% accuracy	Lack of standardized waste data, sensor noise
8	Md. Masrur Masuk Shopnil	Ensemble ML Model (Random Forest, Logistic Regression, SVM)	75% accuracy	Data quality, feature selection, model complexity
9	N.M. Ifham	Decision Tree, Random Forest with One-Hot Encoding	86.11% accuracy	Data quality, feature selection, model complexity
10	K. S. Chowdary	XGBoost Regression, PCA, RFE	73% accuracy	Data variability, environmental dependencies
11	Bhargav Sagiraju	Machine Learning Model for Meat Freshness Prediction	>90% accuracy	Limited to meat products, need for diverse datasets
12	Asha V	CNN, ResNet50, YOLO for Waste Segregation	95% success rate in waste classification	Non-availability of standardized waste statistics, sensor noise
13	Farzana Shaikh	Inception v3 for Waste Image Classification	83.30% accuracy	Need for more training images, real-time processing optimization
14	J. Relin Francis Raj	AI-Powered Smart Waste Bins using LSTM Networks	>97% accuracy in predicting waste buildup trends	Further tuning needed for anomaly detection,

				environmental factors
15	Pavan Manjunath	IoT-Based Food Wastage Management System using ML	Predictions of food wastage with accuracies between 70-90%	Access to technology, cost, data privacy, scaling ability

CHAPTER-3

PROPOSED SYSTEM

3 Methodology of the system

This method starts with loading a dataset, where the data is separated into feature independent variables and a target dependant variable. Here, the algorithm identifies the categorical and numerical columns via select types. With numerical, missing values were imputed as the mean using ‘StandardScaler‘. For categorical values, missing value imputation will be done by the most frequent value and encoded using one-hot encoding. This process, taken care of by a ‘Column Transformer‘, is implemented on both the training and the test set. RFE with a Random Forest Regressor is conducted on the model for its optimization. The transformed data is further used to train an ensemble model, such as Random Forest, Gradient Boosting, Boost, Light GBM, Extra Trees, and a stacking regressor. A comparison is done in MSE, MAE, R-squared (R^2), and RMSE and the results plotted for comparison. (Discuss on proposed architecture.)

