ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Novel TDMKCC Model for decision making in e-Agriculture

Krishna Kumar Verma
Department of Computer Science
A.P.S. University
Rewa (M.P.) India

Navita Shrivastava
Department of Computer Science
A.P.S. University
Rewa (M.P.) India

Abstract: E-Agriculture is one of the wing of e-Governance and application of information communication technology (ICT) for improving the agriculture sector. The implementation, application and adoption of ICT along with utilization of data mining techniques over data generated by these services are necessary for sustainable development and continue growth in agricultural sector. Telecomm sector is playing an important role in e-Agriculture, which is very useful in problem/query identification and solving; this process is initiated as "Kisan Call Center (KCC)" in India. In this research paper a novel model *"Text Data Mining for Kisan Call Centre (TDMKCC)"* has been proposed for knowledge management. We have implemented this proposed model with the help of R-Tool, suggestions and recommendations are also discussed which will definitely help in improving decision making for sustainable growth in Farming or Cultivation practices.

Keywords: Text Data Mining, e-governance, e-Agriculture, Kisan Call Centre, Open data

I. INTRODUCTION

The Department of Agriculture & Cooperation (DAC), Ministry of Agriculture, Govt. of India, launched Kisan Call center (KCC) in 2004 across the country for increasing knowledge accessibility in the form of answer asked by farmer through phone call. The main purpose of this initiative is to respond to issues or problem raised by farming community, in the local language. An operational model of KCC is depicted in Figure 1, which has been adopted for managing farmer calls (questioning and answering) in the form of KCC in India. This operational model has three levels: Level-I is professionally managed basic call center interface with high quality bandwidth and management of local language, Level-II is a response center, where subject matter specialists of agriculture are made available and Level-III is the Nodal Cell Management Group, which is supposed to be managed by a senior officer from the nodal institution [1].

Kisan Call center (KCC) is an important service of e-Agriculture under e-Governance programme. This service can be divided into common applications services as Government to Citizen (G2C), Government to Business (G2B) and Government to Employee (G2E) [2], for instance:

- G2C: Especially this service is launched by government of India (GoI) to farming citizen.
- G2B: Indirectly this service includes application between agriculture departments and local market.
- G2E: This service also includes employee management to coordinate departmental activity.

Text Data Mining (TDM) is a process of extracting text documents from stored database and converting this unstructured or semi structured data set into a specific structured data set, this process includes text operation, tokenization, stop word removing, white space removing, unused word, symbol removing etc followed by data mining techniques. As we know that text mining technique are implemented in various field for knowledge production some applications are opinion-comment mining [3], opinion mining [4], spam collection, blog classification, document classification and clustering etc.

Analysis of text mining technique, issues in this field and their application in multidisciplinary field are discussed by [5]. According to them text mining techniques can be applied in Digital Libraries, Academic and Research Field, Life Science, Social Media, Business Intelligence etc. for efficient knowledge generation.

The remainder of this research paper is organized as follows. We start our discussion with a detailed overview of existing studies available in web in section I & II. In Section III & IV, we discuss our algorithm and give an overview of our proposed novel model with the explanation of text data mining (TDM) in Kisan Call Center open data for knowledge extraction & effective decision making. Section V & VI presents our experimental results, discuss and implications of our findings. Finally we conclude our work in last section.

II. RELATED WORK

An attempt was done by R.S. Chouhan et al. (2011) to analyze the per month call record received by Kisan Call Centre (KCC) of Bhopal (M.P.). In this paper authors have discussed that, the calls received from the agriculture sector includes the maximum number of query, which was related to plant protection followed by production techniques, marketing of farm production, weather forecasting and others.

In this study it was observed that there were 13539 calls received per month by the KCC out of which the maximum calls was belong to agriculture field followed by horticulture & others. Among the different districts of Madhya Pradesh (M.P.) the maximum calls were received from Shivpuri followed by Shajapur, Rajgarh, Sagar, Datia, Chatterpur, Tikamgarh, Ujjain, Betul, Chhindwara, Vidisha, Dewas, Mandsaur and Khargon. It was observed that farmers had not been able to follow the recommendations due to communication gap, lack of education, lack of cooperation from agricultural department, difficulty in the adoption of solution due to non-availability of the suggested inputs in local market, Hence it was proposed that the recommendation must

be tuned with local language, local cultivation system, availability of inputs in the local bazaar with lower costs [6].

