International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952224

Location Based Agricultural Product Recommendation System Using Novel KNN Algorithm

Sachin J, Geethatharani P, Surya M K, Kavin K V
Computer Science and Engineering, Dr. Mahalingam College of Engineering and Technology, Anna University, Pollachi, Tamil Nadu, India

ABSTRACT

It is evident that the need for personalized product recommendation is much needed these days. Generally, product recommender systems are implemented in web servers that make use of data, implicitly obtained as results of the collection of Web browsing patterns of the users. Here, the project's motive is to provide location-based agricultural product recommendation system using a novel KNN algorithm by ensuring effective communication and transparency in agriculture trade marketing among buyers and sellers (farmers). It helps the farmer to fix up the market price by preventing the rue pricing of their products. The farmer can post their products into the application with price and other details like a timestamp of harvesting, color, size, the absence of pest, freshness, ripeness etc. Based on the location, the distance between the seller and buyer is calculated using great circle distance. An improved Novel KNN algorithm is used to find the K Nearest Seller by calculating the distance between the sellers and buyer using a Euclidean distance metric. The details posted by the farmers and buyers are stored and updated in a database dynamically. The recommender system recommends nearest sellers and their agricultural products based on buyer interest. The performance of the system is analyzed in terms of accuracy and mean absolute error.

Keywords : KNN Algorithm, Recommendation, Distance Calculation

I. INTRODUCTION

Agriculture is one of the important sectors in India. Most of the Indian population is engaged in agriculture occupation. India being an agricultural country is still using traditional ways of recommendations for agriculture. Though agriculture is a crucial part of our economy, due to the lack of use of technology and unawareness of scientific methods to farming, farmers in most cases do not get the preferred output. Success in agricultural growth is believed to be based on an ecologically adapted and economically viable agricultural technology which involves a continuous adaptation to available resources.

Farmers basically belong to rural areas and they being unaware of the market conditions.They have to sell their agricultural products in local market. Even many times buyers, dealers of Agricultural products are unaware about the productions, quality, quantities, and availabilities of different agricultural products produced in different parts of the world.In current years, Buyers don't have sufficient time to go to market and buy the product.Everyone in this world, focus to save their time.

Recommender systems suggest items based on users past behavior, preferences and personal data. Because of the diversity of data, the variety of information and the wide range of products, recommender

systems are very essential in order to provide recommendations for products and other items.

A farmer is a person who works under the umbrella of agriculture, producing a variety of food products for consumption. A farmers market is a physical retail marketplace intended to sell products directly by farmers to consumers where farmers are not getting a good market price. In order to avoid rue pricing of their products, a recommended approach using a collaborative model is used to allow farmers to obtain fixed market price and earn good amount of profit. This idea will be helpful for farmers and market traders to sell and buy agriculture products easily and efficiently. This system intends to provide reliable and efficient communication and interaction platform between buyers and sellers (farmers) in the field of agriculture. Farmers can add information about their crops, quantity, price, and location. On this basis of the provided information, agricultural products will be recommended to buyers or merchants. It will be helpful in facilitating an environment for the remunerative agricultural market for the benefit of both sellers and buyers.

1.1.OBJECTIVE

The objective of the project is to develop an application that recommends the agricultural products based on the buyers and sellers location and search related products by using the Novel KNN algorithm.

II. EXISTING SYSTEM

Recommender systems are filters which suggest items or information that might be interesting and relevant to user's preference. The recommendation system collects raw data from environmental factors and the collected data is processed and then predictive modeling techniques are applied for prediction. Collaborative filtering method could naturally filter the data and recommend up-to-date information. This approach is based on collecting and evaluating a large amount of information on user's behavior, activity or preference and conclude what users will like based on their relationship to the other users.

Disadvantages:
Less accuracy
Performance degrades on very large dataset

III. PROPOSED SYSTEM

3.1 Location Based Collaborative Model Using Knn Algorithm

The proposed system is based on the location and memory-based collaborative filtering Model. The collaborative filtering is implemented using the KNN algorithm. This system optimizes the existing structure.This algorithm is developed to meet the high-performance accuracy of prediction than the existing system. The Recommender system need to predict more accurately to frame the definite structure. The existing system performs slowly on large datasets. The proposed system uses the location-based collaborative filtering which increases the performance and accuracy.