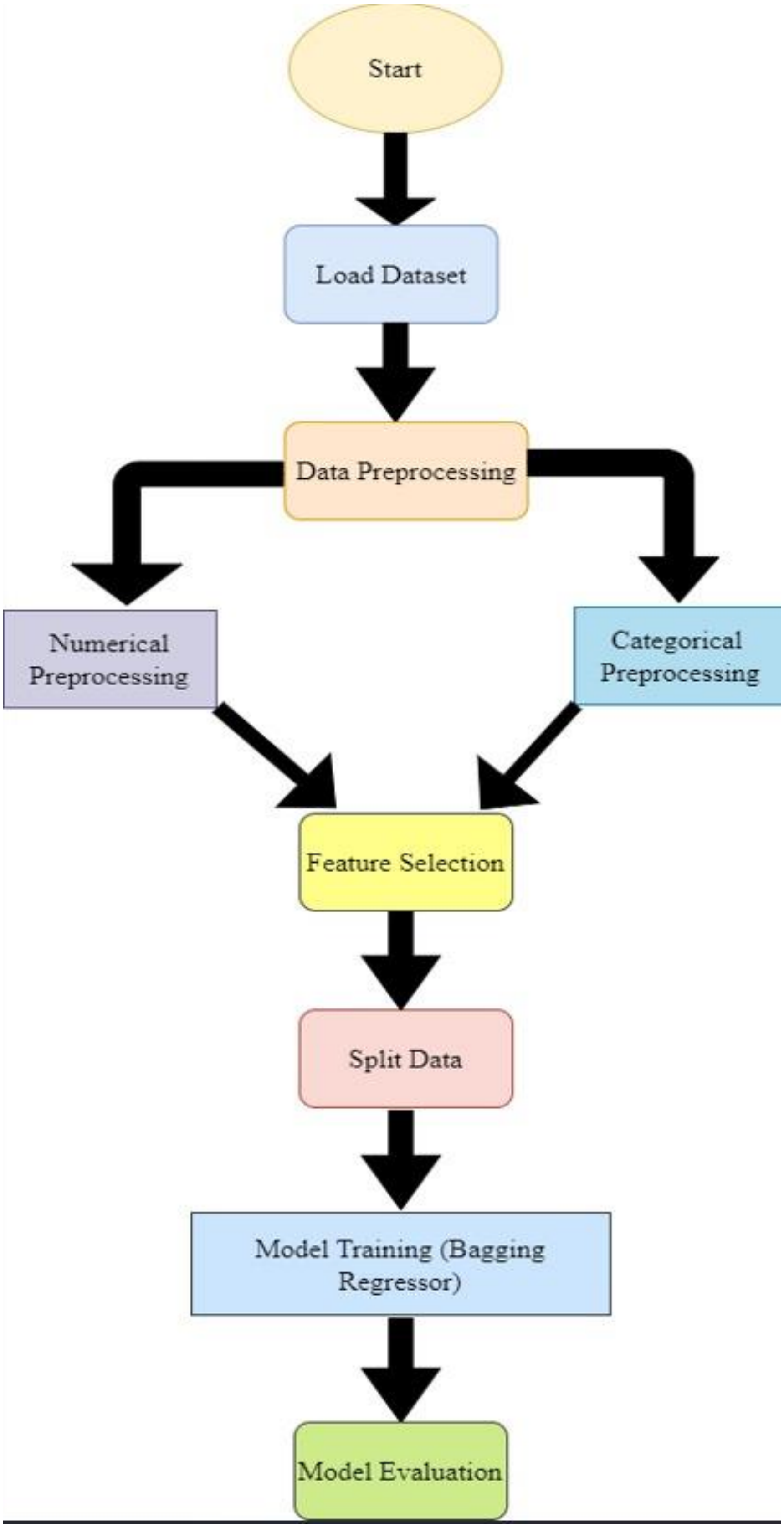

Figure 1: Flow diagram of proposed mode

3.1 Preprocessing

During preprocessing, the data is divided between numerical features and categorical features. We use the select dtypes function to do so. For numerical features, we created a pipeline that included the SimpleImputer to impute missing values with the mean and StandardScaler, scaling the values for standardization. For categorical features, missing values are managed by replacing them with the most frequent value in another SimpleImputer. After this step, categorical variables are converted into the binary indicators one-hot encoding by 4 OneHotEncoder These pre-processing steps for both numeric and categorical data are combined by using Column Transformer. Then applying this pipeline to training data by the fit transform() method which fits the pre-processing steps to this training set, transforming it, respectively. The same transformation is applied to the test set by calling the transform() method. This is going to sync any data between both sets. Missing values, scaling, and categorical encoding are taken care of systematically before applying a model.

3.2 Exploratory Data Analysis

Exploratory Data Analysis can be initiated with the use of a heatmap to analyse the correlation of numerical features. Development in the form of a matrix of correlations- using the 'sns.heatmap()' function from Seaborn explains the associations between variables and can even prove to be helpful in identifying the highly correlated variables that are influencing the model. Annotation of the heatmap with a colour scheme (YlGnBu) helps in better understanding the correlations.

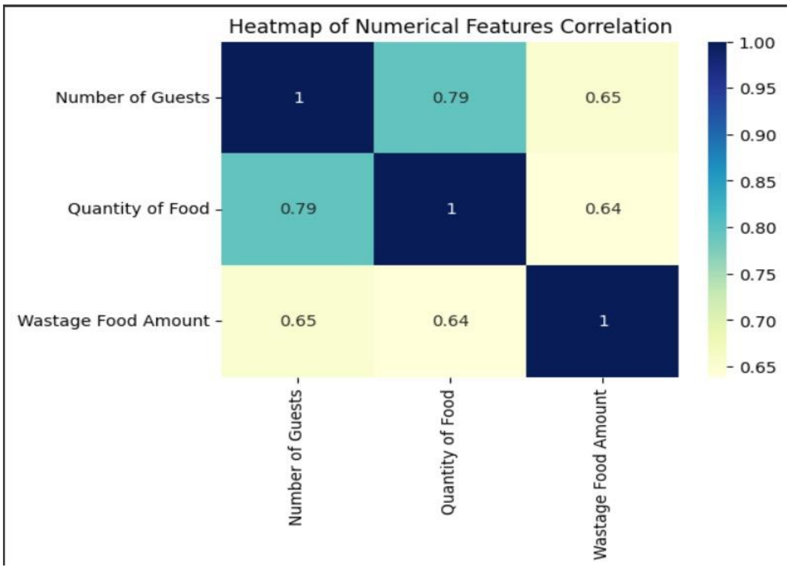

Figure 2: Heatmap to show correlation between numerical features

To compare average food wastage between categories, bar charts are then made. The first bar chart shows the overall food wastage for different event types, showing which event types contribute most to wastage, and the second bar chart plots the geographical location to the average food wastage level of wastage levels across different regions. These plots make use of sns.barplot() with ci=None to prevent the plots from displaying with confidence intervals, then set labels and a title for easy reading.

