Market Segmentation and Analysis of Indian Agricultural Markets

Thukkani Swetha

Ujjwal Chavan

1. Introduction

Agriculture is a crucial sector in India, contributing significantly to GDP and employment. Understanding pricing trends and market segmentation is vital for farmers, policymakers, and businesses to optimize trade, supply chains, and pricing strategies.

This study aims to segment Indian agricultural markets using **machine learning clustering techniques**—**K-Means, Hierarchical Clustering, and DBSCAN**—to identify pricing patterns and market structures across states and commodities.

Objectives

- Identify **market segments** based on pricing variables.
- Analyze the **geographic distribution** of different price segments.
- Compare **clustering techniques** to determine the most effective segmentation model.
- Provide **business recommendations** based on insights derived from the analysis.

---

2. Dataset Overview & Preprocessing

2.1 Dataset Description

The dataset contains **2,238 records**, with the following key attributes:

1. **Geographic Variables**:
    - **State** (e.g., Maharashtra, Karnataka, Tamil Nadu)
    - **District** (e.g., Pune, Bengaluru Rural)
    - **Market** (e.g., Navi Mumbai, KR Market)
2. **Product Variables**:

    - **Commodity** (e.g., Rice, Wheat, Pepper, Tomatoes)
    - **Variety** (e.g., Basmati, Hybrid, Local)
3. **Pricing Variables**:
    - **Min Price** (Minimum price at which the commodity was traded)
    - **Max Price** (Maximum price recorded)
    - **Modal Price** (Most frequently occurring price)
4. **Temporal Variable**:
    - **Arrival Date** (Date of the market entry of the product)

2.2 Data Cleaning & Preprocessing

- **Handling Missing Values**: Missing values were **imputed using median pricing** of similar commodities.
- **Outlier Removal**: Extreme pricing outliers (e.g., ₹1,10,000 for grains) were treated separately for DBSCAN.
- **Scaling**: **MinMaxScaler** was applied to normalize prices for clustering.

---

3. Clustering Methodologies

3.1 K-Means Clustering

Steps Taken:

- **Elbow method** determined **3 clusters** as optimal.
- Features used: **min_price, max_price, modal_price**.
- Clustering assigned each market to a segment.

Results:

- **Cluster 0 (Low Price Market)**: Includes **staples like rice, wheat, leafy vegetables**.
- **Cluster 1 (Mid-Price Market)**: Includes **pulses, fruits**.
- **Cluster 2 (High Price Market)**: Includes **spices like black pepper, cardamom**.

Silhouette Score: 0.764, indicating well-defined clusters.

3.2 Hierarchical Clustering

Steps Taken:

- Applied **Ward's method** to minimize intra-cluster variance.
- Constructed **dendrogram** to visualize cluster relationships.
- Used **PCA (Principal Component Analysis)** to reduce dimensions.

Results:

- Confirmed **three optimal price-based clusters**, **aligning with K-Means**.
- Strong **regional variations** in pricing were detected.

3.3 DBSCAN Clustering

Steps Taken:

- **Density-Based Clustering** applied with:
  - eps=0.5 (Neighborhood size)
  - min_samples=5 (Minimum points per cluster)
- Identified **natural clusters** without forcing a predefined number of clusters.

Results:

- **Detected outliers**, mainly in **high-priced commodities (spices, rare varieties)**.
- **Silhouette Score: 0.514**, indicating **moderate** cluster separation.

---

4. Geographic and Pricing Insights

4.1 Price Distribution by Region

- **South India (Kerala, Tamil Nadu)**:
  - Highest spice prices (Black Pepper: ₹1,20,000 per quintal).

- **Central & Northern India (Madhya Pradesh, Punjab, UP)**:
    - Moderate pricing for **grains and pulses**.
- **Islands (Andaman & Nicobar, Lakshadweep)**:
    - **Higher logistics costs** lead to inflated commodity prices.

4.2 Market Segmentation by Commodity

- **Low-Price Segments**:
    - Grains (Rice, Wheat)
    - Leafy Vegetables (Spinach, Coriander)
- **Mid-Price Segments**:
    - Pulses (Lentils, Chana)
    - Fruits (Mangoes, Bananas)
- **High-Price Segments**:
    - Spices (Black Pepper, Cardamom)
    - Dry Fruits (Almonds, Cashews)

---

5. Comparative Analysis & Findings

Clustering Method	Number of Clusters	Silhouette Score	Strengths	Weaknesses
K-Means	3	**0.764**	Well-separated clusters	Sensitive to outliers
Hierarchical Clustering	3	**0.764**	Produces consistent results	Computationally expensive
DBSCAN	Varies (Density-Based)	**0.514**	Detects outliers well	Less distinct clusters

Key Findings:

1. **K-Means & Hierarchical clustering** provided **clear segmentation**.
2. **DBSCAN** identified **price outliers**, useful for **demand forecasting**.
3. **Geographic pricing variations** impact market segmentation.

---

6. Business Recommendations

6.1 Region-Specific Pricing Strategies

📌 **Grain markets** should **focus on bulk trading strategies**.
📌 **High-priced spice markets** need **value chain optimization** to reduce costs.

6.2 Supply Chain Optimization

📌 Remote areas (**islands**) require **better transportation subsidies**.
📌 **Cold storage & logistics investments** can reduce perishable goods' price fluctuations.

6.3 Outlier Management

📌 **DBSCAN-identified outliers** can help predict **future price surges**.

6.4 Policy Recommendations

📌 **Government price monitoring systems** can help farmers avoid losses.
📌 **Market intelligence dashboards** can help traders make informed pricing decisions.

---

7. Conclusion

This study successfully segmented Indian agricultural markets using **K-Means, Hierarchical, and DBSCAN clustering**. The findings provide crucial insights for **farmers, policymakers, and businesses** to **optimize pricing, logistics, and market strategies**.

Future Scope:

- Applying **Time-Series Forecasting** for future price predictions.
- Integrating **weather patterns** to improve market segmentation models.

---

8. References

1. Kaggle Dataset: *Indian Agriculture Market Data*
2. Clustering Algorithms: *Scikit-learn Documentation*
3. Pricing Trends: *Government of India Agricultural Reports*