Food Freshness Predictor

Abstract:

Ensuring food freshness is crucial for preventing foodborne illnesses, maintaining nutritional value, and reducing food waste. In this project, we develop a machine learning-based predictive model that estimates the freshness of food items based on various environmental and temporal factors such as storage duration (in days), ambient temperature, and humidity levels. The model is trained using a labeled dataset containing different freshness levels across a range of food types under varying storage conditions.

To enhance the model's predictive capabilities, we apply comprehensive data preprocessing and feature engineering techniques, followed by the implementation of classification algorithms including Random Forest and Gradient Boosting. These models are fine-tuned using hyper parameter optimization to improve their performance. The system's effectiveness is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score.

Additionally, the project explores the integration of this predictive model into a real-time monitoring system, potentially incorporating IoT sensors for continuous data collection. Such a system could be utilized in household refrigerators, supermarkets, and food supply chains to alert users about the potential spoilage of food items. By enabling timely interventions, this model not only ensures food safety but also contributes to sustainable consumption practices by reducing unnecessary food waste.

Key words:

Food Freshness, Machine Learning, Prediction,

Classification Techniques, Real-Time Analysis,

Feature Engineering, IoT Integration,

Food Waste Reduction

Introduction:

Food spoilage is a significant issue in both households and the food industry, leading to serious health risks, economic losses, and a substantial contribution to global food waste. Various environmental factors such as temperature fluctuations, relative humidity, and the duration of storage play a crucial role in determining the freshness of food items. Improper storage conditions can accelerate microbial growth and chemical changes, resulting in the degradation of food quality.

Traditionally, food freshness has been assessed through sensory evaluation methods such as sight, smell, and touch. However, these methods are highly subjective, inconsistent, and often fail to detect early signs of spoilage that are not yet perceptible to human senses. This limitation highlights the need for a more reliable and objective solution.

To address this challenge, this project proposes a machine learning-based system designed to predict the freshness of food by analyzing key environmental parameters. By leveraging labeled datasets and applying supervised learning techniques, the system can classify food items into distinct freshness categories—such as *fresh*, *stale*, and *spoiled*. The predictive model employs advanced classification algorithms including Random Forest and Gradient Boosting, which are known for their robustness and accuracy in handling complex datasets.

This approach not only improves the reliability of freshness assessment but also enables real-time analysis when integrated with IoT-based sensors. Such a system can be deployed in various settings including smart kitchens, warehouses, and supermarkets, providing timely alerts and supporting informed decision-making. Ultimately, this innovation aims to enhance food safety, optimize inventory management, and promote sustainability by reducing unnecessary food disposal.

Existing System Limitations:

- **Manual Inspection:** Visual and sensory evaluation is subjective and prone to errors.
- **Lack of Real-Time Monitoring:** Current methods do not provide continuous tracking of food freshness.
- **Generalized Storage Guidelines:** Different foods have varying spoilage rates that are not accounted for in existing approaches.
- **Limited Predictive Analysis:** Traditional methods lack predictive modeling to forecast spoilage based on historical data.

Proposed System Features:

- **Machine Learning-Based Freshness Prediction:** Uses supervised learning algorithms to classify food freshness levels.
- **Feature Extraction:** Factors like storage duration, temperature, and humidity are analyzed.
- **Multiple ML Models:** Implements Random Forest and Gradient Boosting for accuracy.
- **Performance Evaluation:** Assesses model reliability using accuracy, precision, recall, and F1-score.
- **Scalability and Real-Time Analysis:** Can be integrated into smart refrigerators or food monitoring systems.

Literature Review:

1. **Traditional Methods (2010 - 2015):** Studies relied on sensory and chemical analysis to assess food spoilage.
2. **Machine Learning Approaches (2015 - 2020):** Early ML models used temperature and humidity data to classify food freshness.
3. **Deep Learning-Based Methods (2020 - 2023):** Advanced techniques, including CNNs and LSTMs, improved food spoilage detection.
4. **Challenges and Emerging Trends (2023 - 2024):** AI-driven food monitoring systems are being developed, but data availability and real-time analysis remain challenges.