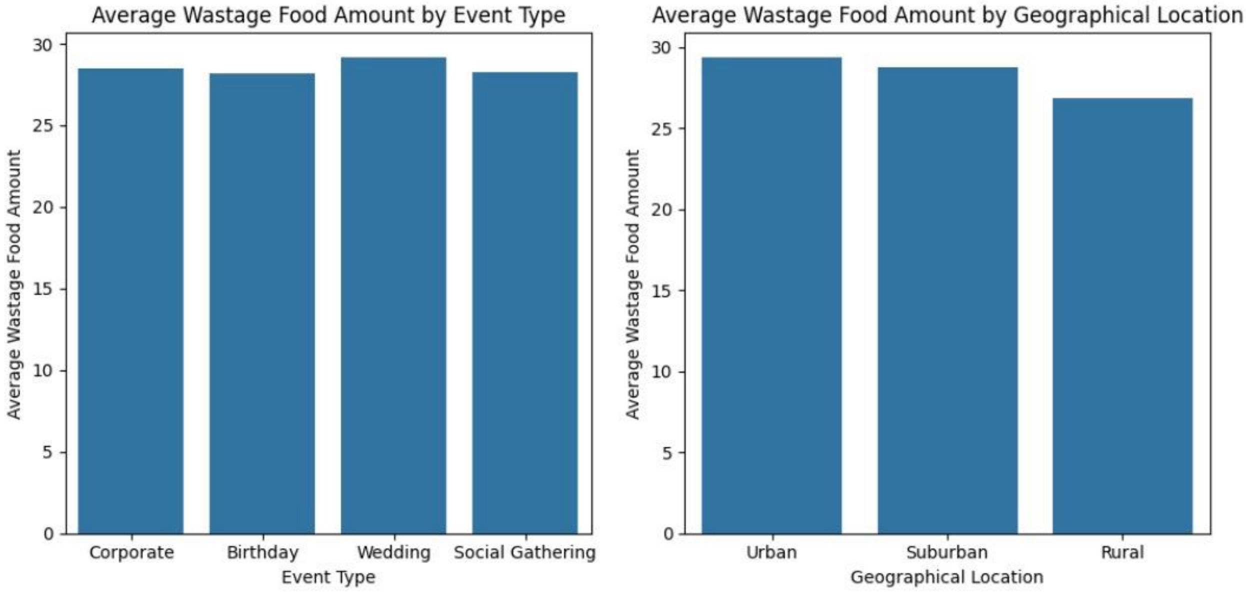

Figure 3. Bar chart for average wastage by event type

3.3 Feature Selection

This methodology, besides improving model interpretability, enhances the performance of the models since all the features are being narrowed down to the most relevant ones, reducing overfitting. A simple, systematic feature selection that could be used for easy processing is Recursive Feature Elimination (RFE), which cuts out several features from the feature set on each iteration. Feature selection mathematically can be defined as

$$minimize\ L(\theta) \quad subjectto \quad ||\theta||_0 \geq m$$

Where: L(θis objective function is weights or coefficients for the features,with m being the number of desired selected features . 0 be the count of non-zero weights in and thus m denote the selected features. RFE is an algorithmic feature selection procedure where features are ranked in terms of importance, and the least relevant features, step by step, are removed.

Starting with fitting a model on all features and ranking them based on the importance of their coefficients or some measure of feature importance, it will merely remove the least relevant feature and refit the model with the remaining ones. It repeats the procedure until it reaches the number of desired features. Mathematically, this can be represented as:

$$minimize\ L(\theta) \quad subjectto \quad \|\theta\|_0 = m$$

The core idea of RFE: minimization of the objective function L(θ . In the case of ’m’ features, the most relevant ones are selected. RFE successively removes those features whose removal induces the smallest increase in the objective function . The RFE algorithm, by its method of iteratively removing features, determines the subset of features from which the variability in the data can best be captured in regards to food wastage.

3.4 Regression using Ensemble Models

In the food waste prediction process, a holistic approach was used in addressing the problems associated with food wastage during event scenarios. The following ensemble machine learning models were adopted: Random Forest, Gradient Boosting, LightGBM, XGBoost, Extra Trees, Bagging Regressor, and Stacking. The emphasis here was on building models with good accuracy and reliability in the prediction of amounts of food waste.

After performing an intensive comparative analysis, the best predictive candidate for food waste belonged to the Bagging Regressor. Bagging is one form of ensemble method, which creates several copies of the same model on the entire given data. It takes the average of the predictions coming from all these models, thus reducing variance and avoiding overfitting.

The objective function for the Bagging Regressor is given by:

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \hat{y}_i) + \lambda \sum_{k=1}^{K} \Omega(f_k)$$

- Where: • $J(\theta)$ is the overall loss function,
- • yi is the actual target value for instance i,
- • yˆI is the predicted value for instance I,
- • L(yi , yˆi) is the loss function measuring the difference between the predicted and actual values (e.g., mean squared error),
- • Ω(fk) is the regularization term for the base learner k,
- • λ is a regularization parameter that controls the strength of the regularization,
- • n is the number of instances,
- • K is the number of base learners.

