International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.2 (2025) pp. 1781-1793
http://www.ijcesen.com

Research Article

ISSN: 2149-9144

Data-Driven Insights: A Critical Analysis of Farmer Call Centre Data Using Machine Learning Techniques

C. Kiruthiga[1]*, K. Dharmarajan[2]

[1]Research Scholar, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Assistant Professor, Dwaraka Doss Goverdhan Doss Vaishnav College, Affiliated to University of Madras, Chennai
* **Corresponding Author Email:** csnkirthi@gmail.com - **ORCID:** 0009-0005-7778-7264

[2]Professor, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai. India
Email: dharmak07@gmail.com - **ORCID:** 0000-0001-8493-1301

Article Info:

DOI: 10.22399/ijcesen.1348
Received : 05 January 2025
Accepted : 07 March 2025

Keywords :

Kishan Call Center,
Machine Learning algorithms,
Natural Language Processing,
Query Processing,
Visualization.

Abstract:

The agricultural sector plays a crucial role in India's economy, society, and environment. Agriculture is the primary source of livelihood for a significant portion of the Indian population, employing over half of the country's workforce. It contributes substantially to the Gross Domestic Product (GDP) and remains a vital sector for rural development and poverty alleviation. Experts use different kinds of smart systems to figure out problems on farms and find possible solutions. The systems help the experts collect and analyze information regarding the issues farmers meet. This study aimed to investigate the query data from Kisan Call Centers (KCCs) from 2020 to 2023 to identify key issues, understand farmers' challenges, and provide data-driven policy and program development insights. Python was used for data processing, Power BI for visualization, and Machine learning algorithms and Natural Language Processing libraries for query analysis.

1. Introduction

The Ministry of Agriculture and Farmers Welfare, Government of India, introduced the "Kisan Call Centres" (KCC) programme in January 2004, an innovative and modern initiative to provide extension information and support to farmers. This includes assistance with crop management, livestock care, pest control, soil health, agricultural technology, market trends, and regulatory compliance, among other topics. This specialized call center provides a dedicated platform for farmers to connect with farming experts, advisors, and professionals, who can offer guidance, solutions, and resources tailored to their needs.
The agriculture sector has witnessed significant growth, driven by advancements in urbanization and agricultural policies. Policymakers primarily focus on increasing food production and ensuring improved market access while addressing the needs of other sectors. Technological advancements have had a transformative impact on agriculture, with innovations in IoT, data analytics, machine learning, and deep learning contributing significantly. Internet of Things (IoT) devices enable smart farming by collecting and analyzing farm data. Farmers often rely on expert guidance and support to enhance their farming practices.

2. Literature Survey

The Kishan Call Center (KCC) has played a crucial role in supporting Indian farmers, offering guidance across various agricultural aspects. Studies analyzing its impact across multiple states have revealed diverse benefits and challenges.
In Telangana's Mahaboobnagar district, farmers reported several advantages. These included financial savings through cost-effective farming practices, technological support for pest and disease outbreak predictions, and improved communication via user-friendly services. According to the analysis, 15.56% of respondents experienced significant benefits, 60% reported moderate benefits, and 24.44% indicated minimal benefits. The research also highlighted a statistically significant positive relationship between the use of mobile phones in farming and the perceived benefits of the KCC services [1-3]. Similarly, in Jabalpur district, Madhya Pradesh, the majority of

farmer queries during the study period focused on Kharif crops, reflecting the seasonal demand for information[4]. In Tamil Nadu's Manachanallur Block, farmers emphasized the need to simplify technical phrases into local languages and proposed the addition of feedback systems and conferencing tools for addressing critical agricultural issues[5]. In Uttar Pradesh's Ayodhya district, an analysis of responses from ten villages revealed that 57.33% of calls were routed to other states, while there was an 85.33% improvement in KCC service efficiency[6]. The integration of advanced technologies has significantly enhanced KCC services. Machine learning (ML), a transformative tool in modern agriculture, has enabled more efficient and sustainable practices. For example, a multi-criteria decision-making method (TOPSIS) was applied to analyze helpline records, identifying key correlations among farmers' concerns and providing policymakers with actionable insights, achieving 73.21% confidence[2].

Natural Language Processing (NLP) has also proven instrumental in analyzing unstructured data, uncovering patterns, trends, and frequently asked queries. Techniques such as MapReduce and DBSCAN have been utilized to cluster these queries, providing valuable insights into farmer needs [7-9]. Deep learning models have further revolutionized forecasting and query resolution. Temporal Convolutional Networks (TCN) demonstrated exceptional performance in predicting mushroom crop yields, achieving an AMSE of 0.0025 and an AMAE of 0.0230[7]. For rice crops, deep learning models outperformed machine learning approaches, with an AMSE of 0.0036 and an AMAE of 0.0225[8]. Query resolution systems, such as KisanQRS, utilizing LSTM-based models, achieved high accuracy (97.52%), F1-score (96.58%), and NDGC (96.20%)[10]. Similarly, the AgriResponse system, featuring three NLP-based models, achieved a top accuracy of 97.12%, showcasing its efficiency in handling farmer queries [11-15].

Despite its notable advantages, KCC faces several challenges, including language barriers and a lack of comprehensive feedback mechanisms. Recommendations to improve its services include translating technical terms into local languages, integrating feedback and conference facilities, and leveraging cutting-edge technologies like ML, NLP, and deep learning for better service delivery. Overall, KCC's adoption of advanced computational methods has greatly enhanced its ability to address farmer concerns, offering a promising pathway for innovation and sustainable agricultural development in India.

3. Methodology

The research workflow was designed to systematically analyze farmer queries and evaluate the effectiveness of Kishan Call Centers (KCC) in addressing agricultural challenges. The methodology focuses on data collection, preprocessing, visualization, and analysis to uncover trends, patterns, and actionable insights.

3.1 Data Collection

Kishan Call Center, daily call logs corresponding to districts of each state in India. The Kishan call center dataset has been downloaded from https://kcc-chakshu.icar.gov.in/. Furthermore, the call records used in this study are from January 2020 to December 2023. The call log contains 35 million data files that have been downloaded. 667118 raw data belonging to Tamil Nadu from call logs were considered for this research. Call logs in CSV File format including 11 attributes details of the attributes are shown in Table 1.