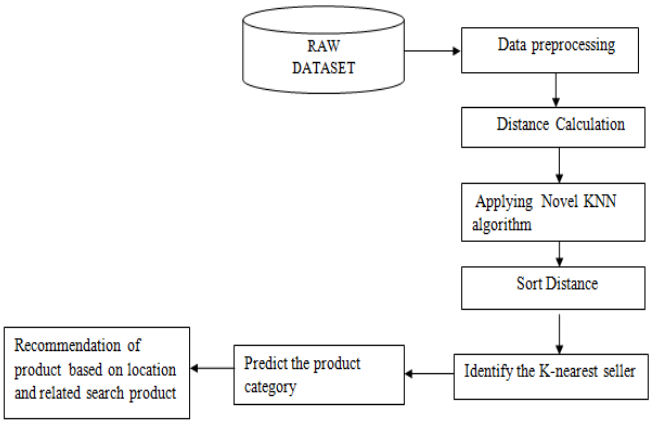

Figure 1. Block diagram of the

Proposed System

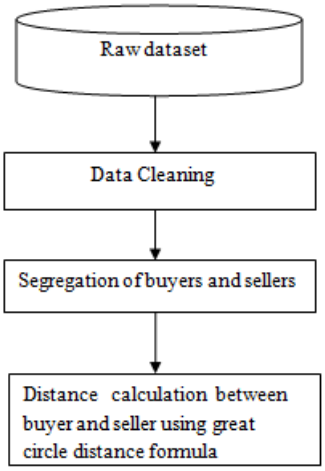

Figure 2. Block diagram of Data preprocessing

IV. METHODS AND MATERIAL

4.1 Data preprocessing

Data preprocessing is the process of data mining technique that transforms the raw data into an understandable format. It identifies the incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or course data.

Each input sequence is preprocessed to generate a set of patterns as training data. The preprocessed data is then used to train the initial weights of the distance. Raw data is preprocessed to get the required format of data. Based on the login type of the users, each sequence of raw datasets is classified. The cleaned data is then used as the training dataset

4.2 Distance Calculation

The distance between each buyer and seller is then calculated using the great circle distance formula. The distance values are then sorted and the sellers are identified based on the distance between the buyers and sellers. For every buyer nearest sellers are identified.

Distance = r * acos[sin(lat1) * sin(lat2) + cos(lat1) * cos(lat2) * cos(lon2 - lon1)]

Equation 1 Great circle distance

Where,

r is the radius of the earth in km.

lat1, lat2 is the latitude coordinate of the location

lon1, lon2 is the longitude coordinate of the location.

4.3 Implementation of the Novel KNN algorithm

The KNN algorithm is implemented and applied on the cleaned and filtered data along with the calculated distance between sellers and buyers. Correlations can be measured by distance metrics such as Euclidean distance. The algorithm uses the training dataset and provides the nearest sellers who are selling the relevant products.

$$\mathrm{d}(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}.$$

Equation 2 Euclidean distance

where n is the number of dimensions. It measures the numerial difference for each corresponding attributes of point p

4.4 Product Recommendation

The K- nearest sellers are then identified and then the product category that the sellers sell are identified and listed to the buyers.K value is defined from 1. The best K value is then chosen based on the accuracy of the predicted value. For example in the sample datasets the K value 3 gives the best accuracy value.

V. RESULTS AND EXPERIMENTS

5.1 Data Set

The KNN approach is tested on the different agricultural product datasets.The data set consists of random data about the product.The length of each set includes Login type, productcode, price, latitude, longitude, Item name, location, name of seller and buyer, phone number.The attributes and its type are listed in Table 1.

Open Government Data Platform(OGD) India is a single-point of access to Datasets/Apps in open format Published by Ministries/Departments of Indian government. Source of Agricultural product dataset.:https://data.gov.in

Table 1 Attribute name and its type

Name	Data type
Product code	Integer
Price	Float
Product label	String
Location	String
Latitude	Double
Longitude	Double
Item name	String

5.2 Evaluation Metric

In this proposed system, the distance between the buyer and sellers are initialized to Zero. The distance is calculated using the great circle distance based on the latitude and longitude coordinates of buyers and sellers.Later the Novel KNN algorithm is used to predict the label. Imbalanced datasets are evaluated using F-Score, because of the dominating effect of the majority class. F-Score considers both the precision and the recall of the test to compute the score.

The performance and accuracy are calculated based on the following measures

Let S, B, Rb denotes,

S Total number of sellers Items in dataset collection

B Set of buyers who got the recommendation

Rb Set of buyers who got the relevant product

TP True positive

FP False positive

FN False negative

5.2.1 Accuracy

It is the ratio of the total number of buyers who got the relevant predicted product labels to the total number of buyers who recommend.

$$\text{Accuracy} = \frac{|R_b|}{|B|}$$

Equation 3 Accuracy

5.2.2 Precision

It is the ratio of the number of buyers in true positive to the sum of the true positive and false positive. It is usually expressed as a percentage. TP and FP values are calculated from the Table

$$\text{Precision} = \frac{|TP|}{|TP + FP|}$$

Equation 4 Precision

5.2.3 Recall

It is the ratio of the number of buyers in true positive to the sum of the true positive and false negative. It is usually expressed as a percentage

$$\text{Recall} = \frac{|TP|}{|TP + FN|}$$

Equation 5 Recall

5.2.4 F-Score

The measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score is

$$F - \text{Measure} = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}}$$

Equation 6 F-Score

The F-score is used for evaluating the accuracy of product recommendation based on features such as product, location, and price. The higher values indicate better prediction accuracy.