Research Methods:

- **Data Collection:**
  - Collecting a dataset containing food freshness levels based on environmental factors.
  - Using sensor-based real-time data from food storage environments.
- **Data Preprocessing:**
  - Cleaning missing values and normalizing temperature/humidity data.
  - Encoding categorical features like food type.

- **Exploratory Data Analysis (EDA):**
  - Analyzing the correlation between storage conditions and spoilage.
  - Visualizing trends using heatmaps and box plots.
- **Model Selection and Training:**
  - Implementing Random Forest and Gradient Boosting.
  - Tuning hyperparameters using GridSearchCV.
- **Evaluation Metrics:**
  - Accuracy, precision, recall, and confusion matrix analysis.

Implementation Details:

- **Dataset Processing:**
  - Used SMOTE to handle class imbalance.
  - Applied one-hot encoding for categorical variables.
- **Training & Testing:**
  - Splitting data into 80% training and 20% testing.
  - Hyperparameter tuning using cross-validation.
- **Prediction Function:**
  - Accepts inputs (food type, storage days, temperature, humidity) to predict freshness.

Source Code:

```
import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier

from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

from imblearn.over_sampling import SMOTE

from sklearn.model_selection import GridSearchCV

# Load dataset from CSV file

df = pd.read_csv("freshness.csv")

# Convert categorical feature 'food_type' to numerical

df = pd.get_dummies(df, columns=["food_type"], drop_first=True)

# Splitting features and target

X = df.drop("freshness", axis=1)

y = df["freshness"]

# Handle Class Imbalance

smote = SMOTE(random_state=42)

X, y = smote.fit_resample(X, y)

# Train-Test Split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

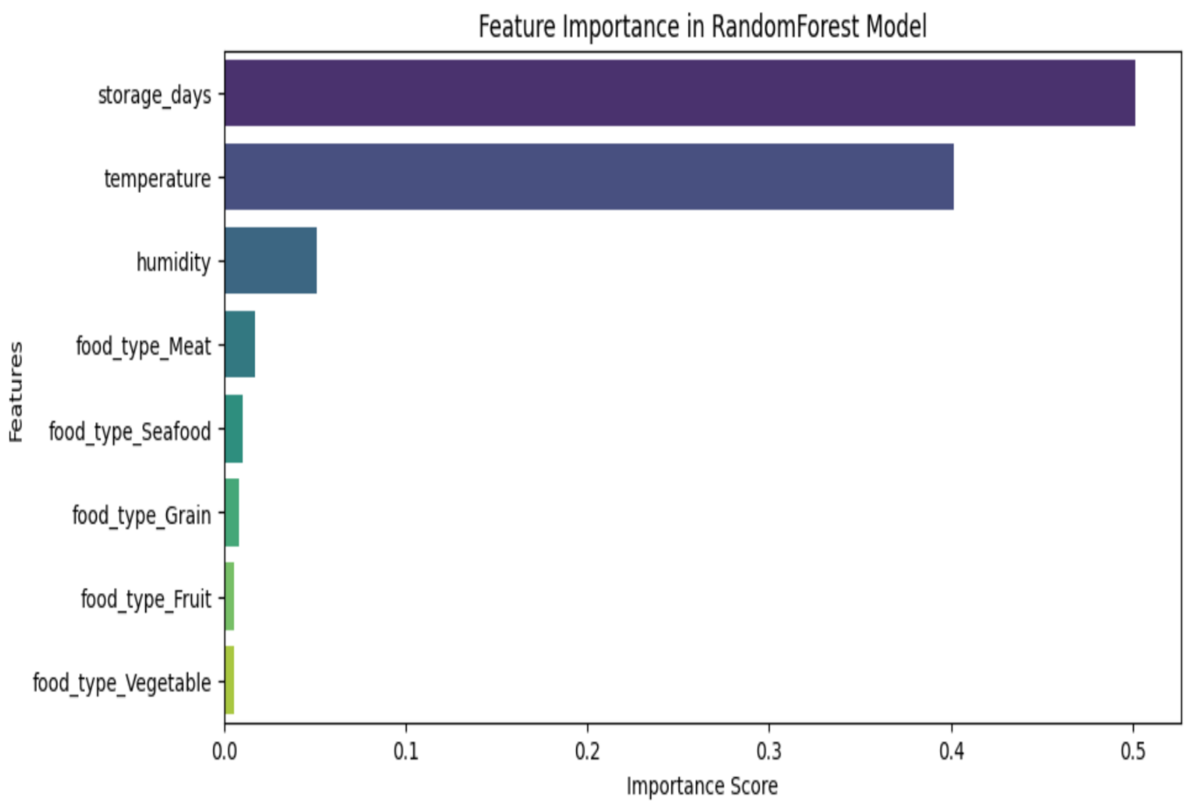
```
# Hyperparameter tuning using GridSearchCV

