Collaborative Recommendation System For Agriculture Sector.

Sapna Jaiswal[1,3,*], *Tejaswi* Kharade[2,**], *Nikita* Kotambe[3,***], and *Dr.Shilpa* Shinde[3,****]

[1]jsapna2011@gmail.com
[2]tejaswisampat1898@gmail.com
[3]nikitakotambe@gmail.com
[4]shilpa.shinde@rait.ac.in

Abstract. Agriculture is one of the most important sector in India and the farmers are one of the most essential members of society. The major economy of the country comes from the agricultural sector. Though there is no end to the woes of Indian Farmers. One of the major causes for the continuing Indian farmer distress is lack of knowledge and benefits of the agricultural programs and schemes proposed by the Government of India.The Collaborative Recommendation System For Agriculture Sector is one such way to solve this problem. There are various workshops conducted to create awareness about the government schemes to the farmers but still the results are not seen as expected. Even if they are aware they are not solved and hence many NGOs and and Institutes have come up with various measures to solve this problem. Our research system focuses on helping the farmers by answering their agricultural queries by generating a profile of basic requirements through a web application and recommends the proposed government schemes developed to help farmers.The recommendation system also periodically update farmers with the recent trends in agricultural field, new Government schemes and programs. Keywords - agriculture, government schemes, web application, recommendation, knn algorithm, cosine similarity, CRSAS- Collaborative Recommendation System For Agriculture Sector

1 Introduction

Growth in agricultural domain are one of the main concerns to the farmers as well as to the economy of our country as it plays the major role in contribution for the same. Agriculture serves a s much as quarter of the Indian economy, employing more than 60 percent of the labour force. Despite of progress in agricultural domain, the problems faced have continued frustrating Indian farmers and the citizens. It is estimated that one-fourth of the Indian produce is lost due to inefficient methodologies used for harvesting, farming, transportation and storage of goods produced of government subsidized crops. Problems due to which farmers are lacking behind consists of Inequality in land distribution, land tenure system, sub-division and fragmentation of holdings, Cropping pattern, instability and fluctuations, and condition of labour. Poor farming technique also gives a way to different problems that ultimately affects the overall economy of the country.

The agricultural sector of India still faces a very vital issue of agricultural growth deceleration. Declining of investment in agricultural research and development stands as one of the core reasons for reduction of growth of the agriculture in India.Thus there needs some technologies and methodologies to work together and bring up a system for farmers as per their problems.

Various schemes do exist based on a single domain such as the type of soil needed for cultivation, different types of crops, other utilities required for agriculture, but there are very few such measures which guides the farmers with proper steps to be undertaken for greater yield, much more profit and also with less expenses and with a helping hand of government. We have come up with combined solution which use many technologies for getting insight of details in more depth and providing the maximum profitable solution. Surveys conducted for the betterment and educational purpose of the people will have the major usage of recommendation system.

To overcome all these problems faced by farmers and ultimately the country, Government of India has proposed and implemented a large number of welfare schemes for farmers. These schemes are to be used by farmers for their growth in agricultural domain. The schemes provide help such as financial help, marketing help, irrigation facilities, insurance to farmers crops, subsidies to farmers and their family. But these schemes need to reach every remote sector of India so that they can get benefits, since due to lack of exposure towards farmers and unawareness of different government schemes the farmers still face various problems though there are various solutions proposed by the government of India. There are various applications related to the recommendation of crops and other problem related to agriculture.[1]

There are ample amount of systems available for farmers that recommend them crops, soil quality,fertilizers to be

---

*e-mail: jsapna2011@gmail.com
**e-mail: tejaswisampat1898@gmail.com
***e-mail: nikitakotambe@gmail.com
****e-mail: shilpa.shinde@rait.ac.in

used, pesticides, urea for soil and also health checkup of soil. There is no such system available for farmers to help with the information of schemes and also helping them choose appropriate scheme. By working on this project we can analyse different queries imposed by farmers and recommend them different government schemes made available by Indian Government to help these farmers.

The system does query analysis from data provided by the farmers in their profile and recommends the scheme based on the feasible solution to their queries. The system also visualises the schemes proposed, implemented, number of farmer queries resolved by the available schemes.

The farmers are unaware of the different government policies/schemes that are helpful for them in various problems related to agriculture. The main objective of this project is to show the amount of queries farmers have in a particular sector such as financial help, crop detection and suggest them different government schemes so that farmers are aided with help and know the process to do so.