Table 1. *Kishan Call Center Dataset Attributes*

S.No.	Attribute Name	Description
1	BlockName	Block name of farmer
2	DistrictName	District name of farmer
3	StateName	State name of farmer
4	CreatedOn	Year, month, date and
5	Season	Season of year
6	Category	Query Category
7	Crop	Target crop of query
8	QueryType	Type of Query
9	Sector	Target sector of query
10	QueryText	Query in textual format
11	KccAns	Query Response

3.2 Data Preprocessing

Data pre-processing involves converting raw data into an intelligible format. Real-world data is often deficient, and inconsistent and is likely to carry many inaccuracies. Data pre-processing is used for resolving such problems. In this study, we have employed the following data preprocessing steps to improve the quality of data. This has been done using Python programming in Google Co-lab. Data cleaning involves addressing missing values, resolving blank entries, and correcting data inconsistencies. Various methods are used to handle missing data, such as removing incorrect state entries, replacing blank values in the "block name" attribute with the most frequently occurring value (mode) within the district, and ignoring certain tuples. Misspellings in the "block name" or "district" fields are corrected to ensure accuracy.

Data integration involves merging data from various sources into a cohesive, unified view. Based on specific requirements, attributes are extracted from the original tables, preprocessed, and combined for further analysis. For instance, in call volume analysis, the year, month, and day attributes are consolidated into a single "date" attribute. Properly integrating data reduces redundancies and inconsistencies, enhancing the efficiency and speed of subsequent data mining processes. The State of Tamil Nadu is the tenth largest state in India by geographical area. TN has 38 districts and 385 blocks. The land area has been classified into seven agro-climatic zones based on soil characteristics, rainfall distribution, irrigation pattern, cropping pattern, and other ecological and social factors. The following are the seven agro-climatic zones of the State. 1) North Eastern 2) North Western 3) Western 4) Cauvery Delta Zone (CDZ) 5) Southern 6) High Rainfall 7) Hilly and High Altitude. A new Column called Agro-Climate has been added, and the districts and blocks have been categorized based on the Tamilnadu Agroclimatic zone. The process of data selection involved narrowing down the dataset to 11 attributes from an initial set of 13 for detailed analysis. Attributes such as "Ans Key," which contained symbol entries, and "Category," which had null values, were excluded. The “Created On” attribute was retained to capture the dates on which calls were made. The analysis primarily targeted calls related to crops, query types, and seasonal trends. Before Preprocessing: The preprocessing steps aim to clean the data, ensuring that only relevant and accurate information is included for further analysis or model training. This also includes converting data into appropriate formats and enriching the dataset with new features such as season and Agro climate, based on domain knowledge. Sample data are shown in table 2.

Table 2. *Sample Data*

Attribute	**Sample Values**
Block Name	Uttirameru
DistrictName	Kancheepuram
Season	Rabi
Agro Climate	North Eastern
Crop	Paddy Dhan
QueryType	Weed Management
Sector	Agriculture
QueryText	Asking about Weed management in paddy

Figure 1 is block diagram of the detailed analysis of the KCC dataset workflow and figure 2 is districts and agroclimatic zone. Figure 3 shows dataset before preprocessing and figure 4 shows algorithm for dataset preprocessing. The preprocessing algorithm begins with loading the KCC dataset, which is stored in a CSV file. The algorithm defines a function discussed in data cleaning, data integration and data selection. This process ensures that the data is in a consistent, meaningful, and usable format for the next steps in the workflow. After Preprocessing: The dataset, has been cleaned, transformed, and enriched for analysis or model training, with improved data quality and added insights. Figure 5 is dataset after preprocessing and figure 6 is visualizing the dynamics of call counts over the years. Figure 7 shows breakdown of call volume: crop wise and season wise insights and figure 8 shows results of unique query type. Figure 9 is top 5 query types vs season and figure 10 is trends in call counts across years and seasons.

3.3 Results and Discussion

The Kishan Call Center (KCC) serves as a vital platform for farmers, providing expert advice, information on government schemes, and solutions to agricultural challenges. By analysing the data generated through these interactions, we can gain valuable insights into the concerns and needs of farmers, identify trends, and improve the services offered by the KCC.

Call Volume vs Season

The graph categorizes the queries by season—Kharif, Rabi, and Zaid—across different years from 2020 to 2023. Rising Demand: The increasing trend in the number of queries indicates that more farmers are seeking assistance from KCCs, possibly due to growing awareness and availability of these services. Season-Specific Support: Kharif Season: With the highest number of queries, there may be a need for more resources and focused support during this season.Rabi and Zaid Seasons: While queries are fewer for these seasons than for Kharif, the increasing trend suggests that these seasons also require substantial support. The highest call averages were Rabi (261722), Kharif (243220), and Zaid (158368).

Call Volume vs Sector

The Sector attribute consists of four values: Agriculture, Horticulture, Animal Husbandry, and Fisheries. Among these, the Agriculture sector recorded 466,179 calls, and the Horticulture sector received the highest number with 1,872,208 calls.

In comparison, Animal Husbandry and Fisheries had significantly fewer calls, with 8,822 and 1,101 respectively, making Fisheries the sector with the lowest call records.

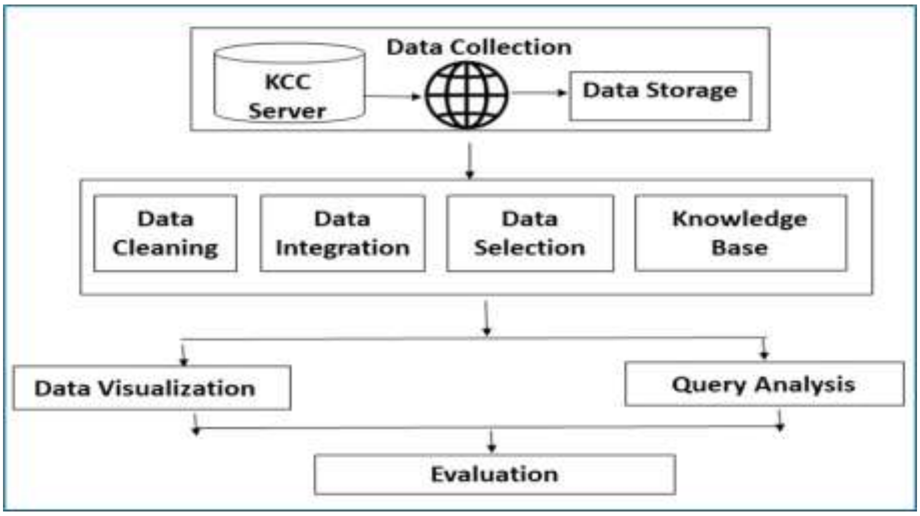

Figure 1. *Block diagram of the detailed analysis of the KCC dataset workflow*

Figure 2. *Districts and Agroclimatic Zone*

```
Attribute: BlockName, Number of items: 703973
Attribute: Category, Number of items: 0
Attribute: Year, Number of items: 705214
Attribute: Month, Number of items: 705214
Attribute: Day, Number of items: 705214
Attribute: Crop, Number of items: 705214
Attribute: DistrictName, Number of items: 705214
Attribute: QueryType, Number of items: 705214
Attribute: Season, Number of items: 0
Attribute: Sector, Number of items: 705214
Attribute: StateName, Number of items: 705214
Attribute: QueryText, Number of items: 705203
Attribute: KccAns, Number of items: 624873
```