A study on primary data of kisan call centers (KCC) had been done by B.R. Sharma et al. (2011) in Himachal Pradesh agriculture. In this research, researchers have selected two different crops I- Apple (a fruit crop) and II- Tomato (vegetable crop) for find out its growth in high hills and mid hills around the state land. This study was based on the primary data collected from the selected 200 farm households – 100 for apple crop and 100 for tomato crop of Shimla and Solan districts of Himachal Pradesh state.

It was observed that out of these 100 households in each cluster village 50 households was progressive and used the services of the KCC and the remaining 50 households did not use this facility. For the tomato crop same patterns was also found. In this research authors have done study on calls recorded from August 2008 to June 2010. The highest calls were recorded for the different diseases in agriculture, animal husbandry, etc, it was revealed that farmers are utilizing this service for extracting more and more knowledge and by using this service they are able to cultivate their crops more efficiently and more scientifically. At the end of this study, it was suggested that there is need to educate farmers more and more by the media and departmental activities [7].

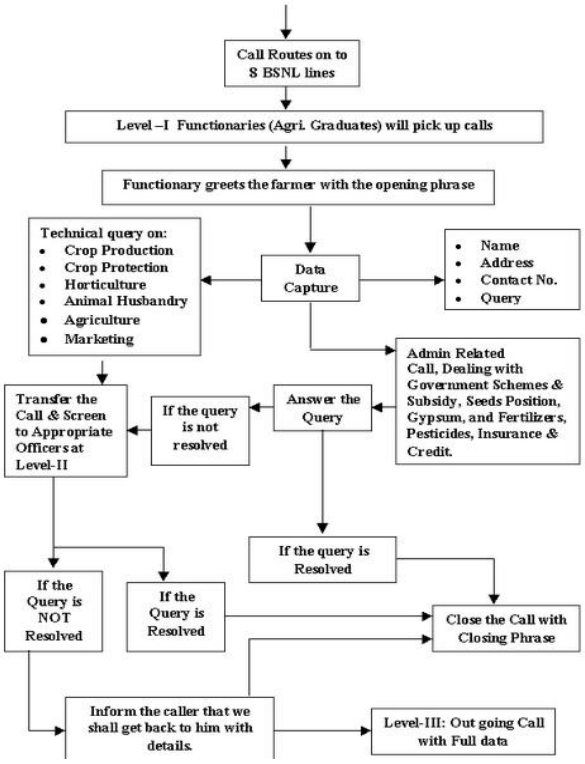

Figure 1: Operational Model [1]

Deepak Shah (2011) discussed issues related to the applications of ICT for economic change in the agricultural sector of India and identified the past and present major ICT initiatives in agriculture. The factors those are responsible for the success of ICT services and the elements of an appropriate framework on agricultural development are also discussed. It has been proved that higher growth in agricultural sector can be achieved only when there is suitable application of ICT with adequate investments from both public and private sectors. Some examples of supply chain management as: the case of e-choupal in Madhya Pradesh and ikisan, Kisan Call Centers (KCCs), Village Knowledge Centres (VKCs), The Gyandoot Project, AGMARKNET, etc. are also covered in this study [8].

Himadri Subrah Saha (2012) has discussed that "Call centers are an integral part of many businesses, and their economic role is significant and growing" and the concept of this call center is playing major role in our agriculture sector thus it can enhance the growth of our economy. Authors also have focused on the steps, which are involved in assessing the technological design; staffing requirements and technical answering is one of the parts of this involvement [9].

Hemavathy Ramasubbian (2015) has discussed the various ICT tools and methodology available for agriculture sector (like wireless technology, global positioning system (GPS), geographic information system (GIS), radio frequency identification (RFID) etc.).

The main focus of this study was Kisan Call Centre's initiative. Operational mechanism for KCC along with schematic diagram and detail information related with this service was also discussed along with covering various levels and it complete process [10].

III. ALGORITHM

Input: A set of databases (D) has the form Di $\{a_1, a_2, \ldots, a_n\}$, where i={1 to n} is representing different integer value for different databases and a1 to an are different attributes.
Output: A set of all frequent terms for textual data.
Methods:
1. Start
2. Make the union of different document and store it in integrated file.

$$IDS < - \bigcup_{i=1}^{n} Di$$

Where IDS stands for integrated data set and i =1 to n is representing different data sets.
3. Project the selected attribute and store it in new file.

$$Text\ File < - \Pi_{Attribute\ \ Name}\ (IDS)$$

4. Create a corpus file from projected new *"Text File"*.
5. Remove the stop words, symbol, white spaces etc.
6. Find the term frequency matrix.
7. Plot the bar chart and create the rule.
8. If document classify correctly and rules are satisfactory.
   then Go to step 10 Else Go to Next.
9. Remove unused words and merge the related words in corpus and Goto Step 5.
10. Stop

Note: Sequence of relational algebra operations - firstly using Union operation and the Project operation. This is not optimized process but for future work; this process generates an integrated data set having all attributes with all rows, can be utilized for generating different type of knowledge.