The structure of this paper is as follows: The section 2 introduces literature survey of various existing papers, section 3 introduces the proposed system, section four explains the experimental results and finally section five concludes the agenda with future work suggestions.

2 Literature Survey

This section comprises of the findings related to our topic. It is basically an evaluative report of information found in the literature relevant to Recommendation Systems based on Soil, Crop, Fertilizers, Insurance, loans. The systems existing are Crop recommendation system, Soil based fertilizer recommendation system, Query Analysis and crop recommendation system, Recommendations with hadoop framework. These systems aim at singular elements such as recommendations only for crop, or fertilizers, or may be some subsidies and insurances for farmers.

1. In, Crop Selection Method to Maximize Crop Yield Rate using Machine Learning [2], the research centers its focus to build a recommendation system that can collect raw data for environmental factors like soil, weather parameters from experienced farmers, agricultural researchers and other stake holders. Based on these parameters, structures are studied and crop is recommended to the farmers. Statistical data analysis and predictive modeling are incorporated in order to anticipate a suitable crop accordingly.Technique named Crop Selection Method (CSM) is used to solve crop selection problem and maximize net yield rate of crop over season and achieves maximum economic growth.

2. In, Web based recommendation system [3], the process focuses on the use of data mining techniques to recommend the farmers regarding crops, crop rotation and identification of appropriate fertilizer. The system uses random forest algorithm for the crop recommendation, and have considered Year of Cultivation of Resultant Crop ,Market Price of Resultant Crop ,Output/Area Ratio of Resultant Crop as

the 3 factors to distributed the rating. For crop rotation recommendation, authors have used FP tree matching frequent pattern observed during past cultivation. Also for fertilizer recommendation, Sufficiency approach is used to check the nutritional values and approaches possible combination of fertilizers to meet the crop requirements and the combination with lowest cost of fertilization will be recommended.

3. In, KrishiMantra: Agricultural Recommendation System [4], the system is used for domain knowledge and used to send recommendations to the farmers based on climate conditions and geographic data. The system shows experimental outcomes as a section of implementation of proposed architecture. The farmer types queries to the query engine, in order to get information for a specific crop. Query asked are matched to GIS data, crop knowledge domain or both, the query is transformed into a semantic web query and after performing reasoning and semantic processing, the result is sent to the mobile device of user.

4. In, Recommendation System With SVM [5], the system proposes a recommendation system for the large amount data in the form of ratings, reviews, opinions, complain, remarks, feedback, and comment about the items such as (product, event, individual and services) using SVM and collaborative filtering. GA is used as optimization technique. Here the recommendation works using hybrid filtering technique to refine different types of factors considering review generated by different reviewers.[6]

3 Proposed System

The underlying section gives a detailed representation of the design which includes the diagrammatical representation along with its explanation.
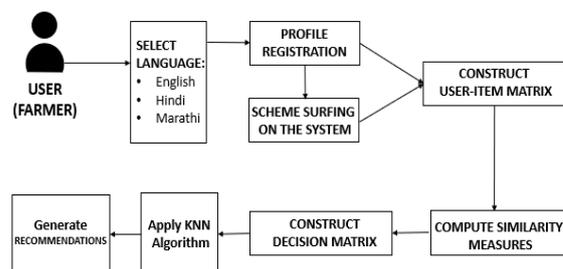

Figure 1. Block Diagram of CRSAS

The target of query analysis is to recognize the problems faced by farmers throughout the year. There is a login and registration page for farmers. Also at times farmers having probability to understand the regional language so we have included Hindi and Marathi languages with the base of English language in the user interface. This will

make the interface user friendly. As there are eligibility criteria for many of the government schemes we have a profile to be filled by the farmers with basic information as the name, age, phone number, location, pan card, adhaarcard number, whether the land they cultivate is owned by the farmer or are tenants. This page makes up the important part.

Figure 2. User Data Set

The above dataset has been taken from Kisan call center(data.gov.in). Kisan call centre is a service provided by the Government organizations for welfare of farmers. Kisan call center answer farmer's queries on a telephone call in the native languages. Kisan call centres are present in 21 different locations in the country covering all Indian states and Union Territories.

Figure 3. Schemes and categories dataset

Technologies and methodologies used for developing the proposed system are based on Machine Learning that consists of Collaborative Filtering with model based approach- knn algorithm, Cosine Similarity, Support Vector Machine and Brute algorithm implemented using python.

3.1 Preprocessing

The data is loaded into Pandas Data frame and then pivot ratings into scheme features. To have best data interpretation, data frame is pivoted as schemeid and userid columns. Null values are filled with zeros.

