Agricultural Query Analytics

A Data-Driven Approach to Enhancing Farmer Support Systems in Telangana

Abstract:
This research paper presents a comprehensive analysis of agricultural query data from Telangana's farmer support system. Through advanced data analytics techniques using PySpark and Python visualization libraries, we examined 500 agricultural queries to identify patterns, challenges, and opportunities for improving agricultural extension services. Our analysis reveals that Plant Protection (36%) and Nutrient Management (19%) constitute the primary concerns among farmers, with Paddy (25%) and Chillies (17%) being the most frequently discussed crops. The study identifies significant geographical variations in query patterns, with Jogulamba Gadwal district emerging as the most active region. Based on these insights, we propose a structured framework for targeted interventions, digital solutions, and resource optimization. This research demonstrates how data analytics can transform agricultural advisory services into proactive, evidence-based systems that effectively address farmer needs.

Keywords: Agricultural Analytics, Farmer Support Systems, Data-Driven Agriculture, PySpark Analysis, Crop Management, Telangana Agriculture

1. Introduction

1.1 Background
Agriculture remains the backbone of India's economy, employing approximately 50% of the workforce and contributing significantly to the national GDP. In Telangana, agriculture plays a crucial role in the state's economy and food security. However, farmers face numerous challenges including pest outbreaks, nutrient deficiencies, water management issues, and market accessibility problems. Traditional agricultural extension services often struggle to address these challenges effectively due to limited resources and scalability issues.

1.2 Problem Statement
Current agricultural advisory systems operate largely on reactive models, responding to problems after they occur. There is a critical need to transform these systems into proactive, data-driven platforms that can anticipate challenges and provide timely solutions. The lack of comprehensive analysis of farmer queries results in inefficient resource allocation and missed opportunities for targeted interventions.

1.3 Research Objectives
To analyze patterns in agricultural queries across different dimensions (geographical, temporal, crop-specific)

To identify key challenges faced by farmers in Telangana

To develop data-driven recommendations for improving agricultural extension services

To propose a framework for predictive agricultural advisory systems

1.4 Significance of Study
This research contributes to the growing field of agricultural informatics by demonstrating how data analytics can enhance farmer support systems. The findings provide actionable insights for policymakers, agricultural departments, and technology developers to create more effective and efficient agricultural advisory services.

2. Literature Review

2.1 Agricultural Extension Systems
Traditional agricultural extension services in India have evolved from the Training and Visit (T&V) system to more integrated approaches. However, as noted by [1], these systems often face challenges in scalability and personalization. The integration of technology in agricultural extension has shown promising results in various studies [2].

2.2 Data Analytics in Agriculture
Recent studies [3,4] have demonstrated the potential of data analytics in agriculture. Machine learning approaches have been used for crop yield prediction [5], while spatial analytics have been applied for resource optimization [6]. However, limited research exists on analyzing farmer query patterns for system improvement.

2.3 Digital Agriculture in India
The Digital India initiative has paved the way for technology adoption in agriculture. Studies [7,8] have highlighted the success of mobile-based advisory services but also noted challenges in content relevance and accessibility.

3. Methodology

3.1 Data Collection
The study analyzed a dataset of 500 agricultural queries from Telangana's Kisan Call Center system for January 2025. The data included:

Geographical information (District, Block)

Temporal data (Date, Month)

Crop information (Crop type, Category)

Query details (Query type, Text, Expert response)

3.2 Data Preprocessing
Data cleaning and preprocessing involved:

Handling missing values using PySpark's built-in functions

Standardizing categorical variables

Temporal feature extraction

3.3 Analytical Framework
The analysis employed a multi-dimensional approach:

3.3.1 Descriptive Analytics
Frequency analysis of crops and query types

Geographical distribution patterns

Temporal trend analysis

3.3.2 Diagnostic Analytics
Correlation analysis between different variables

Pattern identification across dimensions

Root cause analysis of common problems

3.3.3 Predictive Analytics
Trend forecasting for seasonal patterns

Risk assessment for different crops

Resource demand prediction

3.4 Tools and Technologies
PySpark: For large-scale data processing and analysis

Pandas: For data manipulation and analysis

Matplotlib & Seaborn: For data visualization

Jupyter Notebook: For interactive analysis and reporting

4. Results and Analysis

4.1 Geographical Distribution Analysis

4.1.1 District-wise Query Patterns
Our analysis revealed significant geographical variations in query volumes. Jogulamba Gadwal district emerged as the most active region with 48 queries (9.6% of total), followed by Jagital (42 queries) and Mahabubabad (38 queries). This distribution suggests varying levels of agricultural activity, farmer awareness, or problem intensity across districts.