IV. PROPOSED MODEL

In this paper we have proposed novel model named "Text Data Mining for Kisan Call Center" (TDMKCC); which can be utilized for improving the performance & decision making in Kisan Call Center sector as well as in e-agriculture sector, by using text data mining & visualization on historical data. The overall text data mining process, we have suggested, is

schematically illustrated in Figure 2. This model has a set of task and is divided into following five steps:

A. Data Collection
B. Integration of Data
C. Attribute Selection
D. Text Data Mining Process
E. Decision Making

A. Data Collection

It is a process of collection of secondary data from Indian open government data (https://data.gov.in) and converting these data according to our required format using online/ offline tools as text, csv, xls etc. file format. In this research, we have collected KCC data of different districts of Madhya Pradesh state, belongs to month of December 2015 and December 2016.

B. Integration of Data

The integration is an important process of any research, in this model we will integrate different data sets based on our required knowledge, because KCC open data is divided in to three categories. Firstly these data are categorized in to month wise (from Jan to Dec) than state wise (from J&K to Kanyakumari) and at last district wise of every state; so after integration we can generate & predict different types of knowledge from different integrated data set.

For integration we have taken (Year 2015->Month: December->State: Madhya Pradesh-> All Districts) as training data set and (Year 2016->Month: December->State: Madhya Pradesh-> All Districts as testing data set.

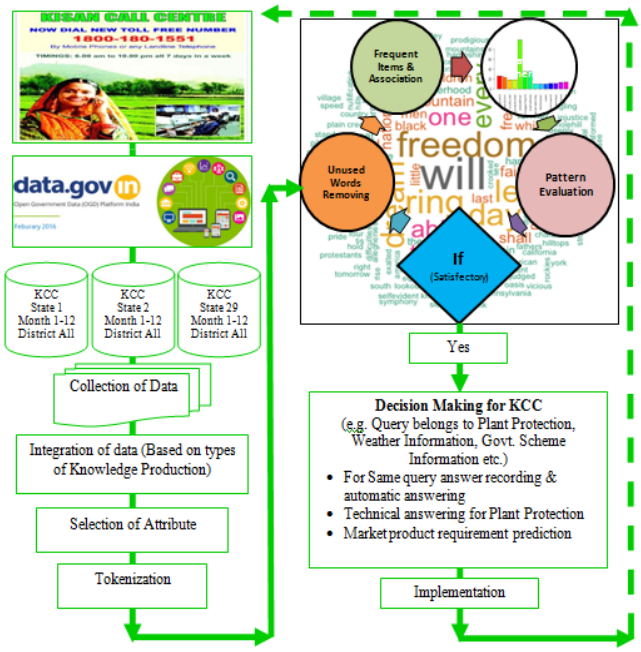

Figure 2: Text and Data Mining Model for Kisan Call Center

C. Selection of Attribute

We have seen that, KCC data warehouse has a set of attributes, mostly these attribute has the values in the form of text so selection of attribute is also another important part of our model. For identification of most frequent query, we can follow this step- firstly select attribute "query type" and try to find out which query is frequently asked by farmer. On the basis of this query we can do further analysis. For example if the frequent query type is about plant protection then we need to identify which plant is mostly affected from deceases in specific month.

D. Text Data Mining Process

After selection of attribute, we apply text data mining process for producing knowledge and pattern discovery. Text data mining process include first text mining and then data mining techniques for knowledge extraction which makes text analysis more and more simple and efficient. This Process includes

- Tokenization
- Creation of Corpus
- Stop word removal
- Unused word removal
- Frequent term identification
- Pattern plotting and
- Result evaluation

E. Decision Making

On the basis of result produce by this model, we would able to improve the kisan call center process. Recording & automatic answering mechanism for most frequent questions from citizen, may play important role for improving process of this sector. Another important factor of this model is that the result produced by this model can be utilized for market management.