Figure 4. Schemes

3.2 Threshold Values Identification

Considering the count of schemes, users(farmers), and scheme ratings, three things has been understood:

Values of lot of user scheme matrix are zeros. The data we dealt with was sparse data, and hence we worked with scipy-sparse-matrix to avoid wastage of memory and avoid problems of overflow.[7]

Figure 5. Count Of Ratings

3.3 Modelling Recommendation Systems

The data is loaded into Pandas Data frame and then pivot ratings into scheme features.

Collaborative Filtering is a technique used in Recommender Systems. Collaborative Filtering is a method of making instinctive prognosis (filtering) about the interests of a user by collecting liking or taste information from many users(collaborating). These forecast are specific to the user, but uses information collected from many users. Uses cases of collaborative Filtering usually involve very large data sets. Collaborative Filtering systems have many forms, but many common systems can be reduced to two steps.[8]

1. Includes Users who share same rating patterns with the active user.

2. Uses the ratings from the people / user similar to step 1 to calculate a prediction for the active user.

This falls under the subtitle of user based collaborative filtering. The use case of this is the user based Nearest Neighbor Algorithm.

Another sub type is the item based Collaborative Filtering proceeds in an item-centric manner.

1. Build an item-item matrix determining the relationships between pairs of items.

2. Includes the taste of the current user by examining the matrix and matching that users data.
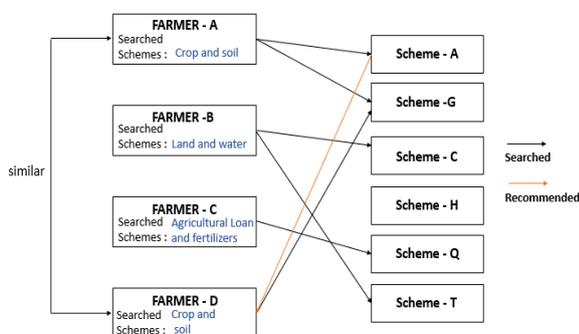

Figure 6. Collaborative Filtering Diagram

Implementing KNN algorithm:

A distance measure needed to determine the "closeness" of the particular instances. KNN uses the similar approach of classifying an object by searching its nearest neighbors and picking the most popular category among the neighbors.[9][10]
Algorithm

1. Read the input data.

2. Select X to the chosen number of neighbors

3. For every value in the data set

    (a) Find the distance between the query value and the current value for the given data set.

    (b) Sum up the distance of value and the index of the value to an ordered collection.

4. Arrange the ordered collection of distances and indices in ascending order by the distances.

5. Select the first X entries from the arranged ordered collection

6. Get the labels of the chosen X entries from collection

7. If it is classification, return the mode of the X- labels.

Following are the limitations of KNN algorithm:

1. Improper value of k reduces the performance.

2. The K Nearest Neighbour Algorithm is slow as it analysis every sample each time.

3. The algorithm is negatively affected by correlated and irrelevant features.

Optimizations in KNN algorithm:

1. We have chosen the best value of k to get optimal result.

2. We have built our own model using cosine similarity instead of methods such as Euclidean or Manhatten distance as we observed it performs better on our data because of higher variance resulting in higher accuracy along with greater classification speedup.

3. The dataset is preprocessed, simple and not correlated [11].

Another modeling algorithm taken into consideration for this project is Support Vector Machine (SVM). The machine learning algorithm is a supervised algorithm being used and implemented for classification challenges on the data sets. This algorithm is mainly used in classification problems. In this SVM algorithm, each data item is plotted as a point in n-dimensional space with each feature value being the value of a particular coordinate. Then the classification is performed by finding the hyper-plane differentiating the two classes very well.

4 Experimental Results

4.1 Analysis of Algorithms

The model has been processed with Support vector machine algorithm which yields less accuracy as compared to the KNN algorithm. After all the analysis we can conclude that k-nearest neighbors algorithm provides better accuracy and precision. The comparison table can better depict the above case [12]:

Algorithm	Accuracy	Precision
K-Nearest Neighbor	0.78	0.87
Support Vector Machine	0.62	0.75