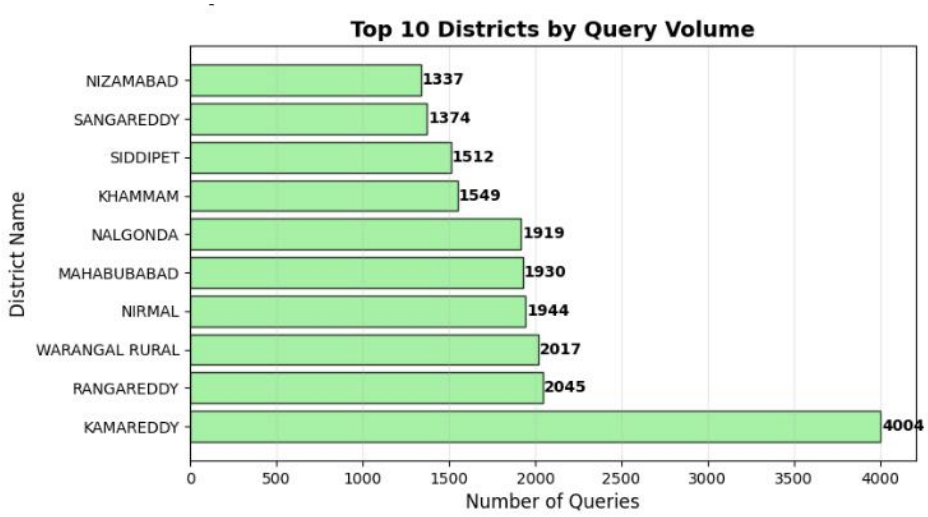
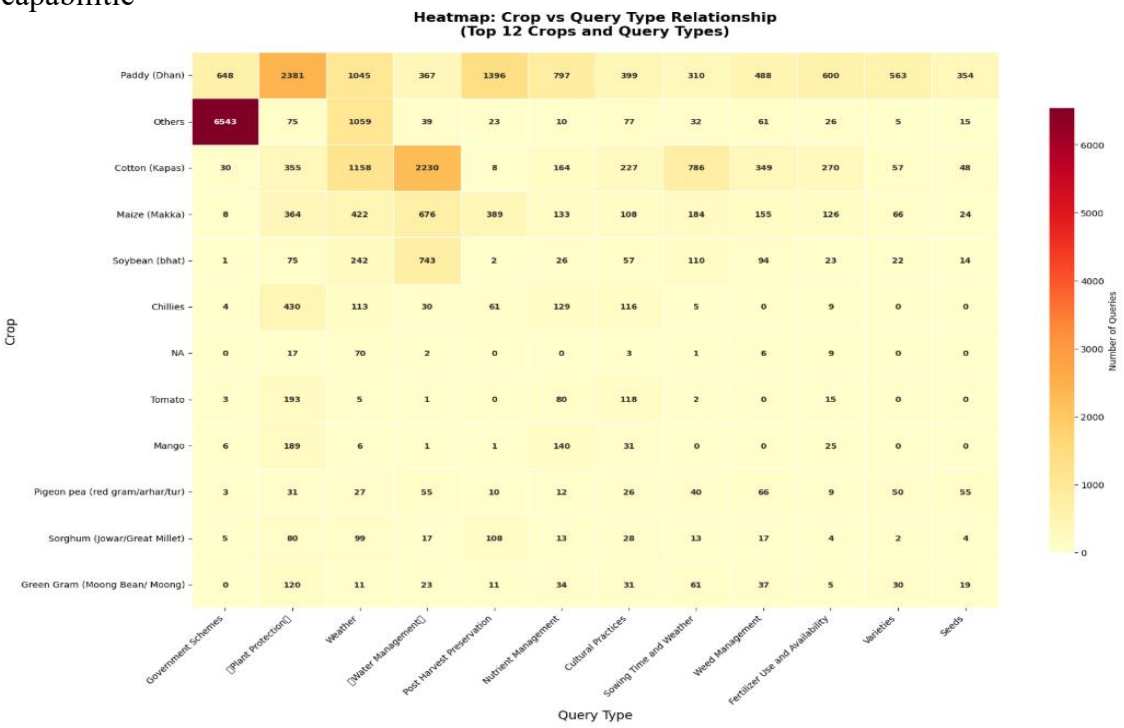