V. IMPLIMENTATION AND VALIDATION

For implementation and validation testing of proposed model, we have used open source R tool along with its text mining and other package. Frequent data pattern for attribute "query type" is shown in following graph for integrated data of year 2015 and 2016 (Figure 3 and Figure 4).

It has been observed that "Plant Protection", "Weather Information", "Government Scheme Information", "Market Information" & "Fertility Uses Availability" are most frequent question asked by farmer in year 2015. Also in year 2016 previously year asked questions are repeated with same frequency, new questions are also arises, which has some low frequency.

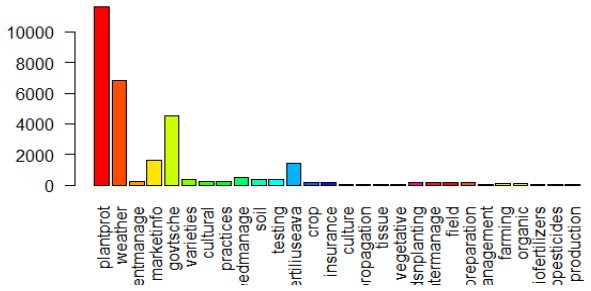

Figure 3: Query Type frequency of M.P. December 2015

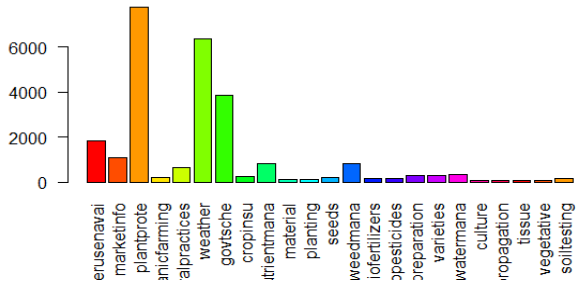

Figure 4: Query Type frequency of M.P. December 2016

The word cloud over "plant protection" data is created (shown in Figure 5 and Figure 6) which shows that fungal attack is major problem of various crops in M.P. in month December.

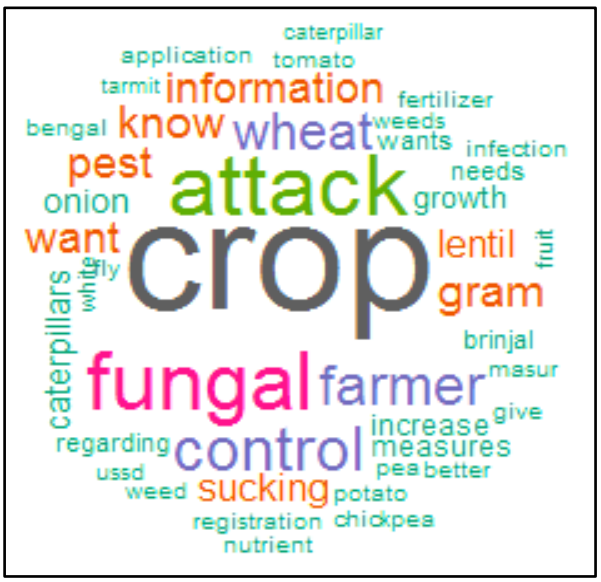

Figure 5: Word Cloud of "Plant Protection"
(IDS December 2015)

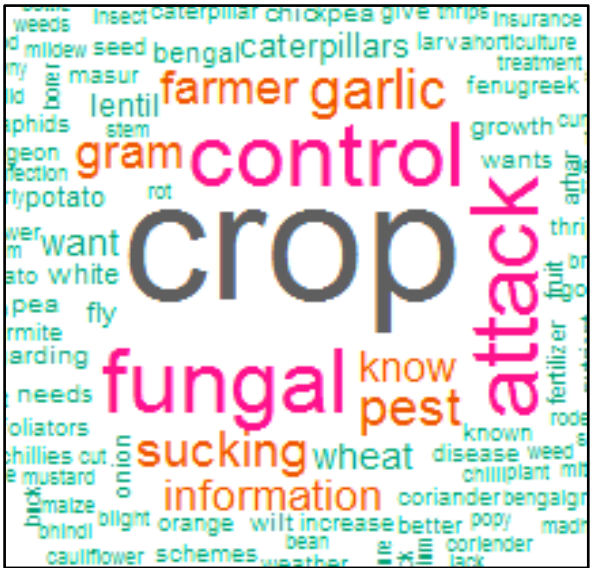

Figure 6: Word Cloud of "Plant Protection"
(IDS December 2016)

VI. Result and Discussion

On the analysis of previous year data, we can predict the type of problems, which will definitely come in next year. With the help of proposed model, we will able to find answer the questions (problems) of farmer in advance.