Figure 3*. Dataset Before Preprocessing*

Algorithm: Dataset Preprocessing for KCC Data

Algorithm 1 Preprocess KCC Dataset

1: **Input:** KCC Dataset
2: **Output:** Processed Dataset
3: **Step 1: Load the CSV file**
Load the dataset using pandas:

$$df = pd.read\_csv('path/to/input.csv')$$

4: **Step 2: Handle Missing and Invalid Values in `BlockName`**
Replace '0' in `df['BlockName']` with NaN:

$$df['BlockName'] = df['BlockName'].replace('0', NaN)$$

Define a function to fill `BlockName` with the mode value within the same `DistrictName`:

$$fill\_block\_mode(row) = \begin{cases} mode(df['BlockName'] \mid df['DistrictName'] = row['DistrictName']) & \text{if NaN} \\ row['BlockName'] & \text{otherwise} \end{cases}$$

Apply the function to update `BlockName`:

$$df['BlockName'] = df.apply(fill\_block\_mode, axis=1)$$

5: **Step 3: Remove Rows with Invalid or Missing `Crop`**

$$df = df[df['Crop'] \neq' 0']df[df['Crop'] \neq' 0']_{df[df['Crop'] \neq' 0']_{df[df['Crop'] \neq' 0']}}$$

6: **Step 4: Drop the `Category` Column if it Exists**

$$\text{if 'Category' in df.columns:} \quad df = df.drop('Category', axis=1)$$

7: **Step 5: Remove Rows with Year = 2024**

$$df = df[df['Year'] \neq 2024]df[df['Year'] \neq 2024]_{df[df['Year'] \neq 2024]_{df[df['Year'] \neq 2024]}}$$

8: **Step 6: Create `season` Column Based on `Month`**
Define a function to determine the season:

$$determine\_season(month) = \begin{cases} 'kharif' & \text{if } month \in \{6, 7, 8, 9, 10\} \\ 'rabi' & \text{if } month \in \{11, 12, 1, 2\} \\ 'zaid' & \text{otherwise} \end{cases}$$

Apply the function to `df['Month']`:

$$df['season'] = df['Month'].apply(determine\_season)$$

9: **Step 7: Create `Agro_climate` Column Based on `DistrictName`**
Define lists for each agro-climatic region:

$$north\_eastern, cdz, high\_rainfall, hilly\_high\_altitude, western, southern$$

Define a function to assign the agro-climate region:

$$determine\_agro\_climate(district) = \begin{cases} 'North\ Eastern' & \text{if } district \in north\_eastern \\ 'CDZ' & \text{if } district \in cdz \\ 'High\ Rainfall' & \text{if } district \in high\_rainfall \\ 'Hilly\ and\ High\ Altitude' & \text{if } district \in hilly\_high\_altitude \\ 'Western' & \text{if } district \in western \\ 'Southern' & \text{if } district \in southern \\ 'North\ Western' & \text{otherwise} \end{cases}$$

2

Apply the function to `df['DistrictName']`:

$$df['Agro\_climate'] = df['DistrictName'].apply(determine\_agro\_climate)$$

Figure 4. *Algorithm for dataset preprocessing*

```
  data = pd.read_csv(file_path)
Attribute: Block, Number of items: 663310
Attribute: Year, Number of items: 663310
Attribute: Crop, Number of items: 663310
Attribute: DistrictName, Number of items: 663310
Attribute: QueryType, Number of items: 663310
Attribute: Season, Number of items: 663310
Attribute: Agro climate, Number of items: 663310
Attribute: Sector, Number of items: 663310
Attribute: QueryText, Number of items: 663310
```