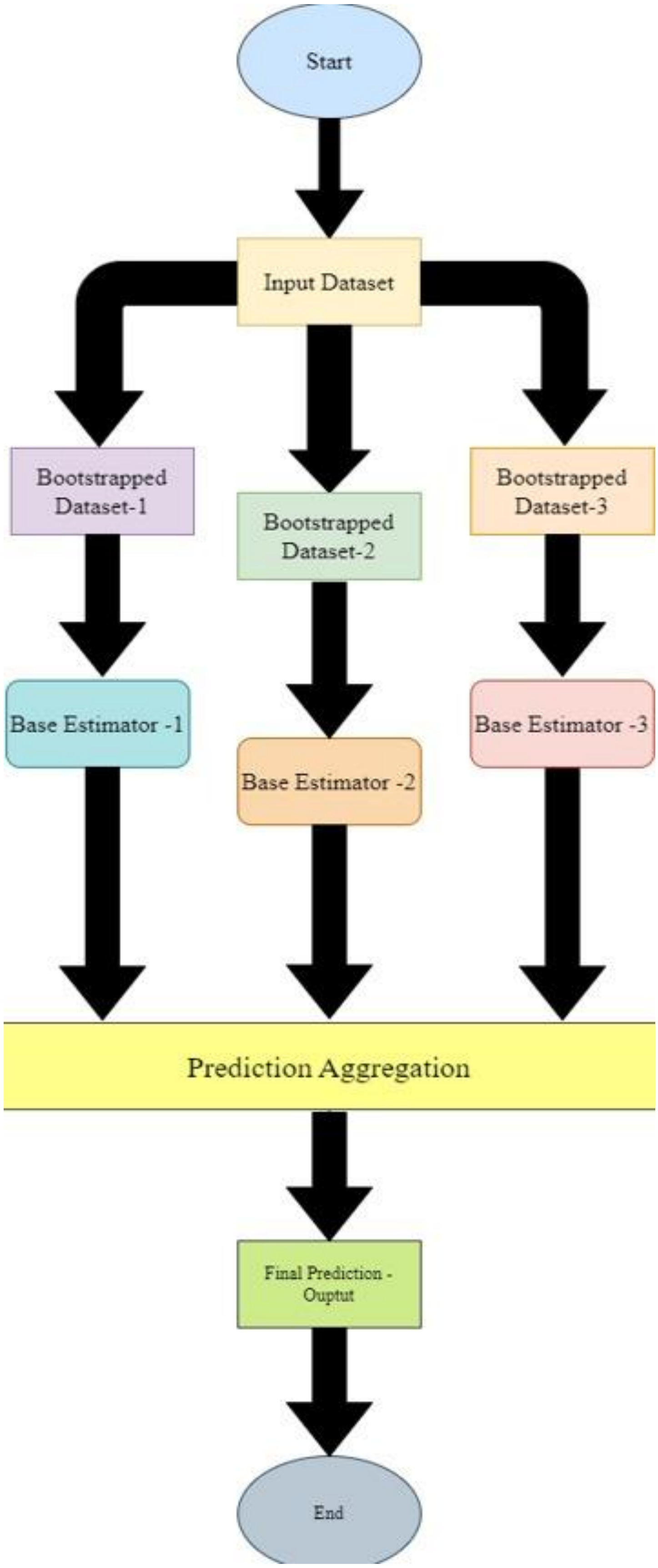

Figure 4: Visualization of Bagging Regressor architecture

CHAPTER-4

IMPLEMENTATION

Source code :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor, ExtraTreesRegressor, BaggingRegressor
from sklearn.feature_selection import RFE, SelectKBest, f_regression
from sklearn.metrics import mean_squared_error, r2_score,mean_absolute_error,root_mean_squared_error
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
import xgboost as xgb
import lightgbm as lgb
from sklearn.ensemble import StackingRegressor
import warnings

# Import necessary libraries
warnings.filterwarnings("ignore")

# Load dataset
data = pd.read_csv('/content/food_wastage_data.csv')
data.head(10)
import seaborn as sns
import matplotlib.pyplot as plt

# Heatmap to show correlation between numerical features
plt.figure(figsize=(6,4))
sns.heatmap(data.corr(numeric_only=True), annot=True, cmap="YlGnBu")
plt.title('Heatmap of Numerical Features Correlation')
```

```
plt.show()

# prompt: insert both the above bar graphs side by side

import matplotlib.pyplot as plt
import seaborn as sns

# Create a figure with two subplots side by side
fig, axes = plt.subplots(1, 2, figsize=(10, 5))

# First subplot: Average wastage by event type
sns.barplot(x='Event Type', y='Wastage Food Amount', data=data, ci=None, ax=axes[0])
axes[0].set_title('Average Wastage Food Amount by Event Type')
axes[0].set_ylabel('Average Wastage Food Amount')
axes[0].set_xlabel('Event Type')

# Second subplot: Average wastage by geographical location
sns.barplot(x='Geographical Location', y='Wastage Food Amount', data=data, ci=None, ax=axes[1])
axes[1].set_title('Average Wastage Food Amount by Geographical Location')
axes[1].set_ylabel('Average Wastage Food Amount')
axes[1].set_xlabel('Geographical Location')

# Adjust layout and display the plot
plt.tight_layout()
plt.show()

X = data.drop('Wastage Food Amount', axis=1)
y = data['Wastage Food Amount']

# Identify categorical and numerical features
categorical_cols = X.select_dtypes(include=['object']).columns
numerical_cols = X.select_dtypes(include=['int64', 'float64']).columns
```

```
# Preprocessing for numerical data
numerical_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='mean')),
    ('scaler', StandardScaler())])

# Preprocessing for categorical data
categorical_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='most_frequent')),
    ('onehot', OneHotEncoder(handle_unknown='ignore'))])

# Bundle preprocessing for numerical and categorical features
preprocessor = ColumnTransformer(
    transformers=[
        ('num', numerical_transformer, numerical_cols),
        ('cat', categorical_transformer, categorical_cols)])

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Apply preprocessing before feature selection
X_train_processed = preprocessor.fit_transform(X_train)
X_test_processed = preprocessor.transform(X_test)
# Random Forest Regressor for feature selection
from sklearn.ensemble import RandomForestRegressor
from sklearn.feature_selection import RFE
import numpy as np