In this process we have collected data from open government data portal (https://data.gov.in) of India & Integrate these data for generation of effective knowledge pattern.

The data has 11 attributes (Season, Sector, Category, Crop, QueryType, QueryText, KCCAns, StateName, DistrictName, BlockName and SubmitDate), we have downloaded data set that is of December 2015 & 2016 respectively, it also belongs to different 50 districts of Madhya Pradesh. These downloaded data sets are converted into CSV (Comma separated value) format and then integrated. After Integration data size becomes 29064 records for 2015 & 25449 records for 2016 respectively.

After this process we applied text and data mining technique over integrated data set. The outcome shows that most frequest query topic asked by formers is "Plant Protection" .

VII. Acknowledgment

I would like to thanks to Ministry of Agriculture and Farmers Welfare Department of Agriculture, Cooperation and Farmers Welfare, Government of India, for providing Kisan Call Center Data as open data in Open Government Data (OGD) Platform India.

VIII. Conclusion

In this research paper, we have introduced the concept of Kisan Call Center process followed by Text Data Mining techniques and proposes a well-defined novel model named "Text Data Mining for Kisan Call Center" (TDMKCC), this model has following features:

1. Extracting the secondary data from e-repository of KCC.
2. Integration of these e-repositories.
3. Text Data Mining process over selected attribute.
4. Knowledge production and pattern evaluation.

Based on experiment outcome Plant Protection, Weather Information, Government Scheme Information, Market Information & Fertility Uses Availability etc. are most frequent questions asked by farmer in the year 2015. Same as in year 2016, previously asked questions are repeated with average/same frequency; new queries are also raised with low frequency. After this, same experiment are performed over sub dataset "Plant Protection" and we observed that query over "fungal attack on crop" is asked by maximum number of farmers in both year.

Via our proposed approach we can say that knowledge produced by this model can be adapted for improving decision making in KCC and its related departments for sustainable development in country.

IX. References

[1] JNKVV, http://jnkvv.org/JNKVV_EXTENSION_SP/EXTENSION_KisanCall Centre.aspx

[2] V.R. Rao," A Framework for e-Government Data Mining Applications (eGDMA) for Effective Citizen Services"-An Indian Perspective. *International Journal of Computer Science and Information Technology Research*, *2*(4), pp.209-225, 2014.

[3] T. Tilahun, & D.P. Sharma, "Design and Development of E-Governance Model for Service Quality Enhancement", *Journal of Data Analysis and Information Processing*, *3*(03), 55, 2015.

[4] G. Stylios, Christodoulakis, D., Besharat, J., Vonitsanou, M., Kotrotsos, I., Koumpouri, A., & Stamou, S., "Public opinion mining for governmental decisions", *Electronic Journal of e-Government*, *8*(2), 203-214, 2010.

[5] R. Talib, Hanif, M. K., Ayesha, S., & Fatima, F., "Text Mining: Techniques, Applications and Issues", *International Journal of Advanced Computer Science & Applications*, *1*(7), 414-418, 2016.

[6] R.S. Chouhan, Dushyant Kumar and H.O. Sharma, "Performance of Kisan Call Centre: A Case Study of Kisan Call Centre of Indian Society of Agribusiness Professionals Bhopal (Madhya Pradesh)", Ind. Jn. of Agri. Econ.Vol.66, No.3, 2011.

[7] B.R. Sharma, Pratap Singh and Amresh Sharma, "Role of Kisan Call Centres in Hill Agriculture", Ind. Jn. of Agri. Econ.Vol.66, No.3, 2011.

[8] D. Shah, "Information and Communication Technology for Agricultural Development in India: An Integrated Approach", Ind. Jn. of Agri. Econ.Vol.66, No.3, 2011.

[9] H. S. Saha, R. Roy and M. M. Hussain, "Capacity Planning and Dimensioning of an Agricultural Call Center", International Journal of Computer and Information Technology, ISSN 2218-5224 (ONLINE), Vol 02, No. 02, 2012.

[10] H. Ramasubbian et al., "Information & Communication Technology (ICT) in Indian Agricultural Sector with special reference to Kisan Call Centre (KCC)", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue 6, ISSN(Online): 2319-8753, ISSN (Print): 2347-6710, 2015.