Figure 5. *Dataset after Preprocessing*

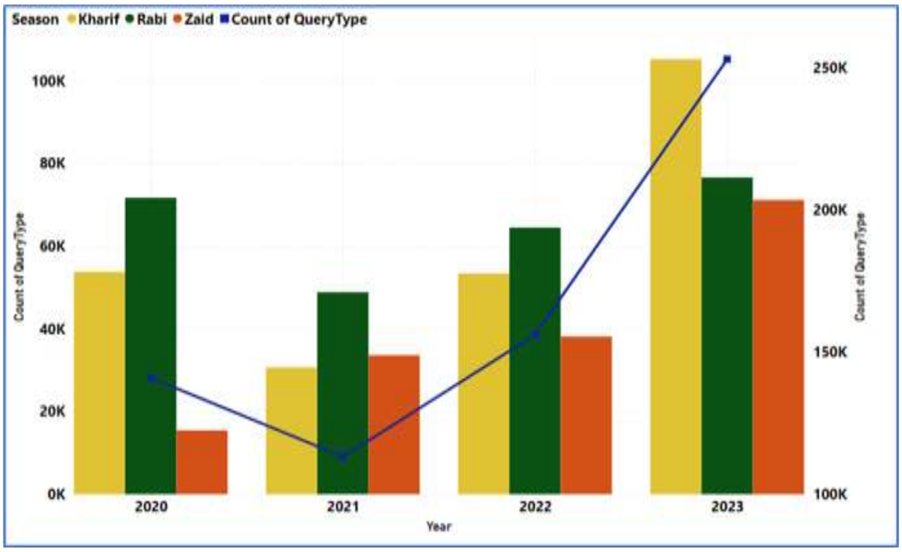

Figure 6. *Visualizing the Dynamics of call counts over the years*

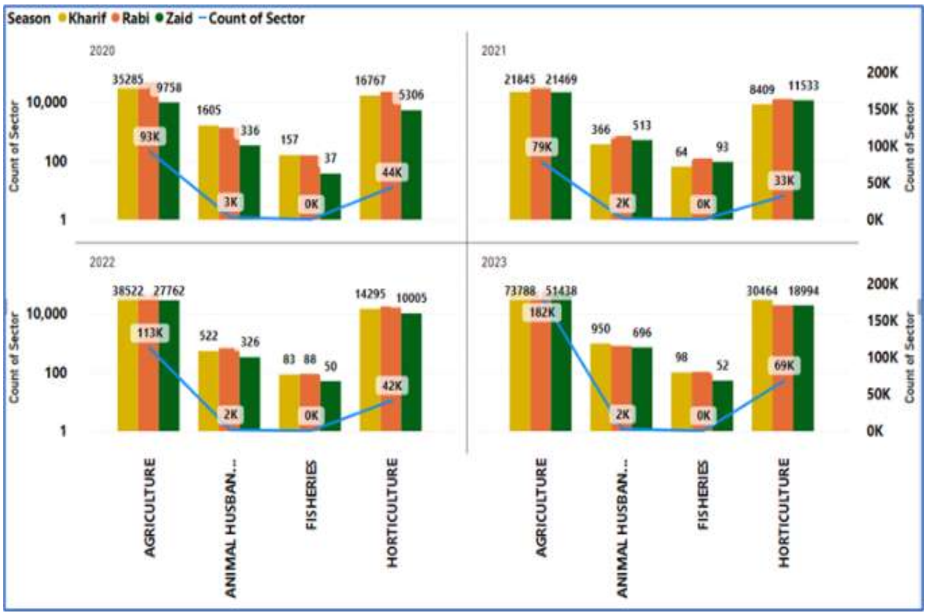

Figure 7. *Breakdown of Call Volume: Crop wise and Season wise Insights*

```
['Weed Management' 'Plant Protection' 'Weather' 'Nutrient Management'
 'Varieties' 'Seeds' 'Sowing Time and Weather' 'Cultural Practices'
 'Market Information' 'Bio-Pesticides and Bio-Fertilizers'
 'Government Schemes' 'Fertilizer Use and Availability' 'Crop Insurance'
 'Animal Production Piggery Goatery Sheep Farming etc'
 'Training and Exposure Visits' 'Seeds and Planting Material'
 'Cattle shed Planning and Management' 'Disease Reporting'
 'Agriculture Mechanization' 'Disease Management' 'Organic Farming'
 'Field Preparation' 'Dairy Production' 'Water Management'
 'Soil Health Card' 'Soil Testing'
 'Vegetative Propagation and Tissue Culture' 'Credit' 'Floriculture'
 'Breeding of freshwater prawn' 'Fishery Nutrition'
 'Fish Fingerling Production' 'Water Management Micro Irrigation'
 'Integrated Farming' 'Livestock Products Processing and Packaging'
 'Medicinal and Aromatic Plants' 'Landscaping' 'Nursery Management'
 'Beekeeping' 'Animal Nutrition' 'Tank Pond and Reservoir Management'
 'Animal Breeding' 'Power Roads etc' 'Post Harvest Preservation'
 'Water Testing for Fish Production' 'Freshwater Pearl Farming'
 'OldSenile Orchard Rejuvenation' 'Spices and Condiment Crops'
 'Post Harvest Management Cleaning Grading Packaging Food Processing Cool Chain etc'
 'Mushroom Production' 'Artificial Insemination' 'Storage'
 'Breeding and culture of ornamental fish' 'Fishery Mechanization'
 'Fish  Marketing' 'Cold Storage' 'Freshwater pearl culture'
 'Plasticulture']
```

Figure 8. *Results of unique query type*

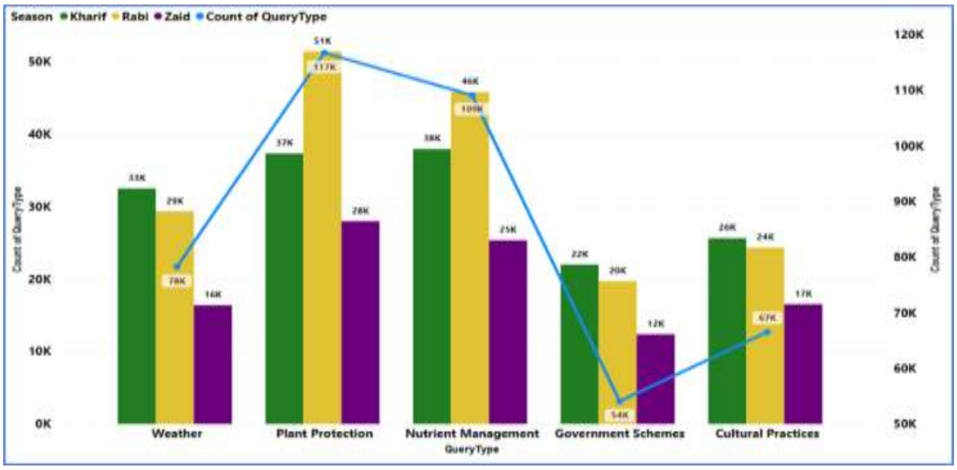

Figure 9. *Top 5 Query Types vs Season*

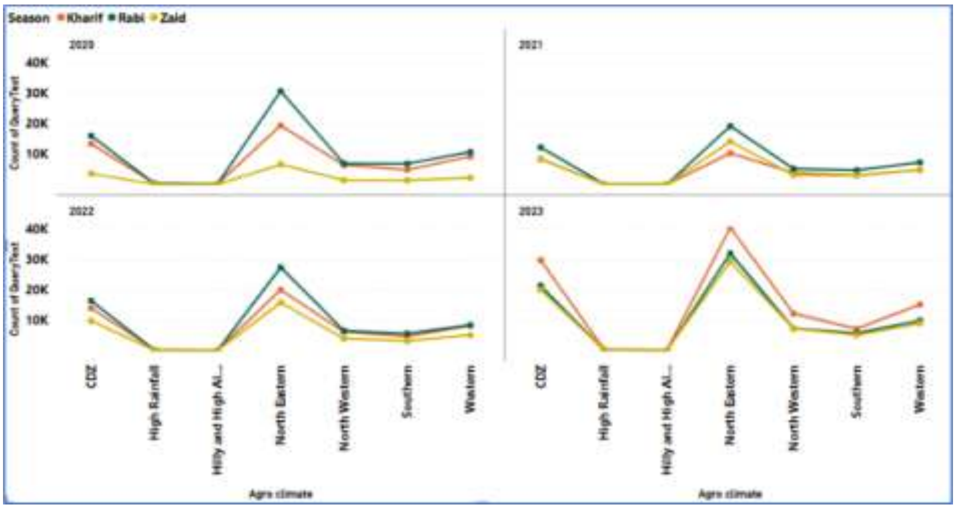

Figure 10*. Trends in call counts across years and seasons*

Call Volume vs Query Type

Calls have been categorized based on crop and query asked by the farmer. The farmer asked 26 different queries in the agriculture sector, 37 in Horticulture, 15 in Animal Husbandry, and 15 in Fisheries. Unique query types have been retrieved from the dataset using Python.

The call count has been calculated and the top 5 query types Plant Protection (116742), Nutrient Management (109134), Weather (78260), Cultural Practices (66541), and Government Scheme (54077) were selected for further analysis.

Call Volume vs Agroclimate

The call data analysis across different Agro-Climate Zones provides valuable insights into the regions that require the most attention and support. Here, we rank these zones based on the number of calls received, highlighting the areas with the highest engagement and potential need for agricultural support and resources. The volume of calls received from different zones can help us understand the specific needs and challenges faced by farmers in each region. Here's a detailed analysis of the number of calls received from various agro-climate zones and their respective ranks based on call volume. Table 3 is call vs agro climate.