# Initialize the Random Forest Regressor
rfr = RandomForestRegressor(n_estimators=100, random_state=42)

# Recursive Feature Elimination (RFE) for feature selection
```

```
# Adjust n_features_to_select as per your requirement
rfe = RFE(estimator=rfr, n_features_to_select=10)

# Fit RFE on the preprocessed training data
X_train_rfe = rfe.fit_transform(X_train_processed, y_train)

# Transform the test data using the selected features
X_test_rfe = rfe.transform(X_test_processed)

# Get the indices of the selected features for RFE
selected_features_rfe_indices = np.where(rfe.support_)[0]

# Get the names of the selected features
# Ensure 'preprocessor' has the correct feature names
feature_names = preprocessor.get_feature_names_out()
selected_features_rfe = [feature_names[i] for i in selected_features_rfe_indices]

# Output the selected feature names after RFE
print("Selected Features using RFE:")
print(selected_features_rfe)



# Dictionary to store evaluation metrics for each model
metrics = {}

# Ensemble models
models = {
    'Random Forest': RandomForestRegressor(n_estimators=100, random_state=42),
    'Gradient Boosting': GradientBoostingRegressor(n_estimators=100, random_state=42),
    'LightGBM': lgb.LGBMRegressor(n_estimators=100, random_state=42),
    'XGBoost': xgb.XGBRegressor(n_estimators=100, random_state=42),
    'Extra Trees': ExtraTreesRegressor(n_estimators=100, random_state=42),
```

```
    'Bagging Regressor': BaggingRegressor(n_estimators=100, random_state=42)

}

# Create a list of estimators for stacking
estimators = [
    ('rf', RandomForestRegressor(n_estimators=100, random_state=42)),
    ('gb', GradientBoostingRegressor(n_estimators=100, random_state=42)),
    ('lgbm', lgb.LGBMRegressor(n_estimators=100, random_state=42)),
    ('xgb', xgb.XGBRegressor(n_estimators=100, random_state=42)),
    ('et', ExtraTreesRegressor(n_estimators=100, random_state=42)),
    ('bag', BaggingRegressor(n_estimators=100, random_state=42))
]

# Create a stacking regressor:
stacking_regressor = StackingRegressor(
    estimators=estimators,
    final_estimator=RandomForestRegressor(n_estimators=100, random_state=42)
)

# Add stacking regressor to the models dictionary
models['Stacking Regressor'] = stacking_regressor

for model_name, model in models.items():
    # Train model on RFE selected features
    model.fit(X_train_rfe, y_train)
    y_pred_rfe = model.predict(X_test_rfe)

    # Evaluate model on RFE selected features
    mse_rfe = mean_squared_error(y_test, y_pred_rfe)
    r2_rfe = r2_score(y_test, y_pred_rfe)
    mae_rfe = mean_absolute_error(y_test, y_pred_rfe)
```

```
    rmse_rfe = np.sqrt(mse_rfe)

    metrics[model_name] = {
        'RFE': {'MSE': mse_rfe, 'R2': r2_rfe, 'MAE': mae_rfe, 'RMSE': rmse_rfe},
    }

# Create a table to display the evaluation metrics for each model
results_df = pd.DataFrame({
    'Model': list(metrics.keys()),
    'MSE ': [metrics[model]['RFE']['MSE'] for model in metrics],
    'MAE ': [metrics[model]['RFE']['MAE'] for model in metrics],
    'R2 ': [metrics[model]['RFE']['R2'] for model in metrics],
    'RMSE': [metrics[model]['RFE']['RMSE'] for model in metrics],
})

# Display the table
print(results_df)



# Extract the metrics for plotting
models = results_df['Model'].tolist()
mse_values = results_df['MSE '].tolist()
mae_values = results_df['MAE '].tolist()
r2_values = results_df['R2 '].tolist()
rmse_values = results_df['RMSE'].tolist()

# Create subplots for each metric
fig, axs = plt.subplots(4, 1, figsize=(6, 15))  # Reduced width to 6
fig.suptitle('Model Performance Metrics')

# Plot MSE
axs[0].plot(models, mse_values, marker='o', linestyle='-')
```

```
axs[0].set_ylabel('MSE')
axs[0].set_title('Mean Squared Error')
axs[0].tick_params(axis='x', rotation=45)

# Plot MAE
axs[1].plot(models, mae_values, marker='o', linestyle='-')
axs[1].set_ylabel('MAE')
axs[1].set_title('Mean Absolute Error')
axs[1].tick_params(axis='x', rotation=45)

# Plot R2
axs[2].plot(models, r2_values, marker='o', linestyle='-')
axs[2].set_ylabel('R2')
axs[2].set_title('R-squared')
axs[2].tick_params(axis='x', rotation=45)

# Plot RMSE
axs[3].plot(models, rmse_values, marker='o', linestyle='-')
axs[3].set_ylabel('RMSE')
axs[3].set_title('Root Mean Squared Error')
axs[3].tick_params(axis='x', rotation=45)

plt.tight_layout()
plt.show()

# prompt: plot MAE,RMSE graphs side by side

# Extract the metrics for plotting
models = results_df['Model'].tolist()
mae_values = results_df['MAE '].tolist()
rmse_values = results_df['RMSE'].tolist()

# Create subplots for MAE and RMSE side by side
```

```
fig, axs = plt.subplots(1, 2, figsize=(12, 5))  # Reduced width to 6
fig.suptitle('Model Performance: MAE and RMSE')

# Plot MAE
axs[0].plot(models, mae_values, marker='o', linestyle='-')
axs[0].set_ylabel('MAE')
axs[0].set_title('Mean Absolute Error')
axs[0].tick_params(axis='x', rotation=45)

# Plot RMSE
axs[1].plot(models, rmse_values, marker='o', linestyle='-')
axs[1].set_ylabel('RMSE')
axs[1].set_title('Root Mean Squared Error')
axs[1].tick_params(axis='x', rotation=45)

plt.tight_layout()
plt.show()
```

CHAPTER-5

EVALUATION AND RESULTS

5 .Results

While evaluating the output of the predictive algorithms, Food waste, several significant metrics are used to measure them performance. These include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2) Score, and Mean Absolute Error.

Table 1. Comparison of evaluation metrics of ML models

Model	MSE	MAE	R^2	RMSE
Random Forest	7.147123	1.542729	0.931051	2.67341
Gradient Boosting	8.904335	2.230108	0.914099	2.984013
LightBGM	7.482658	1.785939	0.927814	2.735445
XGBoost	7.18599	1.521036	0.930676	2.68067
Extra Trees	7.232978	1.526639	0.930223	2.68942
Bagging Regressor	7.139497	1.543352	0.931124	2.671984
Stacking Regressor	7.441588	1.559182	0.92821	2.727928

From the Table:1, it can be seen that Bagging Regressor performed well with good results in all evaluation metrics. It attained the lowest MSE and RMSE values; that means its predictions are closer to the actual food waste values. The MAE value is also competitive, suggesting that the general accuracy of model predictions is good. Among these, Bagging Regressor's R^2 value is the highest, which means it best captures the relationship between input features and food wastage, so this is one of the most reliable models within this analysis.

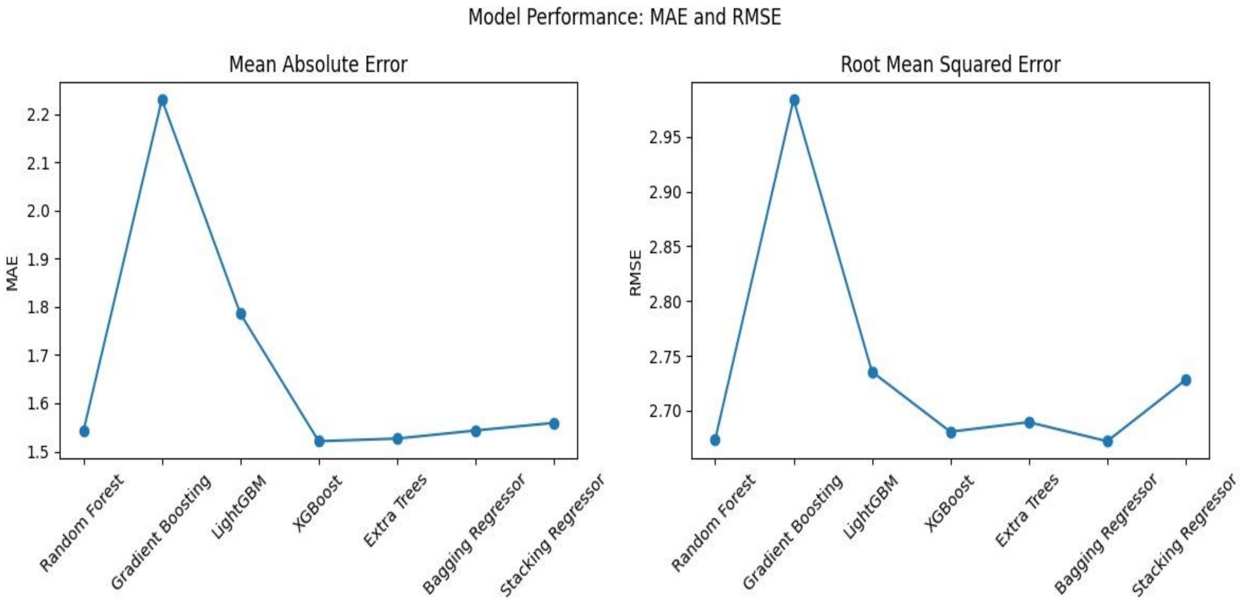

Figure 5: Scatter Plot on the Performance of Models Across Two Key Metrics: [MAE] and [RMSE]

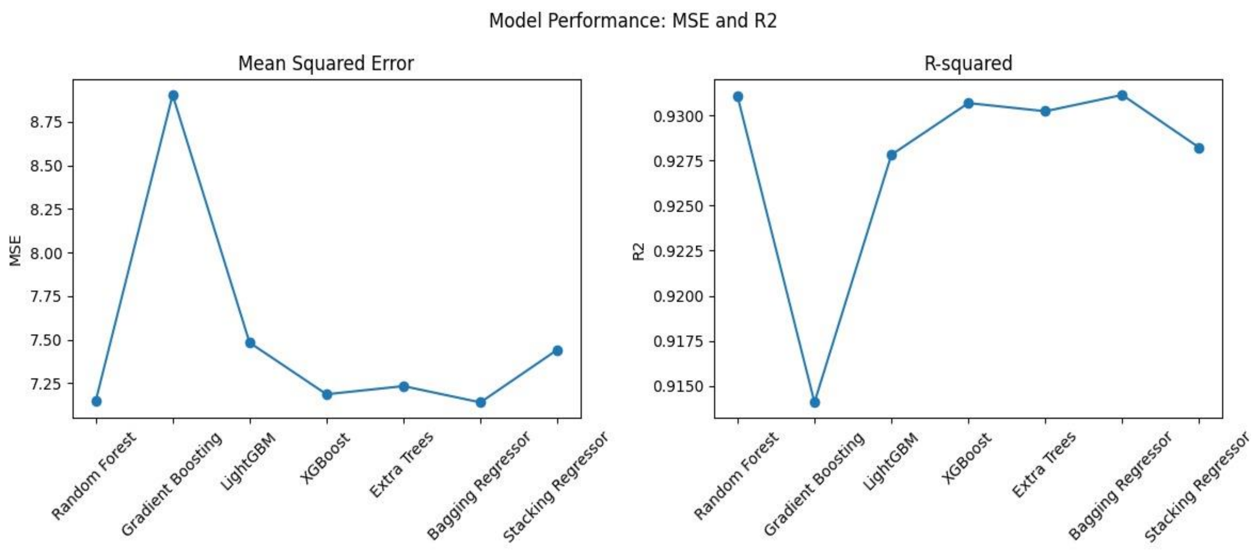

Figure 6: Scatter Plot on the Performance of Models Across Two Key Metrics: [MSE] and [[R2]

As in the plots presented in figures 6 and 7, it was found that the best performing model was Bagging Regressor. This model performed better by giving less MAE, RMSE, and MSE than other models during run, which means that, on average, it contains the smallest error of predictions. Specifically, it performed superbly concerning Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). There, the model of Bagging Regressor outperformed all the other models such as Random Forest, Gradient Boosting, and Stacking. Best models for R-squared (R^2) were developed for Bagging Regressor, XGBoost and Extra Trees. This means that those three models have fitted most of the variance in the data set concerning food waste. Gradient Boosting is the worst of all, and though Stacking was very good, it was not consistent enough as was the case for Bagging Regressor.

CHAPTER-6

CONCLUSION

6.Conclusion

The findings of this study reveal that ensemble learning models can very well predict food waste. Of these, Bagging Regressor performed the best overall, having the lowest MSE at 7.13, MAE at 1.54, and a high R^2 of 0.931, thus beating the predictive performance to be observed with it. This model performed slightly better than the Random Forest and XGBoost models that performed well too by having similar R^2 scores and even slightly lower prediction errors.

The Gradient Boosting model, though very strong, had more exaggerated values for MSE at 8.90 and MAE at 2.23, thus indicating that this approach was unable to capture the underlying patterns of the data well. The Stacking Regressor also proved to be robust as it could not outperform top models and yielded an MSE of 7.44 and R^2 score of 0.928.

CHAPTER-7

REFERENCE

7.References

[1] K. S. Chowdary, L. K. Praneetha, S. Holika, D. B. Priya, S. Venkatraman Phani Kumar and K. Venkata Krishna Kishore, ”Prediction of Food Wastage using XG Boost,” 2024 8th International Conference on Inventive 10 Systems and Control (ICISC), Coimbatore, India, 2024, pp. 307-312, doi: 10.1109/ICISC62624.2024.00059.

[2] K. Harshini, P. K. Madhira, S. Chaitra and G. P. Reddy, ”Enhanced Demand Forecasting System For Food and Raw Materials Using Ensemble Learning,” 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021.

[3] N. M. Ifham, M. I. M. Abdulla, M. R. M. Rifadh, M. N. M. Hussain Ajward, N. H. P. R. S. Swarnakantha and K. Rajendran, ”Machine Learning Based Approach to Reduce Food Wastage,” 2023 5th International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka, 2023, pp. 119-124, doi: 10.1109/ICAC60630.2023.10417516.

[4] M. M. Masuk Shopnil, A. Husna, S. Sultana and M. N. Islam, ”An Ensemble ML Model to Predict the Wastage of Food: Towards Achieving the Food Sustainability,” 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), Gazipur, Bangladesh, 2023, pp. 1-6, doi: 10.1109/NCIM59001.2023.10212669.

[5] M. J. I. Tutul, M. Alam and M. A. H. Wadud, ”Smart Food Monitoring System Based on IoT and Machine Learning,” 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), Gazipur, Bangladesh, 2023, pp. 1-6, doi: 10.1109/NCIM59001.2023.10212608.

[6] S. Yadav, M. Rajendaran, A. Nagpal, A. Karthik, K. Aravinda and B. T. Geetha, ”Exploring the Potential of Machine Learning in Enhancing Agricultural Practices and Food Production,” 2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT), Jabalpur, India, 2024, pp. 663-669, doi: 10.1109/CSNT60213.2024.10546234.

[7] K. Mogilipalem, A. Kumar Poodari, Y. Pulipati and K. B. Sangeetha, ”Food Spoilge Detection using IoT and Machine Learning,” 2024 5th International Conference for Emerging Technology (INCET), Belgaum, India, 2024, pp. 1-4, doi: 10.1109/INCET61516.2024.10593159.

[8] M. S. Aldossari, T. A. Almarshadi, R. A. Alwasel, R. S. Alqahtani and A. S. Alluhaidan, ”Waste Management System,” 2022 International Conference 11 on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2022, pp. 1293-1298, doi: 10.1109/CSCI58124.2022.00232

[9] G. Pant, P. Rawat, S. Kathuria, A. Gehlot and N. Rathor, ”Waste Management System with Technological Intervention: Advantages and Challenges,” 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2023, pp. 1461-1466, doi: 10.1109/ICESC57686.2023.10193334.

[10] Miguel Rodrigues, Vera Migueis, Susana Freitas, Telmo Machado, “Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste,” Journal of Cleaner Production, Volume 435, 2024, Article 140265, ISSN 0959-6526, doi: 10.1016/j.jclepro.2023.140265

[11]Alberto Garre, Mari Carmen Ruiz, Eloy Hontoria, “Application of Machine Learning to support production planning of a food industry in the context of waste generation under uncertainty,” Operations Research Perspectives, Volume 7, 2020, Article 100147, ISSN 2214-7160, doi: 10.1016/j.orp.2020.100147.

[12]Yasanur Kayikci, Sercan Demir, Sachin K. Mangla, Nachiappan Subramanian, Basar Koc, “Data-driven optimal dynamic pricing strategy for reducing perishable food waste at retailers,” Journal of Cleaner Production, Volume 344, 2022, Article 131068, ISSN 0959-6526, doi: 10.1016/j.jclepro.2022.131068.

[13] Freija van Holsteijn, Rene Kemna, “Minimizing food waste by improv- ´ ing storage conditions in household refrigeration,” Resources, Conservation and Recycling, Volume 128, 2018, Pages 25-31, ISSN 0921-3449, doi: 10.1016/j.resconrec.2017.09.012.

[14] Walter Kling, “Food Waste in Distribution and Use,” Journal of Farm Economics, Volume 25, Issue 4, 1943, Pages 848–859, doi: 10.2307/1231591.

[15] O. C. Tusa, “Development and analysis of static and dynamic pricing models for fresh fruit and vegetable retailers with food waste as a key consideration,” Master’s thesis, 2021.

[16]Sandeep Kumar Panda and Madhukar Dwivedi, “Minimizing food wastage using machine learning: A novel approach,” in Smart Intelligent Computing and Applications: Proceedings of

the Third International Conference on Smart Computing and Informatics, Volume 1, Springer Singapore, 2020.

[17]Violeta Stefan, Pernille Haugaard, Liisa Lahteenm¨ aki, “Determinants of˙ consumer food waste behaviour: Two routes to food waste,” Appetite, Volume 96, 2015, doi: 10.1016/j.appet.2015.08.025.

[18] S. B, S. R, S. G. R, S. P. R. V, R. P. H and S. R, "Tracking of Food Waste in Food Supply Chain Using Machine Learning," *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, Chennai, India, 2023, pp. 1-4, doi: 10.1109/ICECONF57129.2023.10083755.

[19] Jagtap, Sandeep, and Shahin Rahimifard. "The digitisation of food manufacturing to reduce waste–Case study of a ready meal factory." *Waste management* 87 (2019): 387-397.

[20] R. Akash and L. S. Anderson, "Maximizing Shelf Life Efficiency: Enhancing Sustainability through Smart Food Processing," *2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)*, Bengaluru, India, 2024, pp. 1-5, doi: 10.1109/ICETCS61022.2024.10543432.

[21] D. Dilkhush Mihirsen, J. T. Joseph and B. Renisha, "Time Series Analysis for Supply Chain Planning in Restaurants," *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Patna, India, 2020, pp. 1-5, doi: 10.1109/ICCCS49678.2020.9276872.

[22] Zarin, Fathima M., and S. Syed Abdul Hameed. "UTILISING ANALYTICAL APPROACH TO BUILD A MODEL FOCUSED ON FOOD WASTAGE PREDICTION." *International Journal on Global Business Management & Research* 8.1 (2019): 45-50.

[23] Nair, Girish. "The food we waste: antecedents of food wastage management behaviour." *International Journal of Social Economics* 48.6 (2021): 826-842.

[24] Schanes, Karin, Karin Dobernig, and Burcu Gözet. "Food waste matters-A systematic review of household food waste practices and their policy implications." *Journal of cleaner production* 182 (2018): 978-991.

[25] Hebrok, Marie, and Casper Boks. "Household food waste: Drivers and potential intervention points for design–An extensive review." *Journal of Cleaner Production* 151 (2017): 380-392.