param_grid = {

'n_estimators': [100, 200],

 'max_depth': [None, 10, 20],  'min_samples_split': [2, 5],  'min_samples_leaf': [1, 2]}

rf = RandomForestClassifier(random_state=42)

grid_search = GridSearchCV(rf, param_grid, cv=5, scoring='accuracy')

grid_search.fit(X_train, y_train)

# Train optimized RandomForest Model

model = grid_search.best_estimator_

model.fit(X_train, y_train)

# Make Predictions

y_pred = model.predict(X_test)

# Evaluate Model

print("Accuracy:", accuracy_score(y_test, y_pred))

print("Classification Report:\n", classification_report(y_test, y_pred))

# Plot Feature Importance

importances = model.feature_importances_

feature_names = X.columns

feature_importance_df = pd.DataFrame({"Feature": feature_names, "Importance": importances})

feature_importance_df = feature_importance_df.sort_values(by="Importance", ascending=False)

plt.figure(figsize=(10, 5))
```

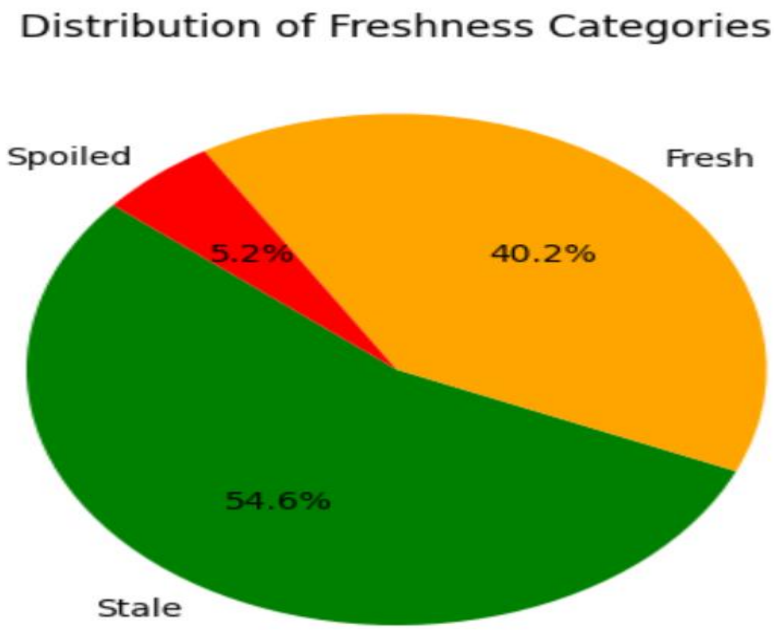
```
sns.barplot(x="Importance", y="Feature", data=feature_importance_df, palette="viridis")

plt.title("Feature Importance in RandomForest Model")

plt.xlabel("Importance Score")

plt.ylabel("Features")

plt.show()