Table 3. *Call vs Agro Climate*

Sno.	Agro-Climate Zone	No. of Calls	Rank
1.	North Eastern	265142	I
2.	Cauvery Delta Zone	173093	II
3.	Western	93903	III
4.	North Western	70332	IV
5.	Southern	55236	V
6.	High Rainfall	3239	VI
7.	Hilly and High Altitude	2365	VII

The North Eastern Zone tops the list with the highest number of calls (265142). The Cauvery Delta Zone, known for its extensive rice cultivation and other crops, ranks second in the number of calls (173093). The Western Zone, which includes diverse agricultural landscapes and cash crops, is third in call volume (93903). Ranking fourth (70332), the North Western Zone also shows considerable engagement. The Southern Zone is ranked fifth (55236). In comparison, the number of calls is lower compared to the top zones. With a lower call volume, the High Rainfall Zone ranks sixth(3239).

The Hilly and High Altitude Zone, which includes challenging terrains and unique agro-climatic conditions, has the lowest number of calls(2365)—drilled Northern Eastern Call logs through the season attribute. More number of calls were made during the Rabi Season till 2022 and calls during the Kharif (40430) season gradually increased in 2023. Table 4 is district belongs to agro climate zone [16].

Call Volume: Season vs Agroclimate

The Kharif season consistently receives the highest number of queries, particularly in 2020 and 2023, followed by Rabi and Zaid. The CDZ (Central Dry Zone) generally has the most queries across the years, with fluctuations in other zones like High Rainfall and North Eastern. Notably, 2021 saw a dip in queries compared to other years, especially in the High Rainfall and North Eastern zones. Each agro-climatic zone exhibits seasonal peaks, such as the Hilly and High Altitude zone during the Kharif season in 2020, highlighting the influence of specific climatic conditions. These patterns imply that regional temperatures, seasonal cycles, and outside variables have a major influence on inquiry volumes, providing information on the demands and difficulties of agriculture in various locales.

Call Volume vs Crop

The Crop column contains a total of 283 distinct crop entries. Paddy Dhan (136516), Groundnut pea nutmung phalli (32555), Black Gram urd bean (29172), Coconut (26883), Cotton Kapas (19972), Maize Makka (19625), Banana (17585), Chillies (15560), Brinjal (14482), Sesame GingellyTil Sesamum (14011), Onion (12963), Jasmine (11477), Sugarcane Noble Cane(10150) are crops received more than 10k Calls for the past four years (figure 11).

```
Crop
Others                            168615
Paddy Dhan                        136516
Groundnut pea nutmung phalli       32555
Black Gram urd bean                29172
Coconut                            26883
                                   ...
Brussils Sprouts                       1
Sponge Gourd                           1
Lehberry                               1
Spinach Palak                          1
Spine Gourd                            1
Name: count, Length: 283, dtype: int64
```

Figure 11. *Result of crop and call Volume*

In addition to the crop-related calls, inquiries related to weather, government schemes, credits, and similar topics are grouped under the crop category labeled as “Others”.This category holds the highest call records among crops(168615).

Crop vs Query Type

The top five crops and query types were selected for an in-depth analysis. Plant Protection and Nutrient Management queries on crops, Government schemes, and Weather Call volume have steadily increased over the years for all crop categories (figure 12).

Crop vs Season

Queries on Paddy dhan, Groundnut pea nutmung phalli & Black gram Urd bean crops were frequently asked by the farmers during the Rabi season. During the kharif season, farmers inquired about the weather and government schemes.