VI. SUMMARY OF RESULTS

The agricultural product labels are predicted based on its relevant search product label and the nearest location for the buyers. It is seen that k value with 3 works best for the agricultural product recommendation dataset. It is seen that the dynamic distance calculation and S-KNN algorithm works better. Thus the product of nearest sellers, are recommended to the buyers based on the related search products. This approach gives about 82% of accuracy on predicting the product label.

Table 2 Accuracy for the given K-Value

K value	Accuracy
2	60%
3	82%
4	68%
5	70%
6	67%
7	60%
8	48%

VII. CONCLUSION

Thus, Novel KNN can be used to predict the product labels and recommend based on the high probability class. The Novel KNN along with the dynamic distance calculation improves the performance of the product recommendation on the agricultural product dataset. However, accuracy of the prediction varies according to the k-value. This approach works well for k-value=3 on this particular dataset.

VIII. ACKNOWLEDGMENT

Our hearty thanks to our guide **Ms. J. Aiswarya** Assistant Professor for her constant support and guidance offered to us during the course of our project by being one among us and all the noble hearts that gave us immense encouragement towards the completion of our project.

IX. REFERENCES

[1]. Zhibian pan, Yidi Wang, Weiping Ku, "A new general nearest neighbor classification based on the mutual neighborhood information", Science Direct on Knowledge-based System, Vol. 2 , pp.142-152, 2017.

[2]. David S.Zamar, Bhushan Gopaluni, Shahab Shokhansaji, "A Constrained K-Means and Nearest Neighbor Approach for Route Optimization in the bale Collection Problem", International Federation of Automatic Control, Elsevier 2017.

[3]. Shichao Zhang, Debo Cheng, Zhenyun Deng, Ming Zong, Xuelin Deng, "A novel KNN algorithm with data-driven k parameter computation", Elsevier, Vol.3 , pp.44-54 , 2017.

[4]. Moon-Hee Park, Jin-Hyuk Hong, and Sung-Bae Cho, "Location-based recommendation system using Bayesian algorithm", Dept. of Computer Science, Yonsei University,pp.120-749,2007.

[5]. Robert Pálovics, Péter Szalai, Júlia Pap, ErzsébetFrigo, Levente Kocsis, AndrásA.Benczúr, "Location-aware online learning for top-k recommendation", Institute for ComputerScience and Control, Hungarian Academy of Sciences, Hungary.

[6]. Shyr-Shen Yu, Shao-Wei Chu, Chuin-Mu Wang, Yung-Kuan Chan, Ting-Cheng Chang, "Two improved k-means algorithms", Elsevier,2017.

[7]. Semem ben Salem, Sami Naouali, Zied Chtourou, "A fast and effective partitional

clustering algorithm for large categorical dataset using a k-means based approach", Vol. 68, pp.463-488, 2018.

[8]. Chiu-ChingTuan, Chi-FuHung∗, Zong-HanWu, "Collaborative location recommendations with dynamic time periods", Department of Electronic Engineering, National Taipei University of Technology, Vol. 35, pp.61-48 ,2017.

[9]. Chiu-ChingTuan, Chi-FuHung∗, Zong-HanWu, "An adaptive point-of-interest recommendation method for location-based social networks based on user activity and spatial features", Department of Electronic Engineering, National Taipei University of Technology, Vol. 38,pp.470-584, 2017.

[10]. Rong Gao, Jing Li*, Bo Du, Xuefei Li, ChengFang Song, Jun Chang, and Donghua Liu1 "Exploring Spatial-Temporal Sequential Influence and Social Information for Location Recommendation", of Computer Science, Wuhan University, Wuhan, Vol. 19, pp.790-794,2018.

[11]. Python math: Distance between two points using latitude and longitude"[Online].Available: https://www.w3resource.com/python-exercises/math/python-math-exercise-27.php.[Accessed July 2018].

Cite this article as :

Sachin J, Geethatharani P, Surya M K, Kavin K V, "Location Based Agricultural Product Recommendation System Using Novel KNN Algorithm", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 945-950, March-April 2019.

Available at doi :
https://doi.org/10.32628/CSEIT1952224

Journal URL : http://ijsrcseit.com/CSEIT1952224