**Food & Nutrient Analysis**

**Team Members**

* Tharinika S – 22BCS130
* Samyuktha CS – 22BCS108
* Shamiha S – 22BCS116

**1. Introduction**

Food Technology and Food Chemistry are interdisciplinary fields focusing on the **composition, processing, preservation, and quality of food**.

* **Food Chemistry**: Studies carbohydrates, proteins, fats, vitamins, minerals, enzymes, and additives.
* **Food Technology**: Applies engineering and scientific principles to improve food safety, preservation, and nutrition.

This project analyzes **nutrient retention and loss** in vegetables when cooked under different conditions and applies **K-Means clustering** to group vegetables based on nutrient stability.

**2. Problem Statement**

Nutrient levels in food degrade due to:

* **Cooking methods** (boiling, steaming, frying, air frying, oven cooking).
* **Environmental factors** (temperature, humidity, oxygen exposure).

Example:

* Steaming retains most vitamins,
* Frying at high temperatures causes vitamin degradation but enhances flavor.

Thus, analyzing nutrient retention helps make **better dietary recommendations**.

**3. Data Collection & Preprocessing**

**Data Sources**

* Datasets with nutrient composition of vegetables (carbohydrates, proteins, fats, vitamins, minerals).
* Cooking conditions (temperature, time, method).

**Preprocessing Steps**

1. **Data Cleaning**: Remove errors/duplicates, handle missing values.
2. **Data Integration**: Merge multiple sheets (Sheet1, Sheet2).
3. **Data Transformation**:
   * Standardize units (mg per 100g).
   * Encode categorical columns (e.g., “Vegetable”).
   * Normalize values for clustering.

**4. Methodology**

**Statistical Approaches**

* **ANOVA** – Compare nutrient loss across cooking methods.
* **T-tests** – Compare before vs. after cooking.
* **Regression** – Predict nutrient loss trends.

**Machine Learning Approach**

* **Supervised Learning** – Predict nutrient retention.
* **Unsupervised Learning (Clustering)** – Group vegetables with similar nutrient behavior.

We chose **K-Means clustering** because:

* Works well on continuous nutrient data.
* Automatically groups vegetables based on hidden patterns.
* Scalable and fast for large datasets.
* Useful for food recommendations (e.g., vegetables with highest vitamin retention).

**5. K-Means Clustering Process**

**Step 1:** Choose number of clusters (K) → Elbow method suggested **K=4**.  
**Step 2:** Initialize centroids randomly.  
**Step 3:** Assign each vegetable to nearest centroid (based on Euclidean distance).  
**Step 4:** Update centroids as mean of cluster points.  
**Step 5:** Repeat until convergence.  
**Step 6:** Interpret clusters.

**Example Cluster Interpretation:**

* **Cluster 1**: Very high nutrient retention (best for steaming).
* **Cluster 2**: High retention (best for minimal water cooking).
* **Cluster 3**: Medium retention (suited for boiling).
* **Cluster 4**: Low retention (nutrient loss in frying).

**Flow Chart:**

A diagram of a flowchart

AI-generated content may be incorrect.

**6. Python Implementation (Sample Code)**

import pandas as pd

from sklearn.cluster import KMeans

from sklearn.preprocessing import StandardScaler

import matplotlib.pyplot as plt

# Load dataset

df = pd.read\_csv("food\_nutrients.csv")

# Select nutrient features

X = df[['Calories', 'Protein', 'Fat', 'Carbs', 'VitaminC']]

# Scale data

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

# KMeans clustering

kmeans = KMeans(n\_clusters=4, random\_state=42)

df['Cluster'] = kmeans.fit\_predict(X\_scaled)

# Visualization

plt.scatter(X\_scaled[:,0], X\_scaled[:,1], c=df['Cluster'], cmap='viridis')

plt.xlabel("Calories (scaled)")

plt.ylabel("Protein (scaled)")

plt.title("K-Means Clustering of Foods")

plt.show()

**7. Advantages of Using K-Means in Nutrient Analysis**

* **Simple & Scalable** – Works on large datasets.
* **Finds hidden patterns** – Groups vegetables with similar cooking responses.
* **Supports dietary recommendations** – Identifies vegetables best suited for nutrient retention.
* **Flexible** – Can adjust K for different levels of grouping.

**8. Applications**

* **Diet planning** – Recommend cooking methods for maximum nutrition.
* **Food research** – Identify vegetables with similar nutrient stability.
* **Consumer apps** – Help users choose healthier cooking methods.
* **Food industry** – Quality control & product development.

**9. Conclusion**

By applying **K-Means clustering** to nutrient datasets, we can identify **patterns in nutrient retention** across different cooking methods. This helps in making data-driven dietary recommendations and enhances food technology research.