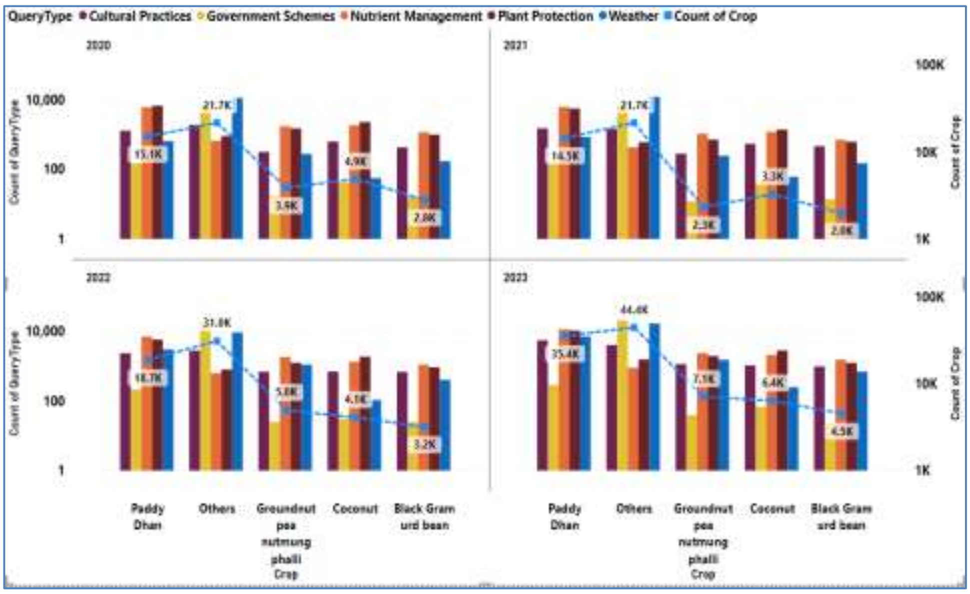

Figure 12. *Patterns in call volume: year vs query type*

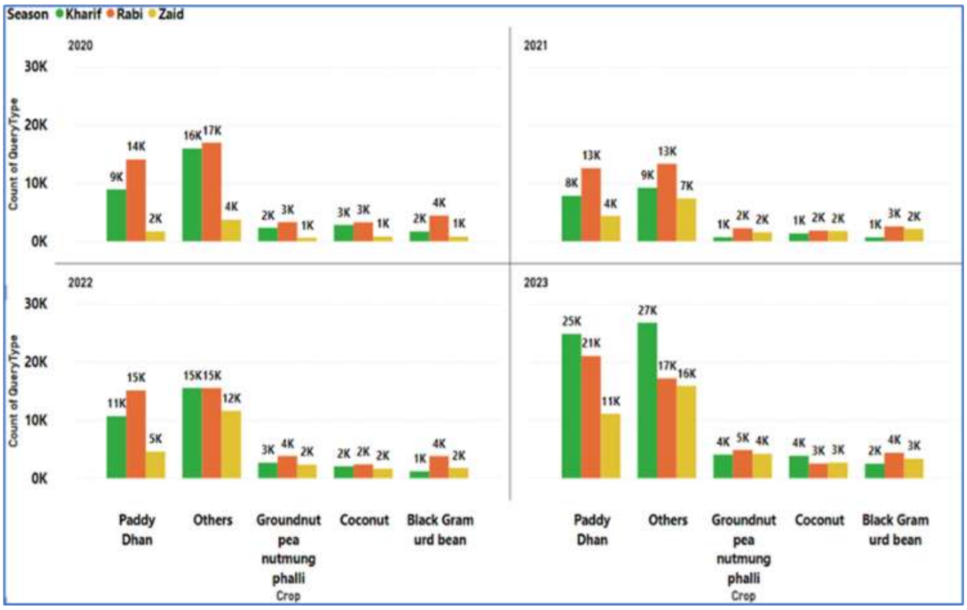

Figure 13. *Call Volume: Crop vs Season*

```
                                                QueryText
0                 asking about weed management in paddy
1           asking about sucking pests management for paddy
2         asked about paddy direct sowing post emergence...
3                             farmer asked query on weather
4                             farmer asked query on weather
...                                                     ...
663305                        farmer asked query on weather
663306               asked about weed management in paddy
663307    asked about pesticide and herbicide applicatio...
663308             asked about available of drones for rent
663309                        farmer asked query on weather
```

Figure 14. *Text processing before processing*

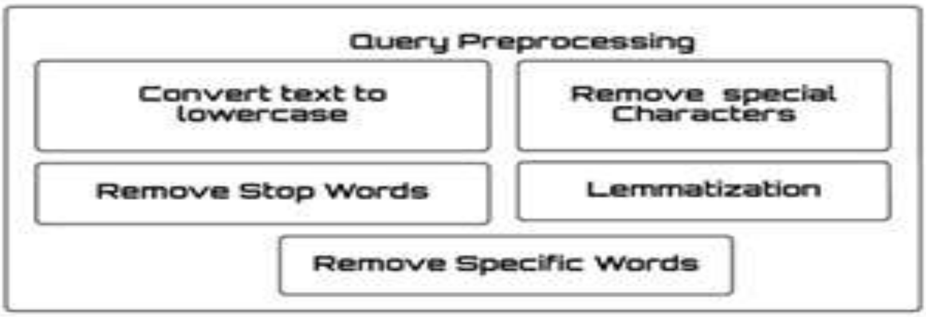

Figure 15. *Query Preprocessing Technique*

```
                                         preprocessed_query
0                                     weed management paddy
1                             sucking pest management paddy
2          paddy direct sowing post emergence weed manage...
3                                      farmer query weather
4                                      farmer query weather
...                                                     ...
663305                                 farmer query weather
663306                                weed management paddy
663307         pesticide herbicide application drone paddy
663308                                 available drone rent
663309                                 farmer query weather

[663310 rows x 2 columns]
Preprocessing complete. Preprocessed queries saved in 'preprocessed_queries.csv'.
```

Figure 16. *Text processing after processing*

Algorithm: Text Preprocessing for Query Text

Algorithm 1 Text Preprocessing Algorithm

1: **Input:** Raw text data (queries) in a CSV file
2: **Output:** Preprocessed queries saved in a new CSV file
3: **Step 1: Load the CSV file**
4: Load the dataset using pandas:

$$\text{Queries} = \text{df['QueryText'].astype(str).tolist()}$$

5: **Step 2: Initialize NLTK tools**
6: Initialize the lemmatizer:

$$\text{lemmatizer} = \text{WordNetLemmatizer()}$$

7: **Step 3: Define preprocessing function**
8: The preprocessing function performs the following steps:
9: **Step 3.1: Convert text to lowercase**

$$\text{Text}_{lower} = \text{Text.lower()}$$

10: **Step 3.2: Remove special characters**

$$\text{Text}_{clean} = \text{re.sub}(r'[^a-zA-Z\backslash s]', '', \text{Text}_{lower})$$

11: **Step 3.3: Tokenize text**

Words = word_tokenize(word_tokenize(word_tokenize(...)))

12: **Step 3.4: Remove stop words**

$$\text{Words}_{filtered} = \{\text{word} \mid \text{word} \in \text{Words}, \text{word} \notin \text{stopwords.words('english')}\}$$

13: **Step 3.5: Lemmatize words**

$$\text{Words}_{lemmatized} = \{\text{lemmatizer.lemmatize(word)} \mid \text{word} \in \text{Words}_{filtered}\}$$

14: **Step 3.6: Remove specific words ('asking', 'asked')**

$$\text{Words}_{final} = \{\text{word} \mid \text{word} \in \text{Words}_{lemmatized}, \text{word} \notin \{'asking','asked'\}\}$$

15: **Step 3.7: Return preprocessed text**

$$\text{Preprocessed Text} = \text{' '.join}(\text{Words}_{final})$$

16: **Step 4: Apply preprocessing to each query**

$$\text{Preprocessed Queries} = \{\text{preprocess(query)} \mid \text{query} \in \text{Queries}\}$$

17: **Step 5: Update the DataFrame**

df['preprocessed_query']df['preprocessed_query']df['preprocessed_query'] ... = Preprocessed Queries

18: **Step 6: Save the updated DataFrame**

df.to_csvdf.to_csv(...)

19: **Step 7: Print the preprocessed queries**

Print(df[['QueryText', 'preprocessed_query']])Print(df[['QueryText', 'preprocessed_query']])Print(df[['Query...

Figure 17*. Algorithm for Text Processing for Query Text*

Table 4. *District belongs to Agro climate zone [16].*

Zone	Districts
North Eastern	Villuppuram, Tiruvannamalai, Cuddalore(Part), Vellore, Kancheepuram, Thiruvallur, Kallakurichi, Chengalpattu, Ranipet, Tirupathur, Chennaimadras
Cauvery Delta Zone (CDZ)	Tiruchirappalli, Thanjavur, Ariyalur, Perambalur, Thiruvarur, Pudukkottai(Part), Mayiladuthurai, Nagapattinam, Cuddalore(Part), Karur(Part),
Western	Erode ,Tirupur,Coimbatore, Dindigul (Part), Pudukkottai(Part), Theni, Karur Pudukkottai(Part), Namakkal(Part).
North Western	Salem,Dharmapuri,Namakkal(Part),Krishnagiri
Southern	Tirunelveli ,Madurai,Virudhunagar,Thoothukudi,Ramanathapuram,Sivaganga
High Rainfall	Kanniyakumari
Hilly and High Altitude	The Nilgiris, Dindigul (Part)Kodaikanal .

Call Volume vs District

Districts were classified based on Agroclimate and listed based on the call volume in decreasing order (figure 13).

District vs Block

Overall call logs were analysed based on the top 5 query types and crops. Table 5 is list of districts and blocks for the top 5 query types based on the call volume.