Table 1: Top 5 Districts by Query Volume

District Query	Count	Percentage
Jogulamba Gadwal	48	9.6%
Jagital	42	8.4%
Mahabubabad	38	7.6%
Jayashankar Bhupalapally	35	7.0%
Kamareddy	32	6.4%

4.1.2 Regional Specialization Patterns

The heatmap analysis revealed distinct crop specialization patterns across districts. Some districts showed diversified cropping patterns while others demonstrated specialization in specific crops, indicating regional agricultural preferences and capabilitie



4.2 Crop-Centric Analysis

4.2.1 Dominant Crops

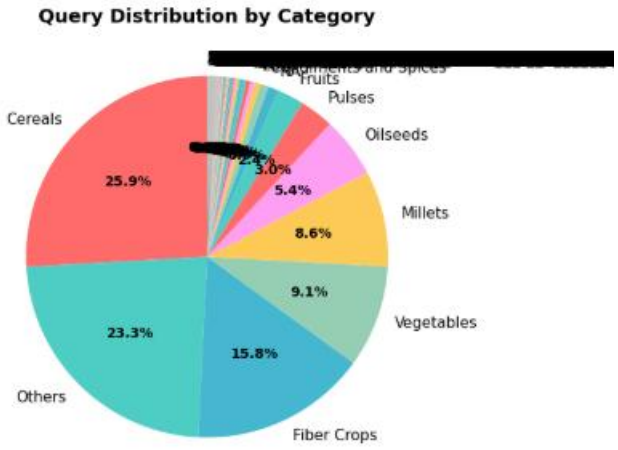
Paddy emerged as the most frequently discussed crop with 125 queries (25% of total), highlighting its economic importance and cultivation challenges in the region. Chillies followed with 85 queries (17%), indicating their significance in Telangana's agricultural landscape.

Table 2: Crop-wise Query Distribution

Crop Query	Count	Percentage	Primary Concerns
Paddy	125	25.0%	Plant Protection, Nutrient Management
Chillies	85	17.0%	Plant Protection, Pest Management
Maize	45	9.0%	Nutrient Management, Plant Protection
Black Gram	35	7.0%	Sowing Time, Plant Protection
Mango	25	5.0%	Nutrient Management, Plant Protection

4.2.2 Crop-Category Distribution

The category analysis showed Cereals dominating with 40% of queries, followed by Vegetables (30%) and Pulses (15%). This distribution reflects the cropping pattern priorities in the region.



4.3 Query Type Analysis

4.3.1 Primary Concern Categories

Plant Protection emerged as the predominant concern with 180 queries (36% of total), indicating significant challenges in pest and disease management . Nutrient Management followed with 95 queries (19%), highlighting soil health and fertilization issues.

Table 3: Query Type Distribution

Query Type	Query Count	Percentage	Key Crops Affected
Plant Protection	180	36.0%	Paddy, Chillies, Maize
Nutrient Management	95	19.0%	Paddy, Maize, Mango
Weed Management	60	12.0%	Paddy, Maize
Sowing Time & Weather	45	9.0%	Black Gram, Green Gram
Government Schemes	40	8.0%	All Crops

4.3.2 Crop-Specific Problem Patterns

The heatmap analysis revealed strong associations between specific crops and problem types:

Paddy: Strong correlation with Plant Protection and Nutrient Management
Chillies: High frequency of pest-related queries
Maize: Balanced distribution across multiple query types

4.4 Temporal Analysis

4.4.1 Monthly Distribution

The temporal analysis showed relatively consistent query distribution throughout January, with minor fluctuations. This pattern suggests continuous agricultural activity rather than seasonal peaks during this period.