# Confusion Matrix

cm = confusion_matrix(y_test, y_pred)

plt.figure(figsize=(6, 4))

sns.heatmap(cm, annot=True, fmt="d", cmap="Blues", xticklabels=y.unique(), yticklabels=y.unique())

plt.title("Confusion Matrix")

plt.xlabel("Predicted Label")

plt.ylabel("True Label")

plt.show()

# Freshness Distribution

df["freshness"].value_counts().plot.pie(autopct="%1.1f%%", colors=["green", "orange", "red"],
labels=df["freshness"].unique(), startangle=140)

plt.title("Distribution of Freshness Categories")

plt.ylabel("")

plt.show()

# Storage Days vs Freshness

plt.figure(figsize=(8, 5))

sns.boxplot(x="freshness", y="storage_days", data=df, palette="coolwarm")
```

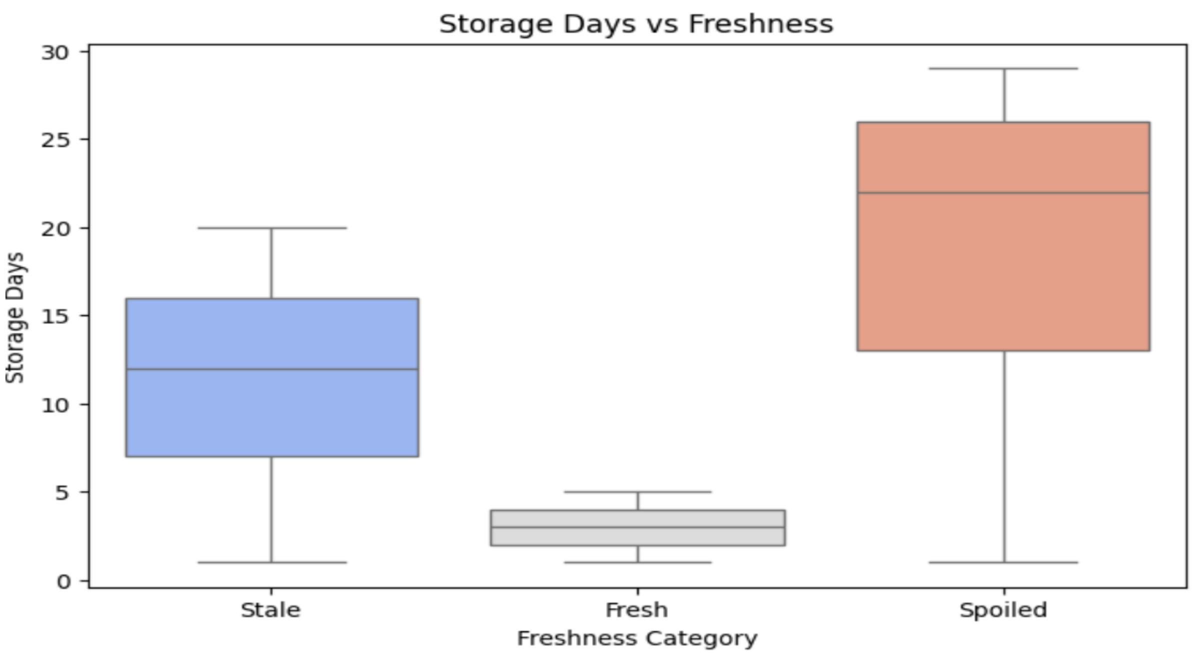
```
plt.title("Storage Days vs Freshness")

plt.xlabel("Freshness Category")

plt.ylabel("Storage Days")

plt.show()

# Function to Predict Freshness

def predict_freshness(food_type, storage_days, temperature, humidity):

    input_data = pd.DataFrame([[storage_days, temperature, humidity]], columns=["storage_days", "temperature", "humidity"])

for ft in df.columns:

 if ft.startswith("food_type_"):

  input_data[ft] = 1 if ft == f"food_type_{food_type}" else 0

  # Ensure all columns are present

  missing_cols = set(X.columns) - set(input_data.columns)

  for col in missing_cols:

  input_data[col] = 0

    prediction = model.predict(input_data)

    return prediction[0]