Query Analysis

Real-world text data often contains inconsistencies such as mixed cases, special characters, irrelevant stopwords, and redundant or inflected words that can negatively impact the accuracy of analytical models. This code addresses these issues by performing preprocessing steps, including converting text to lowercase, removing special characters, tokenizing the text, filtering out stopwords, lemmatizing words to their base forms, and excluding specific unwanted words. Query analysis is conducted on the entire dataset using machine learning and Natural Language Processing (NLP) libraries. Several preprocessing steps were applied to ensure effective query analysis. These steps included tokenization (breaking down queries into individual words or tokens), lowercasing, punctuation and special character removal, stopword removal, and lemmatization (reducing words to their root form). After preprocessing, the textual data is transformed into a clean and standardized format, ready for analysis or use in machine learning tasks.This refined data is now more structured and consistent, making it suitable for applications like clustering, classification, sentiment analysis, or other natural language processing tasks. Preprocessing significantly enhances the quality of the input data, improving the performance and accuracy of subsequent analytical models.

Farmers frequently sought information on plant protection, nutrient management, and weather, emphasizing these as critical areas requiring continuous support. Queries related to crops such as paddy, groundnut, and black gram were particularly prevalent, indicating significant challenges associated with these crops that demand tailored solutions. Seasonal trends were also evident, with Rabi season queries focusing on specific crops, while Kharif season inquiries were more centered around weather-related issues and government schemes.Query analysis is conducted on the entire dataset using machine learning and Natural Language Processing (NLP) libraries. As a result of this processing, the top 10 most frequently asked queries by farmers between 2020 and 2023 were identified.

Figure 14 is text processing before processing and figure 15 is query preprocessing technique. Figure 16 shows text processing after processing and figure 17 shows algorithm for text processing for query text. Figure 18 is the counts of the most common queries asked by farmers.

```
Top 10 Queries:
preprocessed_query
farmer query weather                                   96462
top fertilizer management paddy                         4226
paddy top dressing fertilizer management                3498
leaf folder management paddy                            3190
pm kisan samman nidhi next due detail                   2601
azolla seed availability                                2470
pm kisan samman nidhi scheme status                     2374
pradhan mantri kisan samman nithi                       2362
weed management paddy                                   2276
pradhan mantri kisan samman nidhi pmkisan information   2078
Name: count, dtype: int64
```

***Figure 18**. The counts of the most common queries asked by farmers.*

4. Limitations

The limitations and practical implications were in different query types, such as button shedding in cconut query in nutrient and Plant Production—crab and Snail management in Plant Protection, and Varieties on Black Gram urd bean. Weather-related queries were recorded in the “Others” and specific Crop category.

4. Conclusions

The analysis covered query trends from Kisan Call Centers (KCCs) from 2020 to 2023, categorized by season, sector, crop type, and agro-climatic zones. The KCC dataset provides valuable insights, but it is essential to consider its limitations in terms of data quality, temporal and geographical scope, and sectoral representation. By addressing these limitations and leveraging the practical implications, policymakers and programme designers can enhance agricultural support services, leading to more effective and tailored interventions that meet the diverse needs of farmers. The findings from this research provide valuable data-driven insights for policymakers and program designers. The high volume of queries during the Kharif season highlights the need for increased resource allocation and focused support during this period.

Additionally, understanding the specific requirements of different agro-climatic zones can help develop region-specific policies and programs, ensuring a more effective and targeted approach to addressing farmers' concerns. In the future, to identify the specific problem of the enhanced query analysis by grouping similar queries to identify patterns across them, improving response strategies, and minimizing redundancy. Machine lerning is applied in some fields and reported [17-25].

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

Table 5. *Crop vs Query vs Block, District*

Crop and Query type based on top 5 Call Volume	District and Block
Others 1. Government Schemes 2. Weather 3. Seeds 4. Cultural Practices 5. Credit	1. Government Schemes- Kallakurichi, Villuppuram Dist., Northern Eastern Zone 2. Weather-Polur, Tiruvannamalai Dist., Nothern Eastern Zone 3. Seeds- Chennai(Azolla Seeds), Nothern Eastern Zone 4. Cultural Practices- Chengalpattu, Chengalpattu Dist., Nothern Eastern Zone 5. Credit- Chengalpattu, Chengalpattu Dist., Nothern Eastern Zone
Paddy 1. Nutrient Management 2. Plant Protection 3. Varieties 4. Weather 5. Cultural Practices	1. Nutrient management – Chengam, Tiruvannamalai Dist., Northern Eastern Zone 2. Plant production – Kattumannarkoil,Cuddalore Dist., CDZ Zone 3. Varieties – Polur, Tiruvannamalai Dist., Nothern Eastern Zone 4. Weather – Polur, Tiruvannamalai Dist., Nothern Eastern Zone 5. Cultural Practices – Polur, Tiruvannamalai Dist., Nothern Eastern Zone
Groundnut pea nutmung phalli 1. Nutrient Management 2. Plant Protection 3. Market Information 4. Weather 5. Sowing Time and Weather	1. Nutrient management-Jayamkondam,Ariyalur Dist., CDZ 2. Plant protection-T palur, Ariyalur Dist.,CDZ 3. Market information-Chennai, Nothern Eastern Zone 4. Weather-Jayamkondam, Ariyalur Dist.,CDZ 5. Sowing time and weather -Keelpennathur, Tiruvannamalai Dist., Nothern Eastern Zone
Black Gram urd bean 1. Sowing Time and Weather 2. Nutrient Management 3. Plant Protection 4. Varieties 5. Market Information	1. Sowing Time and Weather-Kallakurichi, Villuppuram Dist., Northern Eastern Zone 2. Nutrient Management-Vikravandi, Villuppuram Dist., Northern Eastern Zone 3. Plant Protection- Kallakurichi, Villuppuram Dist., Northern Eastern Zone 4. Varieties-Vikravandi, Villuppuram Dist., Nothern Eastern Zone 5. Market Information-Kallakurichi, Villuppuram Dist., Northern Eastern Zone
Coconut 1. Plant Protection 2. Nutrient Management 3. Cultural Practices 4. Market Information 5. Seeds and Planting Material	1. Plant Protection-Udumalaipettai, Coimbatore dist., Western zone 2. Nutrient Management- Udumalaipettai, Coimbatore dist., Western zone 3. Cultural Practices- Udumalaipettai, Coimbatore dist., Western zone 4. Market Information-Dharapuram, Coimbatore dist., Western zone 5. Seeds and Planting Material-Pollachi North, Coimbatore dist., Western zone

- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Anima, P & Aneeshkumar, A. (2023). Implementation of Sequential Pattern Neural Classifier in E-Commerce Data Behavioral Characteristic Extraction. *Indian Journal Of Science And Technology*. 16. 1438-144, https://doi.org/ 10.17485/IJST/v16i19.94

[2] Samarth Godara, Durga Toshniwal, (2020) Sequential pattern mining combined multi-criteria decision-making for farmers' queries characterization, *Computers and Electronics in Agriculture*, 173; 105448 https://doi.org/10.1016/j.compag.2020.105448

[3] Kavitha, S., & Anandaraja, N. (2018). Kisan Call Centre Services to the Farming Community: An Analysis. *Journal of Extension Education*, 29(3). https://doi.org/10.26725/JEE.2017.3.29.5910-5916

[4] Koshy, S. M., & Kishore Kumar, N. (2017). Attitude of Farmers towards Kisan Call Centres. Journal of Extension Education, 28(4). https://doi.org/10.26725/JEE.2016.4.28.5753-5759

[5] D. P. Ramasamy, K. Geetha, V. Radhakrishnan, and Iniya V., (2024). Comprehensive Analysis of Kisan Call Centre in Tiruchirappalli District of Tamil Nadu, India, *AJAEES*, 42(1);62–67.

[6] Yadav , V., R. K. Doharey, J. Vishwakarma, N. R. Meena, M. Kumar, and Y. Kumar. (2024). Cultivating Solutions: Farmer-Driven Constraints and Suggestions for Kisan Call Centre Advisories. *Journal of Experimental Agriculture International*, 46(2);141-6, https://doi:10.9734/jeai/2024/v46i22316

[7] Samarth Godara, Durga Toshniwal, Rajender Parsad, Ram Swaroop Bana, Deepak Singh, Jatin Bedi, Abimanyu Jhajhria, Jai Prakash Singh Dabas,

Sudeep Marwaha, (2022). AgriMine:A Deep Learning integrated Spatio-temporal analytics framework for diagnosing nationwide agricultural issues using farmers' helpline data, *Computers and Electronics in Agriculture*, 201;107308, https://doi.org/10.1016/j.compag.2022.107308

[8] Samarth Godara, Durga Toshniwal, (2022) Deep Learning-based query-count forecasting using farmers' helpline data, *Computers and Electronics in Agriculture*,196;106875, https://doi.org/10.1016/j.compag.2022.106875

[9] V. K. Viswanath, C. G. V. Madhuri, C. Raviteja, S. Saravanan and M. Venugopalan, (2018). Hadoop and Natural Language Processing Based Analysis on Kisan Call Center (KCC) Data," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India*, 2018, pp. 1142-1151, https://doi: 10.1109/ICACCI.2018.8554531

[10] Mohammad Zia Ur Rehman, Devraj Raghuvanshi, Nagendra Kumar, (2023) Kisan QRS: A deep learning-based automated query-response system for agricultural decision-making, *Computers and Electronics in Agriculture*,213;108180, https://doi.org/10.1016/j.compag.2023.108180

[11] S. Godara, J. Bedi, R. Parsad, D. Singh, R. S. Bana and S. Marwaha, (2024). AgriResponse: A Real-Time Agricultural Query-Response Generation System for Assisting Nationwide Farmers, *IEEE Access*, 12;294-311, https://doi: 10.1109/ACCESS.2023.3339253.

[12] Koshy, S. M., & Kishore Kumar, N. (2017). Attitude of Farmers towards Kisan Call Centres. *Journal of Extension Education*, 28(4). https://doi.org/10.26725/JEE.2016.4.28.5753-5759

[13] Kavitha, S., & Anandaraja, N. (2018). Kisan Call Centre Services to the Farming Community: An Analysis. *Journal of Extension Education*, 29(3). https://doi.org/10.26725/JEE.2017.3.29.5910-5916

[14] Goyal, Shashikant & Jirli, Basavaprabhu. (2021). Perceived Problems and Suggestions of Farmers regarding *Kisan Call Centre*.

[15] Puja Sinha, Meera Kumari, Sandeep Kumar and Ramnath Kumar (2019). Ray Constraints in Pulse Cultivation Perceived by the Farmers of Tal Land in Patna District of Bihar, *India Int. J. Curr.Microbiol.App.Sci*. 8(8);2991-2997 DOI: https://doi.org/10.20546/ijcmas.2019.808.346

[16] Olujimi, P.A., Ade-Ibijola, A. (2023) NLP techniques for automating responses to customer queries: a systematic review. *Discov Artif Intell* 3, 20. https://doi.org/10.1007/s44163-023-00065-5

[17] D. Naga Jyothi, & Uma N. Dulhare. (2025). Understanding and Analysing Causal Relations through Modelling using Causal Machine Learning. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.1018

[18] N.B. Mahesh Kumar, T. Chithrakumar, T. Thangarasan, J. Dhanasekar, & P. Logamurthy. (2025). AI-Powered Early Detection and Prevention System for Student Dropout Risk. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.839

[19] P. Rathika, S. Yamunadevi, P. Ponni, V. Parthipan, & P. Anju. (2024). Developing an AI-Powered Interactive Virtual Tutor for Enhanced Learning Experiences. *International Journal of Computational and Experimental Science and Engineering*, 10(4). https://doi.org/10.22399/ijcesen.782

[20] Krishna Kumaar Ragothaman. (2025). Smart Distribution in E-Commerce: Harnessing Machine Learning and Deep Learning Approaches for Improved Logistics. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.1157

[21] K. Tamilselvan, , M. N. S., A. Saranya, D. Abdul Jaleel, Er. Tatiraju V. Rajani Kanth, & S.D. Govardhan. (2025). Optimizing data processing in big data systems using hybrid machine learning techniques. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.936

[22] Naresh Babu KOSURI, & Suneetha MANNE. (2024). Revolutionizing Facial Recognition: A Dolphin Glowworm Hybrid Approach for Masked and Unmasked Scenarios. *International Journal of Computational and Experimental Science and Engineering*, 10(4). https://doi.org/10.22399/ijcesen.560

[23] J. Prakash, R. Swathiramya, G. Balambigai, R. Menaha, & J.S. Abhirami. (2024). AI-Driven Real-Time Feedback System for Enhanced Student Support: Leveraging Sentiment Analysis and Machine Learning Algorithms. *International Journal of Computational and Experimental Science and Engineering*, 10(4). https://doi.org/10.22399/ijcesen.780

[24] Ibeh, C. V., & Adegbola, A. (2025). AI and Machine Learning for Sustainable Energy: Predictive Modelling, Optimization and Socioeconomic Impact In The USA. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.19

[25] Olola, T. M., & Olatunde, T. I. (2025). Artificial Intelligence in Financial and Supply Chain Optimization: Predictive Analytics for Business Growth and Market Stability in The USA. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.18