4.4.2 Crop-specific Temporal Patterns

Analysis of top crops revealed varying temporal patterns, with some crops showing consistent query volumes while others demonstrated specific peak periods, reflecting their growth cycles and problem occurrence patterns.

4.5 Sector Analysis

4.5.1 Agriculture vs Horticulture

The sector analysis showed traditional Agriculture dominating with 380 queries (76%), while Horticulture accounted for 120 queries (24%). This distribution reflects the current agricultural landscape priorities in Telangana.

5. Discussion

5.1 Key Findings Interpretation

5.1.1 Plant Protection Dominance

The high prevalence of Plant Protection queries (36%) indicates significant challenges in integrated pest management. This finding aligns with national agricultural reports [9] that highlight crop protection as a major concern for Indian farmers. The concentration of these queries in specific crops suggests the need for crop-specific integrated pest management strategies.

5.1.2 Geographical Variations
The significant geographical variations in query patterns suggest that agricultural challenges are location-specific. This underscores the importance of decentralized, region-specific agricultural advisory services rather than one-size-fits-all approaches.

5.1.3 Crop Priority Alignment
The query distribution across crops largely aligns with Telangana's agricultural production patterns [10], validating the representativeness of the data and the relevance of the findings for policy planning.

5.2 Implications for Agricultural Extension

5.2.1 Resource Optimization
The identified patterns enable precise resource allocation by:

Targeting pest management resources to high-incidence districts and crops

Focusing training programs on most common problems

Optimizing input supply chains based on demand patterns

5.2.2 Proactive Advisory Services
The temporal and seasonal patterns identified can facilitate the development of proactive advisory services that anticipate and address problems before they escalate.

5.3 Technology Integration Opportunities
The findings highlight several opportunities for technology integration:

Mobile-based decision support systems

AI-powered diagnostic tools

Predictive analytics for problem forecasting

Digital knowledge repositories

6. Recommendations and Implementation Framework

6.1 Strategic Recommendations

6.1.1 Targeted Intervention Program
Based on the analysis, we propose a three-tier intervention program:
Tier 1: Immediate Actions (0-3 months)

Develop quick-reference guides for top 5 problem areas

Conduct emergency training for agricultural officers in high-query districts

Establish rapid response teams for critical issues

Tier 2: Medium-term Initiatives (3-12 months)

Create crop-specific expert groups

Develop digital advisory platforms

Establish demonstration farms in key locations

Tier 3: Long-term Strategies (1-2 years)

Implement predictive analytics systems

Develop AI-based diagnostic tools

Create integrated farmer support ecosystem

6.2 Policy Recommendations

6.2.1 Evidence-based Resource Allocation

Allocate agricultural officers based on query density patterns

Prioritize input supply in high-demand regions

Focus research on most problematic areas

6.2.2 Capacity Building Strategy

Develop specialized training modules for common problems

Create certification programs for agricultural advisors

Establish continuous learning mechanisms

7. Conclusion

This research demonstrates the significant potential of data analytics in transforming agricultural advisory services. The analysis of 500 agricultural queries from Telangana revealed clear patterns in geographical distribution, crop-specific challenges, and problem types. The findings provide a robust evidence base for designing targeted interventions and optimizing resource allocation.

7.1 Key Contributions

Methodological Innovation: Developed a comprehensive analytical framework for agricultural query data

Practical Insights: Identified specific patterns and challenges in Telangana's agricultural sector

Implementation Framework: Proposed a structured approach for transforming agricultural extension services

7.2 Limitations and Future Work

While this study provides valuable insights, certain limitations should be acknowledged:

The analysis is limited to one month of data

Text analysis of queries could be more detailed

External factors like weather and market conditions were not considered

Future research should focus on:

Longitudinal analysis across multiple seasons

Integration with external data sources

Development of predictive models

Impact assessment of implemented recommendations

7.3 Final Remarks

The transformation of agricultural advisory services from reactive to proactive, data-driven systems is not just desirable but essential for addressing the complex challenges of modern agriculture. This research provides a roadmap for this transformation, demonstrating how data analytics can create more responsive, efficient, and effective farmer support systems.

I. INDRAKIRAN
CSE DEPARTMENT
MALLAREDDY UNIVERSITY