Table 1. Comparison Table Of Algorithms

After the queries are processed and schemes are recommended based on the categories of scheme asked by the farmer. Now, the training data has high dimensionality. Instead of using Euclidean distance, we had used cosine similarity for nearest neighbor search, since Euclidean distance suffers from "curse-of-dimentionality".
The analysis of the system are done and the graphical results are displayed.
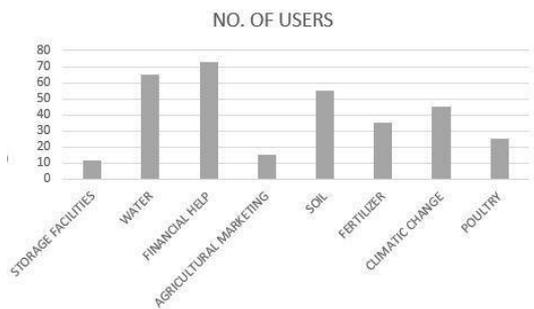
Figure 7. Graphical representation of analysis using cosine similarity

4.2 Precision and Accuracy Measure

Cosine similarity equation :

$$Similarity(A,B) = (A*B)/(|A|*|B|)$$

Accuracy : 0.7824075179015128
Precision : 0.8731884057971014

Benchmark for threshold value is k=3(considering the most appropriate value of k=sqrt(n)/2 where n is number of schemes for farmers)
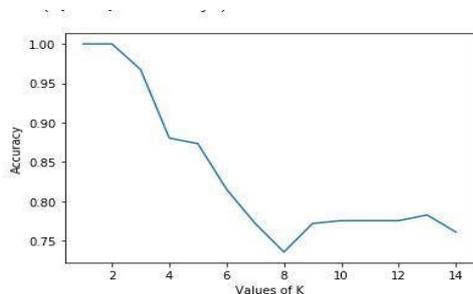

Figure 8. Precision status

5 Conclusion And Future Work

In this paper, the Query Analysis has been done using Collaborative filtering using KNN Algorithm with Gradient Ascent Singular Valued Decomposition and Support Vector Machine. After comparing the results obtained from KNN and SVM implementation, it is observed that KNN performs intuitively better. The accuracy obtained nearly stands at 87percent.Future work is aimed on working with larger data sets and advanced algorithms . The same application can be linked with other government relief facilities to provide one stop solution for farmers.

References

[1] Understanding Machine Learning: From Theory to Algorithms By: Shai Shalev-Shwartz and Shai Ben-David, 2014.

[2] Rakesh Kumar, M.P. Singh, Prabhat Kumar. Crop Selection Method to Maximize Crop Yield Rate using MachineLearning Technique.ieeexplore.ieee.org/document, 2015.

[3] Kiran Shinde , Jerrin Andrei , Amey Oke : Web Based Recommendation System for Farmers, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 3 1444 – 1448, 2015.

[4] Vikas Kumar, Vishal Dave, Rahul Bhadauriya, Sanjay Chaudhary : KrishiMantra: Agricultural Recommendation System , Proceedings of the 3rd ACM Symposium on Computing for Development, 2013.

[5] Sung Hwan Min and Ingoo Han : Recommender Systems Using Support Vector Machines, Proceedings of the 5th international conference on Web Engineering, 2005.

[6] Pandiyan, O. El-Hassan, A. Hassan Khamis, and P.Rajasekaran, "Ontology with SVM based diagnosis of tuberculosis and statistical analysis". International Journal of Medical and Health Sciences Research,2016 ,Vol.3, No.3, pp.37-43, ISSN(e): 2313-2752, 10.18488/journal.9/2016.3.3/9.3.37.43, 2016 Asian Medical Journals, 2016.

[7] Bailey, T. Jain, A." A note on distance-weighted k-Nearest Neighbors rules", IEEE Trans. Systems, Man, Cybernetics, 8:311-313, 1978.

[8] Breese, J.S., Heckerman, D., and Kadie, C. : Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998.

[9] Sadegh Bafandeh Imandoust And Mohammad Bolandraftar :Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background ,S B Imandoust et al. Int. Journal of Engineering Research and Applications Vol. 3, Issue 5, pp.605-610, 2013.

[10] Gou, J., Du, L. Zhang, Y. amp; Xiong, T. quot;A New Distance-weighted k-nearest Neighbor Classifierquot;, Journal of Information amp; Computational Science, 9(6): 1429-1436, 2012.

[11] Hellman, M.E., quot; The nearest neighbor classification rule with a reject option quot;, IEEE Trans. Systems, Man, Cybernetics, 3: 179-185, 1970.

[12] Grčar M., Fortuna B., Mladenič D., Grobelnik M. kNN Versus SVM in the Collaborative Filtering Framework. In: Batagelj V., Bock HH., Ferligoj A., Žiberna A. (eds) Data Science and Classification. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg, 2000.