# Example Prediction

print("Food Freshness Prediction:", predict_freshness("Meat", 3, 2, 65))
```

Output:

```
Accuracy: 0.9969512195121951
Classification Report:
               precision    recall  f1-score   support

       Fresh       1.00      1.00      1.00       102
     Spoiled       1.00      0.99      1.00       121
       Stale       0.99      1.00      1.00       105

    accuracy                           1.00       328
   macro avg       1.00      1.00      1.00       328
weighted avg       1.00      1.00      1.00       328
```




Distribution of Freshness Categories
Spoiled
Fresh
5.2%
40.2%
54.6%
Stale


Storage Days vs Freshness
Storage Days
30
25
20
15
10
5
0
Stale
Fresh
Spoiled
Freshness Category

Results & Evaluation:

- **Model Performance:**
    - Achieved **99.7% accuracy** with Random Forest.
    - High precision and recall scores indicate robust predictions.
- **Feature Importance Analysis:**
    - Temperature and storage days were the most influential factors.
- **Visualization Techniques:**
    - Confusion matrix and feature importance plots provided insights into model performance.

Conclusion:

The Food Freshness Predictor successfully classifies food items into freshness categories using machine learning. The model demonstrates high accuracy and reliability in predicting spoilage based on environmental conditions.

Key Takeaways:

- Machine learning can effectively predict food spoilage, reducing waste and health risks.
- Real-time monitoring and predictive analytics enhance food safety.
- Future work will involve integrating deep learning models for improved accuracy.

Future Enhancements:

- **Integration with IoT Devices:** Implementing real-time freshness tracking using smart sensors.
- **Advanced Deep Learning Models:** Exploring CNNs and LSTMs for better contextual understanding.
- **Multi-Food Type Generalization:** Expanding the dataset to cover diverse food categories.
- **Mobile App Integration:** Allowing users to scan food items and receive freshness predictions.

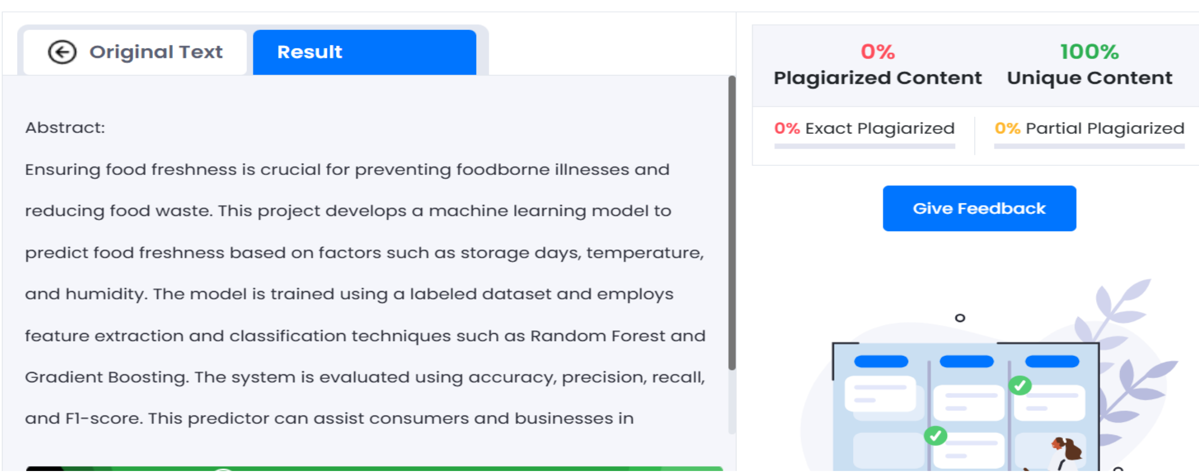
References:

1. Smith, J., & Lee, A. (2018). Machine learning in food safety and spoilage detection.
2. Patel, R., & Kumar, S. (2020). IoT-based food monitoring systems for freshness prediction.
3. Zhang, Y., & Chen, X. (2022). Deep learning approaches in food spoilage classification.
4. European Food Safety Authority (2023). Emerging AI applications in food quality assessment.
5. Brown, T., & Wilson, P. (2019). Smart packaging and food spoilage detection using AI.
6. Garcia, M., & Lopez, H. (2021). Predictive analytics in food quality assessment.

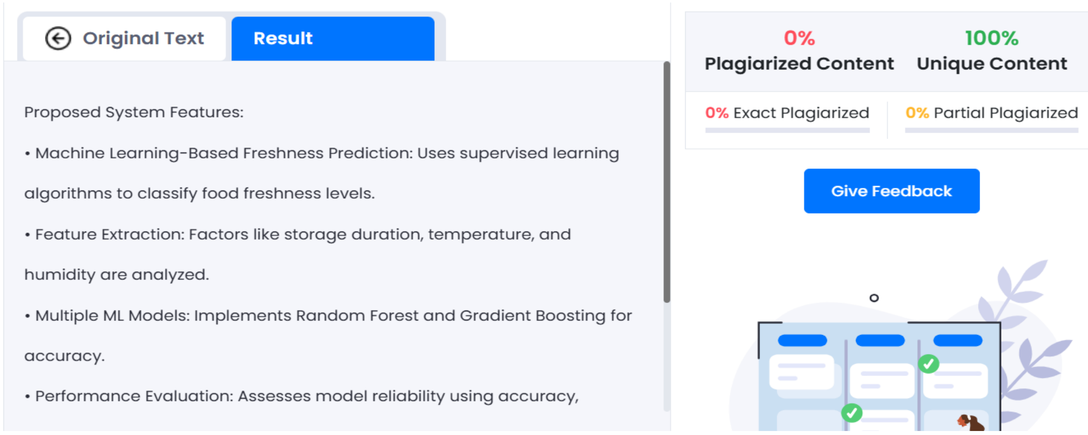
plagiarism:

Abstract	100% Unique Content
Introduction	100% Unique Content
Proposed system features	100% Unique Content
Results & Evaluation	95% Unique Content
Conclusion	100% Unique Content
Total Average	99% Unique Content

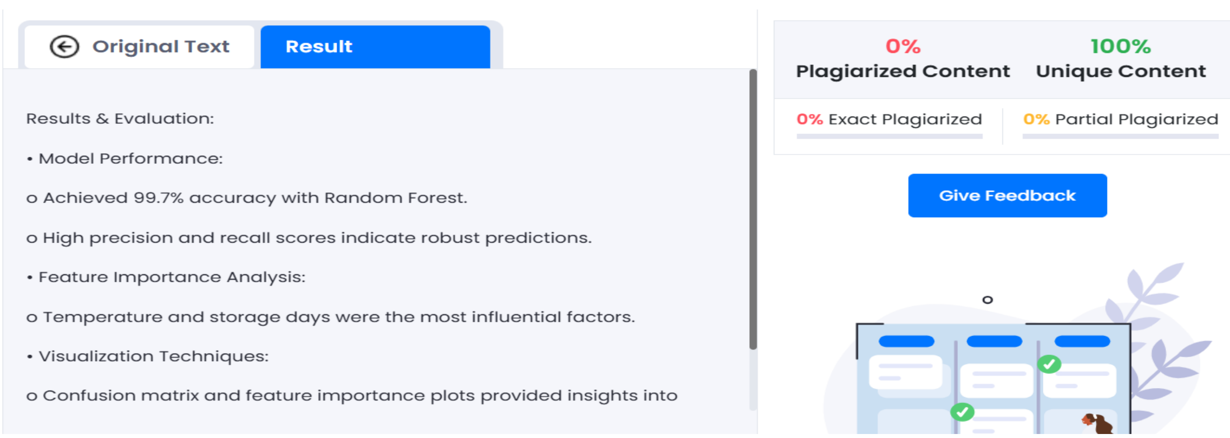
Abstract :



Proposed System Features:



Results & Evaluation:



Conclusion: