

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/358295584>

# Machine Learning based Recommender Systems for Crop Selection: A Systematic Literature Review

Preprint · January 2022

DOI: 10.21203/rs.3.rs-1224662/v1

CITATIONS

0

READS

1,117

5 authors, including:



**Younes OMMANE**

Mohammed VI Polytechnic University

6 PUBLICATIONS 1 CITATION

SEE PROFILE



**Mohamed Amine Rhanbouri**

Université Mohammed VI Polytechnique

1 PUBLICATION 0 CITATIONS

SEE PROFILE



**Hicham Chouikh**

1 PUBLICATION 0 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Adaptive Ridge regression [View project](#)



Predictive density estimation [View project](#)

# Machine Learning based Recommender Systems for Crop Selection: A Systematic Literature Review

Younes OMMANE (✉ [younes.ommame@um6p.ma](mailto:younes.ommame@um6p.ma))

Université Mohammed VI Polytechnique

Mohamed Amine RHANBOURI

Université Mohammed VI Polytechnique

Hicham CHOUIKH

Université Mohammed VI Polytechnique

Mourad JBENE

Université Mohammed VI Polytechnique

Ikram CHAIRI

Université Mohammed VI Polytechnique

---

## Research Article

**Keywords:** Recommender Systems, Collaborative Filtering, Crop recommendation, Crop selection, Machine Learning, Agricultural management practices

**Posted Date:** February 2nd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1224662/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

Crop selection (CS) is one of the most critical elements that affects the final yield directly. As a result, selecting an appropriate crop is always a critical decision that a farmer must make, considering environmental factors. Choosing an appropriate crop for a given farm is a difficult decision including a plethora of variables that influence the final yield. Experts are frequently consulted to assist farmers with CS; but, as this alternative is time consuming and expensive, it is not available to many farms. The use of recommender systems (RSs) in agricultural management has recently brought some captivating and promising results. We propose a systematic literature review (SLR) in this article to find and provide the most relevant and high-quality publications addressing the crop recommendation (CR) question. The core concept of this SLR is inspired from the guidelines of PRISMA 2020. The different CR approaches are discussed, as well as all the most important input features for recommendation, which are determined and classified. We also identified some of the biggest hurdles to using CR in agriculture. Besides, we made an inventory of the most used techniques for CR. Further, we made an inventory of evaluation criteria and evaluation approaches.

## 1. Introduction

Given the population growth in the last decade, it is becoming extremely difficult to ignore the importance of agriculture, except developed countries and also emerging countries such as India, which is a good example in matter of giving importance to Agricultural development since its independence, which even brought about the so-called “*Green Revolution*” in the late 1960s. However, achieving an improved and sustainable agricultural production largely depends on the advancement of the agricultural management, hence, the advance of agricultural research and its effective applications at farms, through the transfer of technology and innovation. The access to the right information at the right time gives farmers the ability to make accurate and beneficial decisions that positively affect their livelihoods. Considering these facts, RSs algorithms quickly became the most widely used solution, to face the growing amount of agricultural digital information. RSs has become an active research area during last recent years in various domains such as marketing, e.g. streaming platforms, Net e-commerce websites, advertising on social media. Several works showed the effectiveness of using those models to better recommend items having the best chance to be sold to a user.

RSs experienced a huge growth for their enormous benefits in supporting user’s needs, through finding the most suitable items based on information extracted from a collection of data. These systems also play an important role in decision-making, helping users to maximize profits or minimize risks, e.g. Amazon store. Today, RSs are used in many digital companies, such as: Google, Yahoo, Netflix, etc., RSs are applied in different areas, such as: healthcare systems, education, customer segmentation, fraud detection, and financial banking [1].

RSs are being used in CR to provide farmers with better decisions. However, the CR field does not have a detailed classification scheme for its algorithms and features, mainly due to the diverse approaches proposed in literature, as well as the absence of an SLR dedicated to this issue, where most studies reviewed only the use of machine learning (ML) in crop yield prediction [2, 3]. Therefore, it’s difficult and confusing to choose a RS algorithm and input parameters that fit one’s need, when developing a crop RS. In addition, researchers may find it challenging to track the use and the trends of RSs algorithms in agriculture. For these reasons, this SLR comes to fill this gap the best possible. We decided to present a fair, unbiased, and credible SLR, that identifies all relevant and high-quality studies addressing the integration of RSs in CS with the following main objectives:

- Identify trends of RSs algorithms in CR.
- Classify the main techniques that were used in CR.
- Classify the main input features.
- Identify evaluation criteria and evaluation approaches that have been used.
- Specify the current challenges.

This paper is organized as follows: Section 2 gives a general overview of RSs. Section 3 presents the SLR’s protocol. In section 4, an analysis of the results is provided for CR. Section 5 covers a discussion of the current achievements and current challenges. Finally, we conclude the analysis and present our future work.

## 2. Overview Of Existing Recommender Systems

The objective of a recommender system is to provide the user with relevant recommendations according to their preferences. It drastically reduces the time it takes for the user to search for the items that are most interesting to them, and also to find items that they are likely to like but that they might not have paid attention to. Recommendation systems have been defined in several ways. The most popular and general definition that we quote here is that of *Robin Burke* [4], which is as follows: “a recommender system is a system capable of providing personalized recommendations or of guiding the user to interesting or useful resources within a large data space.” The information domain for a general recommendation system consists of a list of users who have expressed their preferences for various items. A preference expressed by a user for an item is called a rating, and is often represented by a triplet (user, item, rating). These notes can take different forms. However, most systems use ratings in the form of a scale of 1 to 5, or binary ratings (like/dislike). The set of triples (user, item, note) form what is called the note matrix. The pairs (user, item) where the user did not give a score for the item are unknown values in the matrix. Firstly, there is non-customized recommendation systems that do not depend on the user for making recommendations. In non-customized recommendation, the

used algorithms are: Top Popular, which recommends the top items (e.g., movies) with the highest ratings, and Product Association, which recommends the best combinations of items that are frequently bought together [5].

Customized approaches are other techniques that provide a recommendation to users based on ratings or content information. All these techniques are either focused on characteristics of an individual (or group) or characteristics of the product or service they are buying for example. In customized RS the following methods are used:

Knowledge-based (KB) filtering is a technique that employs explicit knowledge to identify user's preference based on knowledge of items, user, and matching between both. It is specifically effective in cases with less data on activity history of users [6].

RSs that are based on demographic information of users suggest a list of items that have good feedback from the users that are demographically similar [7]. The advantage of a demographic technique is that it does not require a history of user's feedback.

The vast majority of prior research had adopted the collaborative filtering (CF) approach, which employs a family of algorithms that calculates the utility and the rating of an item for a given user. Two general classes as illustrated in Figure.1 were suggested in literature [8], Memory-Based (MB) algorithms, where the predicted value is estimated as a simple linear combination of ratings and weights, either explicitly or implicitly. The weights can reflect distance, correlation, or similarity between either users or items, this similarity function is the hyper-parameter that affects mainly the prediction quality. Several authors adopted Pearson, Euclidean and Cosine functions as similarity criteria, others used the development of genetic algorithms to find the most suitable combination of weight vectors [9, 10]. Typical examples of this approach are neighborhood-based CF and item-based/user-based top-N recommendations.

MB algorithm are not always as fast and scalable as they are meant to be, especially in the context of actual systems that generates real-time recommendations based on very large datasets. To achieve these goals, MB RSs are used. MB CF involves building a model based on the dataset of ratings and have been used as a "model" to make recommendations without having to use the complete dataset every time. This approach potentially offers the benefit from a probabilistic perspective, either by the calculation of the expected value of a rating given the user's historical data, or using dimensionality reduction techniques, such as: matrix factorization (MF); in order to model the latent factor space. Studies in literature have examined the possibility of exploiting deep neural networks (DNN) architectures [11, 12, 13, 14], Convolutional Neural Networks [15, 16], Recurrent Neural Networks [17] and Auto-Encoders [18, 19]. To capture more complex and nonlinear relations above the ratings dataset, there are many MB CF algorithms. Bayesian networks, clustering models, latent semantic models, such as: Singular Value Decomposition (SVD), Probabilistic Latent Semantic (PLS) analysis [20], Multiple Multiplicative Factor (MMF), Latent Dirichlet Allocation (LDA) [21] and Markov Decision Process (MDP) based models that includes Contextual Bandits approach [22, 23] and Reinforcement Learning (RL) [24].

Nonetheless, CF algorithms suffer from three common problems, namely:

- **Cold start** reflects the inability of suggesting recommendation to a new user or item due to data absence.
- **Sparsity** problem occurs when available data are insufficient to identify similar users.
- **Scalability** problem happens when the RS's performance and latency decrease drastically with an increase in the number of users and items of the system.

And finally, One way to transcend these hassles is the combination of CF and Content-Based Filtering (CBF) methods, which are another family of algorithms that makes recommendations based on user preferences of product features, in a Hybrid Filtering (HF) technique [25, 26].

Due to the successful use of RSs in various advertisement sectors, they have been applied to solve a variety of problems in the agricultural sector. In [27] authors proposed an ontology based on an RS that helps to identify the pests affecting a crop and their treatments. In [28] a cultivation calendar RS for wheat cultivation in Egypt based on climate data is developed. In another work [29], a hybrid technique for recommendation of agricultural products to buyers is used. In [30] a CF web-based RS was designed to provide help, such as: financial help, irrigation facilities and insurance to the farmer's crops.

CS is one of the fundamental issues that have a strong influence on farmer's revenue, and the applications of recommendation techniques has shown a significant progress recently, especially for these specific tasks. Henceforth, the remainder of this analysis will focus on scientific papers CR.

### 3. Research Formulation

To present a clear review of the recommendation techniques applied in agriculture, we followed the SLR protocol adopted in PRISMA [31], in its latest version (2020). In the following sections we formulate the Research Questions (RQs) which we will try to respond in this SLR, then, we explain the adopted search strategy for collecting papers, followed by the exclusion criteria that serve as a filter to select relevant papers for review, and finally in the data extraction phase, the information needed for the analysis of the selected papers is extracted.

#### 1. Questions Formulation

In order to accomplish the objective of this SLR and get a full analysis of CR techniques, we defined the following RQs:

- **RQ1.** How did research about CR evolved over time?

- **RQ2.** What are the main techniques used in literature for CR?
- **RQ3.** What are the main input features exploited for CR?
- **RQ4.** Which evaluation parameters and evaluation approaches have been used?
- **RQ5.** What are the current challenges in CR?

## 2. Search strategy

To identify relevant studies, we first started by performing a selection of the major scientific databases, such as: Google Scholar, ACM, Springer Link, IEEE, Wiley, Emerald, etc.

There are some synonyms that indicate CR Systems. In this SLR, we consider terms that replace “recommendation” by “selection” or “suggestion”. To retrieve studies that use new techniques that are based on agricultural data, we used the terms “Artificial Intelligence”, “Machine Learning”, “Deep Learning”. All these terms were featured in the Search Query (SQ), which is presented as follows:

Query = (Crop AND (Recommendation OR Selection OR Suggestion) AND (“Artificial intelligence” OR “Machine Learning” OR “Deep Learning”))

## 3. Exclusion criteria (EC)

To strengthen the validity of the SLR, we considered only studies published between 2010 and 2020, and we maintained those that have no evaluation section due to the lack of research publications about the CR field. We adopted the following EC:

- **EC1.** Studies must be peer-reviewed articles or proceedings.
- **EC2.** Studies must be published in a conference, journal, press, etc.
- **EC3.** Letters, notes, and patents are not included in the review.
- **EC4.** Graduate reports are not considered for review.
- **EC5.** We considered only studies in English.
- **EC6.** Studies that do not describe their proposed approach in a proper way were not considered in the review.

## 4. Data collection process

To answer the aforementioned RQs, data from the selected articles have been collected and structured. The extracted information focused on verifying whether

- Paper reference
- Year and type of publication
- The Indexing Database
- The country origin of study
- Models used to address the problem
- The inventory and description of crops and features used in each study
- Performance measures used to evaluate the proposed models

# 4. Results

This section presents the harvest of the selection process. Then, we present the information matrix along with results of each RQ.

## 1. Study selection

Based on the obtained results from the aforesaid scientific databases, 89 results were identified. We followed a filtering process to eliminate the articles that do not match our selection criteria. We first excluded 18 records because they were not indexed in well-known academic indexing services. Afterwards, we started scrutinizing the articles’ content by reading the title, abstract and keywords, after that we kept 50 articles for further investigation, in addition to 2 more articles that were found in references. By full reading of the 52 articles of this study, we excluded another 12 articles, as they are either not clearly relevant, or are out of our scope. Thus, we ended up with 40 papers for synthesis and analysis. Fig 2 illustrates a flow chart of the paper filtering process.

## 2. Literature Review Matrix (LRM)

One of the major problems that the farmers face in the beginning of every agricultural season is the selection of a suitable crop that would produce a better yield. This process is usually done based on the farmer’s experience, or with the help of an agronomist. Assisting this process was the objective of a huge number of papers throughout the past few years. Table 4.2 puts forward the LRM presenting in detail the studied articles that addressed this problem, based on different approaches and techniques.

Reference	Year	Contribution	Dataset	Models & techniques	Strengths
[32]	2014	In this paper, a Multi-Level Linguistic Fuzzy Decision Network (LFDN) method is applied to a real case dataset to decide the cultivate crop among four crops.	<b>Crops:</b> Wheat, Corn, Rice, and Faba bean. <b>Features:</b> Temperature, Water, Marketing and Soil.	<b>Multi-Level LFDN.</b>	The method can rank the ac to select the appropriate alter
[33]	2014	This work Integrated artificial neural networks (ANN) with geo-graphical information system (GIS) to assess the suitability of land to cultivate a selected crop.	<b>Crops:</b> Rice. <b>Features:</b> Rainfall, Temperature, Elevation and Slope.	<b>ANN, and Back-propagation.</b>	High consistency in predictin
[34]	2015	This paper presents a technique named CS method to select a sequence of crops based on crop proprieties to improve net yields rate of crops to be planted over a season.	<b>Crops:</b> Seasonal crops, Whole year crops, Short time plantation crops and longtime plantation crops. <b>Features:</b> Geography of a region, Weather conditions, Soil type and Soil texture.	<b>CSM.</b>	CSM method retrieves all pos be sown at a given time start
[35]	2015	In this paper, a rule system is developed to help farmers make decisions about the choice of rice varieties using the crop and the properties.	<b>Crops:</b> 118 rice varieties. <b>Features:</b> 7 Features.	<b>Rule system.</b>	The set of production rul KB and farmer's land profile t varieties.

Reference	Year	Contribution	Dataset
[36]	2016	In this article a hybrid soft decision model has been developed to take decisions on agriculture crop, that can be cultivated in each experimental land.	<b>Crops:</b> paddy, groundnut, sugar- cane, cumbu and ragi. <b>Features:</b> Twenty- seven input criteria, namely s input (6 sub criteria), support, facilities, and risk
[37]	2016	The proposed system is designed to predict the best suit- able crops for a given farm, and to suggest farming strategies, such as: mixed cropping, spacing, irrigation, seed treatment, etc. along with fertilizers and pesticides.	<b>Crops:</b> 44 crops that have been considered <b>Features:</b> Crop name, suitable rain- fall, temperature, cost, soil, and pH.
[38]	2016	The authors applied the majority voting technique using Random tree, CHAID, K- nearest neighbors (KNN) and Naiïve Bayes as base learners for CR.	<b>Crops:</b> Millet, groundnut, pulses, cotton, vegeta sorghum, sugarcane, coriander. <b>Features:</b> Depth, Texture, pH, Soil Color, Permea holding and Erosion.
[39]	2017	Proposed a system that can detect the user's location then recommend top-k crops based on the seasonal information and crop production rate (CPR) of each crop of similar farms.	<b>Crops:</b> Not mentioned. <b>Features:</b> Crop Growing period database, Thermal zone database, Physiographic database, Seasonal crop database and CPR database.
[40]	2017	In this article, an at- tempt is developed to predict crop yield and price that a farmer can obtain from his land, by analyzing patterns in past data.	<b>Crops:</b> Not mentioned. <b>Features:</b> Crop areas, types of crop cultivated, n and the overall crops consumed.

Reference	Year	Contribution	Dataset
[41]	2017	In this study, a FL- based expert system is proposed to auto- mate the CS for farmers based on parameters, such as, the climatic and soil conditions.	<b>Crops:</b> 20 crops. <b>Features:</b> 23 features.
[42]	2017	In this paper, a decision-making tool is developed for selecting the suitable crop that can be cultivated in each agricultural land.	<b>Crops:</b> Paddy, groundnut, and <b>Features:</b> 26 input variables
[43]	2017	This paper develops a fuzzy based agricultural decision support system which helps farmers to make wise decisions regarding CS.	<b>Crops:</b> Not mentioned. <b>Features:</b> 15 parameters.
[44]	2017	This paper proposed two mathematical formulations, the first one for the determination of crop-mix that maximizes the farmer's expected profit, and the second model that maximizes the average expected profit under a predefined quantile of worst realization.	<b>Crops:</b> Corn, Wheat, Soy, Bar <b>Features:</b> the land available 1 required for each crop, the co
[45]	2018	This paper Suggested using deep neural network for agricultural CS and yield prediction.	<b>Crops:</b> Aus rice, Aman rice, B <b>Features:</b> 46 parameters

Reference	Year	Contribution	Dataset
[46]	2018	The article proposes a new system for CR based on an ensembling technique.	<b>Crops:</b> Cotton, Sugarcane, Rice, Wheat. <b>Fe</b> Soil Type, pH value of the soil, NPK conten the soil, Porosity of the soil, Average rainfa Surface temperature, Sowing season.
[47]	2018	This paper presents an intelligent system, called Agro-Consultant, which assists farmers in making decisions about which crop to grow.	<b>Crops:</b> 20 crops. <b>Features:</b> Soil Type, Soil pH, Precipitation, Temperature, Location par
[48]	2018	This paper Investigated the predictive performance of different data mining classification algorithms to recommend the best crop for better yield, based on a classification of soil under different ecological zones.	<b>Crops:</b> Not mentioned. <b>Features:</b> pH, Organic Matter, K, EC, Zn, Fe, Cu and texture.
[49]	2018	The article proposes a system that gives the farmer a prior idea regarding the yield of a particular crop by predicting the pro- duction rate according to the location of the farmer and the past data of weather conditions.	<b>Crops:</b> Rice. <b>Features:</b> Temperature, humi location, and rainfall.
[50]	2019	This study proposed to design a KB solution for building an inference engine for recommending suitable crops for a farm.	<b>Crops:</b> Not mentioned. <b>Features:</b> Elevation, temperature, fertilizer rainfall, field type, seed type and soil.

Reference	Year	Contribution	Dataset
[51]	2019	A new datamining technique was proposed to cluster crops based on their suitability compared to soil nature of different areas.	<b>Crops:</b> 10 Crops. <b>Features:</b> Soil, crop, tempe
[52]	2019	In this paper, a multi-class classification- based decision model has been developed to assist the farmer in selecting suitable crops using rough, fuzzy, and soft set approaches.	<b>Crops:</b> Paddy, groundnut, sugarcane, cun <b>Features:</b> 27 features.
[53]	2019	This study proposes an intelligent agriculture platform that manages and analyses sensors data to monitor environmental factors, which provides the farmer a better understanding of crop suitability.	<b>Crops:</b> Celery, water spinac <b>Features:</b> Temperature, h conductivity (EC), soil mois
[54]	2019	This paper suggests the using of ANN and SVM for crop prediction considering the environmental parameters.	<b>Crops:</b> Not mentioned. <b>Features:</b> Rainfall, minimu
[55]	2019	The paper proposes a mobile application that will allow farmers to predict the region's production for a specific crop.	<b>Crops:</b> Not mention <b>Features:</b> Soil type, temp

Reference	Year	Contribution	Dataset
[56]	2019	This study applies learning vector quantization (LVQ), which is part of the ANN method, to provide recommendations from three types of plants.	<b>Crops:</b> Rice, corn, and soybeans <b>Features:</b> Altitude, rainfall, temp
[57]	2019	This paper presents a stand-alone crop recommending device that detects soil quality and recommends a list of crops based on a FL models.	<b>Crops:</b> 30 Crops. <b>Features:</b> pH level, soil moisture
[58]	2019	This paper presents a hybrid crop RS based on a combination of a CF technique and a case-based reasoning.	<b>Crops:</b> Not mentioned. <b>Features:</b> Temperature data, rair evapotranspiration.
[59]	2019	This work proposes a model that can predict soil series with land type, according to which it can suggest suitable crops.	<b>Crops:</b> Not mentioned. <b>Features:</b> Soil dataset
[60]	2019	The article addressed the problem of selection of best suitable crop for a farm, by applying different classification algorithms.	<b>Crops:</b> 15 Crops. <b>Features:</b> Soil color, pH, average
[61]	2019	This work proposes an ontology-based recommendation system for crop suitability recommendation based on region and soil type.	<b>Crops:</b> Not mentioned. <b>Features:</b> Soil characteristics, w



Reference	Year	Contribution	Dataset
[62]	2019	This paper proposed a hybrid RS based on two classification algorithms by considering various attributes.	<b>Crops:</b> 24 Crops. <b>Features:</b> 15 Features.
[63]	2020	The article proposes using hierarchical fuzzy model to reduce the classical system complexity with the huge number of generated rules.	<b>Crops:</b> Not mentioned. <b>Features:</b> Sand, silt, clay, nitrogen, phosphorus, potassium, soil color, soil pH, water resources.
[64]	2020	This work developed an application that helps selecting the most convenient type of crops in a certain zone considering the climate conditions of that zone, the production, and the needed resources for each crop.	<b>Crops:</b> Peach, Pear, Apricot, and Almond. <b>Features:</b> Relative humidity, Radiati direction, Cooling units, Sunlight, Rainfall, Accumulated radiation, and Wind
[65]	2020	This paper proposed an implementation of a fuzzy-based rough set approach to help farmers in deciding on CS in their agriculture land.	<b>Crops:</b> 23 Crops. <b>Features:</b> 16 Features.
[66]	2020	This work proposes a CR system according to multiple properties of the crop and land.	<b>Crops:</b> 24 crops. <b>Features:</b> Soil types, pH, Electric Conductivity, Organic Carbon, Nitrogen, Phosphorus, Sulfur, Zinc, Boron, Iron, Manganese, Copper.

Reference	Year	Contribution	Dataset
[67]	2020	This paper proposes a CS method to maximize crop yield based on weather and soil parameters.	<b>Crops:</b> 10 Crops. <b>Features:</b> Soil type, soil nutrients, soil p conditions.
[68]	2020	This study Proposes a clustering center optimized algorithm by SMOTE, then use an ensemble of RF and weighed SVM to predict the recommended crop.	The data includes 1530 soil samp 13 types of cultivated land crops.
[69]	2020	This paper treats the integration of AHP and POPSIS with GIS to determine most suitable crops for parcels for land consolidation areas.	<b>Crops:</b> Corn, Clover, Sugar beet and Wh <b>Features:</b> 63 Land Map Units and their topographical, and socio-economic features.
[70]	2020	This article proposes a system for predicting the crop which has maximum yield per unit area in a district.	A dataset published by the Government of Maharashtra, India, cor points. <b>Features:</b> 7 parameters for the years 19
[71]	2020	This article suggests a FL-based CR system to assist farmers in selecting suitable crops.	<b>Crops:</b> Paddy, Jute, Potato, Tobacco, Wheat, Sesamum, Mustard and Green <b>Features:</b> 11 soil parameters, elev And rainfall.

### 3. How did research about CR evolved over time?

The uprise of new technologies in solving agricultural problems is an immi- nent fact. We can see in Figure 3a the distribution of selected studies over time, and we notice a remarkable increase in the number of studies related to CR from 2014 to 2020. Figure 3b illustrates the distribution of papers based on source database. Most of the selected papers were published in *IEEE* or *Springer*, and fewer papers were found in *Wiley* database. Furthermore, Figure 3c presents the

proportion of each type of publication, nearly 62% of papers were published in imminent conferences, and about 31% comes from journal issues, while only 7.14% are book chapters, which enforce the quality of the publications included in the SLR.

#### *4. What are the main techniques that were used in the literature for CR?*

RSs are generally classified into three types: CBF, CF, or HF. The CF based model was used the most in our literature review, as discussed in section 2, CF tries to compress the entire database into a model, then performs its recommendation task by applying reference mechanism into this model. We identify two common approaches for MBCF: clustering and classification. Clustering CF assumes that users of the same group have the same interest, so they are partitioned into groups called “clusters”. The authors in [61] proposed K-means clustering (KMC) algorithm, which is an unsupervised learning algorithm used to find out fertilizers with NPK contents that are the nearest to the requirements of a specific crop. First it calculates the required amount of the fertilizer, then the algorithm forms clusters of nearby fertilizers based on the Euclidean distance. Therefore, fertilizers in clusters with minimum distance are recommended to farmer. The recommendation task can be viewed as a multiclass classification problem, which uses a classifier supervised learning algorithm that maps the input data to a specific output, a variety of these classifiers were tested on agricultural data. In this context, this study [48] carried out a comparative experiment on data instances from Kasur district, Pakistan for soil classification using J48, BF Tree and OneR, that are a variety of DT based models, which is the most used technique in our literature survey. Besides Naive Bayes Classifier (NBC) that has a significant outcome, mainly because it encodes dependencies among different features by which it connects the causal relationships between items. On the other hand, [54] has investigated the use of SVM and ANNs. The results indicate that the ANN model captures non-linearities among features of the dataset, marking the best accuracy and prediction rate compared to SVM. Another technique for CS that can improve the model's accuracy is ensemble learning, [46] exploited this technique in order to build a model that combines the predictions of multiple ML algorithms together and recommend the right crop with a high accuracy. The independent base learners used in the ensemble model are RF, NBC, and Lin-ear SVM. Each classifier provides its own set of class labels with an acceptable accuracy. The labels class of individual base learners are combined using the majority voting technique. The CR system classifies the input soil dataset into the recommendable crop type, Kharif and Rabi (Autumn and Spring). One of the most promising models in CF is FL, which extracts IF-THEN rules from the provided data using a membership function and linguistic variables that expresses the human knowledge. The authors in [49] proposed a fuzzy based model that uses 27 rules with 3 modalities: Low, Medium, and High. In this traditional single-layer fuzzy system, the rules are exponentially increased when the system's parameter increased, and a larger rule base will affect the system performance and transparency. Therefore, [63] developed a multi-layer system by using the fuzzy hierarchical approach. The hierarchical fuzzy model was applied in the same Mamdani[1] fuzzy inference system for a suitable CR system. The CR system has 12 input variables, and it was decomposed into six fuzzy subsystems, then arranged by priority. The results show that a hierarchical CR system provides a better performance than a traditional fuzzy CR.

CFMB models have a low frequency in the reviewed studies, even though they represent the most common approach in RSs. Yield prediction is based on similarity relationships among items (farms or crops), in terms of collected production yield. For instance, [39] proposed a model consisting of calculating the PCS between farms using the information stored in the crop growing period database, the thermal zone database, and the physiographic database, then, they select the top-n similar farms. The seasonal information and CPR of each crop of the similar farms are used for filtering the first appropriate list to the context. Finally, they recommend the top-k crops to each user respectively. Another study [49] used the similarity approach, in which they developed a system that gives the farmer a prior idea regarding the yield of a particular crop, by predicting the production rate. The COS measure is used to find the similar farmers in terms of location from the database. Then, the resulting farmers that are like the querying farmer form the database for the fuzzy algorithm.

CBF, one of the most significant models in RSs, which are of a high importance for CR, as well as yield estimation. Examples of CBF model applications included in [66], the model of this study is based on the contents that use soil and crop properties, then suggests the list of five high priority crops based on the corresponding properties between the crop and the land for matching soil properties. The algorithm takes two inputs, the land soil details and the re-quired property value for each crop. Primitively, the algorithm computes the similarity between the land and the crop, based on their properties to predefine a range. If the comparison falls in a predefined range, they generate a rank for the combination of crops and lands. In another study [51], authors developed a new data mining technique to cluster the crop based on the suitability of a crop against the soil nature of areas. Features are extracted from the datasets using five different feature extraction metrics, such as, pH distance calculation, NPK (Macro nutrition distance calculation), MICRONUT (Micro nutrition distance calculation), water requirements, and temperature requirements. Then, the crops are clustered using hierarchical clustering based on the vectors into three groups, namely: most suitable, less suitable, or least suitable.

HF is another significant category of models used in CR. In the first study [64], authors presented a new method, which is integrated within an IOT system, that is developed to advise farmers which crop type will generate more y clustering technique is proposed to the obtained groups that have been characterized by their weather conditions. The extracted knowledge forms the model and the rules engine. Finally, the RS generates an ordered list of crops that are suitable in descending order. In the second study [37], the authors developed a CR (hybrid) system, which utilizes FL to choose from 44 crop rules. The system is based on FL, which gets input from an ANNs based weather prediction module. An agricultural named entity recognition module is developed using conditional random field to extract crop conditions data. Further, cost prediction is established based on a LR equation to aid in ranking the crops recommended.

Table 3 shows how many studies describe an approach in each of the classes described earlier in section 2, as well as the studies themselves according to the approach category. As an outcome, a significant number of CF approaches when developing RSs are observed. Over half of the reviewed studies indicated that CF is the most used approach, with a stronger emphasis on a MB method. Perhaps, the availability of historical datasets of farmers linked to the marked dominance of CF in the last years.

Figure 4 traces the timeline of publications, this latter confirms that CF with a MB method has a continual growth. The graph shows that there has been a slight increase in the two recent years in this field and the number of studies is likely to increase after 2020.

Another important conclusion drawn from Table 3 and Figure 4 is the scarcity of research efforts focused on other filtering methods. Nevertheless, some studies showed that the CBF and HF give more accurate recommendations overall than all other types of filtering. However, throughout the years, the research pace on these types of filtering has been relatively low.

Classification of RS	Number of studies	References
CF / Model-based	28	[32, 33, 34, 38, 40, 41, 42, 46, 52, 61, 53, 54, 55, 56, 59, 47, 63, 57, 69, 62, 68, 71, 43, 65, 60, 70, 67, 45]
CBF	5	[35, 48, 51, 66, 44]
HF	5	[36, 37, 50, 58, 64]
CF/MB	2	[39, 49]

Table 3: Articles by type of recommendation technique.

[1] First introduced as a method to create a control system by synthesizing a set of linguistic control rules obtained from experienced human operators. In a Mamdani system, the output of each rule is a fuzzy set.

Table 5 shows the distribution of applied ML algorithms in this study. Some papers applied more than one ML algorithm. Peculiarly, the most applied ML algorithm is DT-based. However, this SLR does not differentiate between different DT-based algorithms (J48, Part, RF, etc...) in the analysis. The other widely used algorithms are SVM and FL algorithms. Some ML algorithms had a low rank in this SLR despite their popularity. It is the case of the similarity methods or regression algorithms. Thus, these algorithms are not being investigated enough, which opens opportunities for future studies in CR field, to fill this gap.

ML Algorithm	Number of studies	References
DT	13	[38, 52, 48, 47, 46, 61, 59, 50, 68, 60, 70, 67, 45]
FL	11	[32, 37, 41, 52, 63, 49, 57, 64, 71, 43, 65]
SVM	8	[46, 52, 54, 55, 59, 68, 60, 45]
ANN	6	[33, 37, 54, 59, 58, 45]
NBC	4	[38, 46, 52, 48]
Regression	4	[40, 60, 45, 69]
KNN	3	[38, 59, 47]
Ensemble	2	[38, 46]
KMC	1	[61]
PCS	1	[39]
COS	1	[49]
LVQ	1	[56]

Table 5: Number of articles by type of ML algorithm used.

##### 5. What are the main input features?

ML models are data-depending models, without a constitution of high- quality training data, even the most performant algorithms theoretically will not give the expected results. Indeed, robust ML models can be useless when they are trained on inadequate, inaccurate, or irrelevant data. In the same con- text a wide variety of inputs were suggested in the reviewed articles, Figure 5 shows the classification of these parameters in six categories, viz.:

**Geography:** This category of inputs indicates the agroclimatic regions, which is a land unit suitable for a certain range of crops and cultivars. Table 7 shows that 19 papers built their RS using geographic data among other variables which confirms the importance of this type of inputs, mainly because it works as an identifier that is unique to every farm.

**Weather conditions (WCs):** Weather plays a major role in determining the success of agricultural pursuits. For farmers, timing is critical in the obtainment of resources, such as: fertilizer and seed, but also forecasting likely weather in the upcoming season, informing on how much irrigation is needed, as well as temperature that can affect crop growth. These factors can be determined by recording hourly, daily, or weekly, temperature, rainfall, solar radiation, wind speed, evaporation, relative humidity, and evapotranspiration. In this SLR, WCs were used in 75% of the reviewed articles as Table 7 indicate.

**Soil propriety (SP):** All soils contain mineral particles, organic matter, water, and air. The combinations of these components determines the soil's quality, which depends both on its physical properties (texture, color, type, porosity, bulk density, etc.) and chemical properties (soil pH, soil salinity, nutrients availability, soil electrical conductivity etc.). Table 7 confirms that soil characteristics are the mandatory inputs on which researchers-built crop RSs.

## Soil physical properties:

- **Soil texture:** Refers to the size of the particles that make up the soil and depends on the proportion of sand, silt and clay-sized particles and organic matter in the soil, it can influence whether soils are free draining, whether they hold water and how easy it is for plant roots to grow.
- **Soil color:** The surface soil varies from almost white through shades of brown and grey to black. Light color indicates low organic matter content while dark color indicates a high organic matter content.
- **Soil type:** It describes the way the sand, silt and clay particles are clumped together. Organic matter (decaying plants and animals) and soil organisms like earthworms and bacteria influence soil structure. It is important for plant growth, regulating the movement of air and water, influencing root development and affecting nutrient availability.
- **Soil porosity:** It refers to the pores within the soil. Porosity influences the movement of air and water. Healthy soils have many pores between and within the aggregates. Poor quality soils have few visible pores, cracks, or holes.
- **Bulk density:** It is the proportion of the weight of a soil relative to its volume. Bulk density is an indicator of the amount of pore space available within individual soil horizons and it reflects the soil's ability to function for structural support, water and solute movement, and soil aeration.

## Soil chemical properties:

- **Soil pH:** Soil reactivity is expressed in terms of pH and is a measure of the acidity or alkalinity of the soil. More precisely, it is a measure of hydrogen ion concentration in an aqueous solution and ranges in soils from 3.5 (very acid) to 9.5 (very alkaline). The effect of pH is to remove from the soil or to make available certain ions.
- **Soil salinity:** It is the salt content in the soil; the process of increasing the salt content is known as salinization. Salts occur naturally within soils and water. Salination can be caused by natural processes such as mineral weathering or by the gradual withdrawal of an ocean.
- **Nutrients availability:** Sixteen nutrients are essential for plant growth and living organisms in the soil. These fall in two different categories namely macro- and micronutrients. The macronutrients include Carbon (C), Oxygen (O), Hydrogen (H), Nitrogen (N), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulphur (S) and are the most essential nutrients to plant development whereby a high quantity of these is needed. The micronutrients on the other hand are needed in smaller amounts, however they are still crucial for plant development and growth, these include Iron (Fe), Zinc (Zn), Manganese (Mn), Boron (B), Copper (Cu), Molybdenum (Mo) and Chlorine (Cl). Nearly all plant nutrients are taken up in ionic forms from the soil solution as cations or as anions.
- **Soil Electrical Conductivity (SEC):** It is an indirect measurement that correlates very well with several soil physical and chemical properties. Electrical conductivity is the ability of a material to conduct (transmit) an electrical current. As measuring soil electrical conductivity is easier, less expensive, and faster than other soil properties measurements, it can be used as a good tool for obtaining useful information about soil.
- **Crop propriety (CP):** Some crops are very labor-intensive. Some crops require more skill than others. Some crops are riskier than others (high profit if it's a good year but high chance of crop failure if the weather is bad), and some farmers are more able to cope with those risks. Each crop has its suitable amount needed of nutrients, optimal weather conditions and optimal soil properties. Unfortunately, there is no universal structure or data source for this kind of crop information, so researchers in different reviewed papers use data mining techniques to extract knowledge from raw data, where FL shows high quality results, because of its rules generating model.
- **CPR:** There are a lot of crop types produced in farms not all of them are suitable for producing in all areas. So, considering CPR of each one of them for every farm is very important to recommend and predict the crop productivity. Almost 90% of the reviewed papers are using supervised learning, where crop yield or crop profitability, in ton/hectare or kg/hectare, were used as the dependent variable.
- **Market:** Even with a high yield, decision about recommending the crop cannot be taken without knowing its price for the period of sell, as well as what is its cost. The price of a specific crop is determined through demand/supply in the market; however, it can be predicted using historical data. In the

other hand, cost can only be given by the farmer himself, ultimately it remains difficult to gather such data. Table 7 confirms this claim with just four papers using market information.

Feature class	Number of studies	References
WCs	20	[32, 33, 35, 36, 37, 40, 41, 42, 39, 50, 54, 57, 58, 64, 71, 43, 60, 67, 44, 45]
SP	19	[32, 35, 36, 37, 38, 40, 41, 42, 39, 50, 57, 69, 68, 71, 43, 65, 66, 60, 45]
Geography	17	[35, 33, 36, 37, 38, 40, 41, 42, 39, 50, 54, 69, 71, 65, 70, 67, 45]
CPR	9	[32, 35, 40, 39, 50, 64, 70, 44, 45]
CP	9	[35, 39, 50, 54, 57, 69, 64, 67, 44]
Market	4	[32, 35, 37, 40]

Table 7: Distribution of papers by feature classes.

Table 9 presents the number of papers for each variable; it indicates that from all the features cited above temperature and rainfall are the widely used parameters. This finding is coherent with the fact that WCs have an important impact on the CPR and determine the soil's sustainability, nevertheless, it remains necessary to extract other information to build an efficient CR. This information was grouped previously in the soil property category and they are pH-value, soil type and nutrient availability where they were cited respectively in [15], [11] and [10]. Less important variables occurred in a range of 1 to 6 and they are a mix of all categories of features such as Elevation for geography, salinity for soil characteristics and humidity in weather conditions. The number of articles included in this SLR, could give a relevant order of variable importance evolved in a CR algorithm which is statistically supported by the limited number of research papers in precision agriculture dedicated to CR.

Feature	Number of studies	References
Temperature	17	[46, 52, 61, 49, 54, 55, 51, 56, 47, 57, 58, 64, 43, 60, 70, 67, 45]
Rainfall	16	[46, 61, 52, 49, 54, 55, 51, 56, 47, 63, 64, 71, 60, 70, 67, 45]
pH-value	14	[46, 48, 51, 56, 59, 47, 63, 57, 69, 43, 65, 66, 60, 67]
Soil type	10	[46, 52, 48, 61, 54, 51, 47, 65, 67, 45]
Nutrients	9	[46, 61, 51, 59, 68, 43, 65, 66, 45]
Humidity	6	[49, 53, 58, 64, 43, 45]
Yield rate	5	[61, 49, 55, 64, 70]
EC soil	5	[48, 53, 63, 69, 66]
Salinity	5	[53, 59, 69, 43, 65]
Crop type	4	[54, 34, 35, 50]
Pressure	2	[53, 58]
Soil color	2	[63, 60]
Elevation	2	[71, 45]
Soil porosity	1	[46]

Table 9: Distribution of papers by features.

#### 6. Which evaluation parameters and evaluation approaches have been used?

Several evaluation metrics have been used. Table 11 gives information about metrics used for evaluation techniques in the reviewed studies. This SLR restricted to the CS, which make classification metrics such as: Accuracy, Precision or Recall, the most popular performance metrics used in the studies of this SLR.

Accuracy is the proportion of true results among the total number of cases examined, which has the highest number of occurrences in our SLR, more than precision, which refers to the fraction of relevant recommendation among the retrieved crops, and recall, which refers to the fraction of retrieved recommendation among all relevant crops.

Another important result is the existing of studies evaluating their results by using regression error metrics such as: RMSE, MSE and the mean absolute error (MAE) metrics. The reason is that a study may use CPR as an output of the model developed, then, it chooses the crop which has the highest rate.

What is striking in the Table 11 is the remarkable amount of papers that are not evaluated by any performance criteria. The most likely cause of this result is the difficulty to verify whether the recommended crop is truly the correct or this reason, most of studies seek the help of experts or farmers to judge the relevance of the suggested crops [50].

Metric	Number of studies	References
Accuracy	19	[46, 52, 48, 54, 56, 59, 47, 35, 36, 38, 42, 50, 71, 65, 60, 45]
Precision	10	[52, 48, 36, 37, 39, 58, 68, 65]
Recall	7	[48, 37, 39, 52, 58, 68, 65]
F-measure	5	[37, 52, 68, 65]
Sensitivity	2	[52, 36]
Specificity	2	[52, 36]
MSE	2	[33, 45]
RMSE	2	[48, 70]
MAE	1	[48]
No metric	11	[61, 49, 53, 55, 51, 63, 32, 34, 40, 41, 57, 64, 43, 67]

Table 11: Distribution of articles by performance metrics.

## 7. What are the current challenges in CR?

This section puts into terms challenges encountered in the extant research, these challenges are observed through three layers: the proposed algorithm, the used data and the evaluation preferences. Most studies have almost exclusively focused on the exploitation of CF algorithms for classification and clustering, more precisely: DTs, SVM, ANNs and NBC. Several gaps and shortcomings were identified in these techniques, namely the *cold start* problem, where the system cannot draw any inferences for users or items about which has not yet gathered sufficient information. A closer look to the literature, reveals that these proposed adaptations are very classical, since the field of RS has known a significant improvement due to entertainment companies, new algorithms were developed. For Instance, the Netflix Prize was an open competition for the best CF algorithm to predict user ratings for movies [72]. On September 21, 2009, the grand prize was given to the BellKor's Pragmatic Chaos team, which bested Netflix's own algorithm for predicting ratings by 10.06% [73]. During this competition MF became widely known due to its effectiveness, and important steps were taken in later years towards some very successful algorithms, which share the same basis of latent factor and user/item representations. Unfortunately, there was no article that suggested an implementation of these modern techniques for CR. A potential barrier that may face researchers in their quest to solve this delay is the available data, Figure 6 shows that more than 50% of the reviewed papers were published by Asian researchers, and more precisely Indian researchers, where SP parameters, CPR parameters and parameters for a very long period of time and by different states/districts, are available in government official websites. Notwithstanding, these data are inaccessible for foreign scientists. Furthermore, the proposed structure of data presents another challenge, where the vast majority of the well performed RS are fitted by a user/item rating matrix, and the user's demographic data for hybrid systems, where ratings are either explicit or implicit (number of clicks, number of page visits, the number of times a song was played, etc.) So different complex preprocessing techniques are required. Even though, the *sparsity* remains a potential challenge to deal with. Finally, the high correlation degree that was observed for performance metrics and evaluations of the model, based on anonymous experts raises a very important question about the reliability of these results, hence the proposed algorithm.

## 5. Discussion

During the last decade, the use of technology to enhance the agricultural processes has been very remarkable. For CR, we can see clearly from Table 4.2 that there has been some success in this direction. Exploration of the used techniques shows us that most of the times, the problem is formulated as a classification problem, where algorithms as DTs and NBC, etc. give remarkable results. Fuzzy systems have been used in other cases, to model the uncertainty

in input variables that were categorized and analyzed to facilitate the choice for future researchers. This study reveals also some challenges that are being faced when creating a CS method, like the unavailability of the data, more specifically benchmarking datasets to compare the models, the input variables are quite different from a study to another, and the difficulty in measuring the performance of the proposed methods; some papers compare their model predictions to what the farmer has actually cultivated, while others compare to domain experts recommendations. Historical evidence shows that great scientific achievements were guided by industrial needs, unfortunately, precision agriculture and CR haven't yet gained the attention that it deserves from different stakeholders, basically in emergent countries where agriculture is the most valuable resource. Major improvements in agricultural domain will certainly appear by integrating the successful algorithms of RS that were developed for entertainment companies, whom humankind might benefit differently. Finally, in this paper we've studied 40 well selected articles from different reliable sources. Nevertheless, this number remains statistically insignificant and more similar works are needed to illuminate the path for new researchers that are willing to innovate, achieve, and contribute to the field.

## 6. Conclusion And Future Work

In this SLR we have presented a detailed analysis of 40 articles published from 2010 to 2020 about the CR problem as well as the main achievements, and current challenges. The SLR was conducted with the aim of providing insights into the kind of solutions that were proposed in the recent years for the CS task. Such insights are valuable in suggesting new directions for research studies and in providing a good understanding of the recent research trends. For our feature work, we look forward to proposing new methods that are inspired from the current development of RS in other domains and tackling the present specific challenges in agriculture.

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

We confirm our consent for the publication of the present paper.

### Availability of data and materials

Not applicable.

### Competing interests

No, we declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

### Funding

This research was funded by AgriEdge of OCP foundation in the framework of the Digital Farming Project, between Mohammed VI Polytechnic University (UM6P), Benguerir, Kingdom of Morocco and Massachusetts Institute of Technology (MIT), Boston, USA.

### Author's contribution

The authors equally contributed.

### Acknowledgement

We acknowledge Agriedge (a company specialized in precision agriculture) for their financial support, as well as the Mohammed VI Polytechnic university (UM6P) for the material and administrative support. We also thank the reviewers for their valuable tips that to an improved version of this article.

## References

1. I. Portugal, P. Alencar, D. Cowan, The use of machine learning algorithms in recommender systems: A systematic review, *Expert Systems with Applications* 97 (2018) 205–227. doi:10.1016/j.eswa.2017.12.020.
2. K. A. Klompenburg T., C. C., Crop yield prediction using machine learning: A systematic literature review, *Computers and Electronics in Agriculture* 177 (January) (2020) 105709. doi:10.1016/j.compag.2020.105709.
3. A. Chlingaryan, S. Sukkariah, B. Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, *Computers and Electronics in Agriculture* 151 (2018) 61–69. doi:10.1016/j.compag.2018.05.012.
4. R. Burke, Hybrid recommender systems: Survey and experiments, *User Modeling and User-Adapted Interaction* 12 (2002) 331–370. doi:10.1023/A:1021240730564.
5. A. Poriya, T. Bhagat, N. Patel, R. Sharma, Non-personalized recommender systems and user-based collaborative recommender systems, *International Journal of Applied Information Systems* 6 (9) (2014) 22–27. doi:10.1.1.428.6731.
6. R. Burke, Knowledge-based recommender systems, <https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day6/burke-elis00.pdf> (2000).
7. E. Aïmeur, G. Brassard, J. M. Fernandez, F. S. M. Onana, Privacy-preserving demographic filtering, in: *Proceedings of the 2006 ACM Symposium on Applied Computing*, Association for Computing Machinery, New York, NY, USA, 2006, p. 872–878. doi:10.1145/1141277.1141479.
8. J. S. Breese, D. Heckerman, C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, in: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, Morgan Kaufmann Publishers Inc., 1998, p. 43–52. doi:10.5555/2074094.2074100.
9. J. Bobadilla, F. Ortega, A. Hernando, J. Alcalá, Knowledge-Based Systems Improving collaborative filtering recommender system results and performance using genetic algorithms, *Knowledge-Based Systems* 24 (8) (2011) 1310–1316. doi:10.1016/j.knosys.2011.06.005.

10. N. Tsapatsoulis, O. Georgiou, Investigating the scalability of algorithms, the role of similarity metric and the list of suggested items construction scheme in recommender systems, *Inter* 29. doi:10.1142/S0218213012400180.
11. H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, H. Shah, Wide amp; deep learning for recommender systems, in: *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS 2016*, Association for Computing Machinery, New York, NY, USA, 2016, p. 7–10. doi:10.1145/2988450.2988454.
12. H.-J. Xue, X. Dai, J. Zhang, S. Huang, J. Chen, Deep matrix factorization models for recommender systems, in: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, International Joint Conference on Artificial Intelligence, 2017, pp. 3203–3209. doi:10.24963/ijcai.2017/447.
13. X. Wang, X. He, M. Wang, F. Feng, T.-S. Chua, Neural graph collaborative filtering, in: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'19*, Association for Computing Machinery, 2019, p. 165–174. doi:10.1145/3331184.3331267.
14. K. R. P. Kumar, B. Bhasker, DNNRec: A novel deep learning-based hybrid recommender system, *Expert Systems with Applications* 144 (2020). doi:10.1016/j.eswa.2019.113054.
15. A. v. d. Oord, S. Dieleman, B. Schrauwen, Deep content-based mu- sic recommendation, in: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13*, Curran Associates Inc., Red Hook, NY, USA, 2013, p. 2643–2651. doi:10.5555/2999792.2999907.
16. D. Kim, C. Park, J. Oh, S. Lee, H. Yu, Convolutional matrix factorization for document context-aware recommendation, in: *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16*, Association for Computing Machinery, New York, NY, USA, 2016, p. 233–240. doi:10.1145/2959100.2959165.
17. G. Srivastav, S. Kant, Review on e-Learning Environment Development and context aware recommendation systems using Deep Learning, *International Conference on Recent Developments in Control, Automation and Power Engineering, RDCAPE (2019)*. doi:10.1109/RDCAPE47089.2019.8979066.
18. S. Sedhain, A. K. Menon, S. Sanner, L. Xie, Autorec: Autoencoders meet collaborative filtering, in: *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, Association for Computing Machinery, New York, NY, USA, 2015, p. 111–112. doi:10.1145/2740908.2742726.
19. D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, Variational au- toencoders for collaborative filtering, in: *Proceedings of the 2018 World Wide Web Conference, WWW '18*, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018, p. 689–698. doi:10.1145/3178876.3186150.
20. L. Li, D. Wang, T. Li, D. Knox, B. Padmanabhan, Scene: A scalable two- stage personalized news recommendation system, in: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, New York, NY, USA, 2011, p. 125–134. doi:10.1145/2009916.2009937.
21. S. Purushotham, Y. Liu, C.-C. J. Kuo, Collaborative topic regression with social matrix factorization for recommendation systems, in: *Proceedings of the 29th International Conference on International Conference on Machine Learning*, Omnipress, Madison, WI, USA, 2012, p. 691–698. doi:10.5555/3042573.3042664.
22. L. Li, W. Chu, J. Langford, R. E. Schapire, A contextual-bandit approach to personalized news article recommendation, *Proceedings of the 19<sup>th</sup> International Conference on World Wide Web (2010)*. doi:10.1145/1772690.1772758.
23. L. Song, C. Tekin, M. van der Schaar, Online learning in large-scale contextual recommender systems, *IEEE Transactions on Services Computing* 9 (3) (2016) 433–445. doi:10.1109/TSC.2014.2365795.
24. G. Zheng, F. Zhang, Z. Zheng, Y. Xiang, N. J. Yuan, X. Xie, Z. Li, Dm: A deep reinforcement learning framework for news recommendation, in: *Proceedings of the World Wide Web Conference, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018*, p. 167–176. doi:10.1145/3178876.3185994.
25. L. M. De Campos, J. M. Fernández-Luna, J. F. Huete, M. A. Rueda- Morales, combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks, *International Journal of Approximate Reasoning* 51 (7) (2010) 785–799. doi:10.1016/j.ijar.2010.04.001.
26. V. Kant, K. K. Bharadwaj, Enhancing Recommendation quality of content-based filtering through collaborative predictions and fuzzy similarity measures, *Procedia Engineering* 38 (2012) 939–944. doi:10.1016/j.proeng.2012.06.118.
27. J. Lacasta, F. J. Lopez-Pellicer, B. Espejo-García, J. Nogueras-Iso, F. J. Zarazaga-Soria, Agricultural recommendation system for crop protection, *Computers and Electronics in Agriculture* 152 (June) (2018) 82–89. doi:10.1016/j.compag.2018.06.049.
28. M. A. Salam, M. A. Mahmood, Y. M. Awad, M. Hazman, N. El Bendary, A. E. Hassanien, M. F. Tolba, S. M. Saleh, Climate Recommender System for Wheat Cultivation in North Egyptian Sinai Peninsula, in: *Advances in Intelligent Systems and Computing*, Vol. 303, Springer Verlag, 2014, pp. 121–130. doi:10.1007/978-3-319-08156-413.
29. A. Iorshase, O. I. Charles, A Well-Built Hybrid Recommender System for Agricultural Products in Benue State of Nigeria, *Journal of Software Engineering and Applications* 08 (11) (2015) 581–589. doi:10.4236/jsea.2015.811055.
30. S. Jaiswal, T. Kharade, N. Kotambe, S. Shinde, Collaborative Recommendation System for Agriculture Sector, *ITM Web of Conferences* 32 (2020) 03034. doi:10.1051/itmconf/20203203034.
31. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, D. Moher, Updating guidance for reporting systematic reviews: development of the prisma 2020 statement, *Journal of Clinical Epidemiology* 134 (2021) 103–112. doi:10.1016/j.jclinepi.2021.02.003.



32. B. M. Elomda, H. A. Hefny, F. Ashmawy, A Multi-Level Linguistic Fuzzy Decision Network, in: *Advances in Intelligent Systems and Computing*, Springer, Cham, 2015. doi:10.1007/978-3-319-11310-4.
33. F. Farnood Ahmadi, N. Farsad Layegh, Integration of artificial neural network and geographical information system for intelligent assessment of land suitability, *Geographical Modelling* 014-1801-z.
34. R. Kumar, M. P. Singh, P. Kumar, J. P. Singh, Crop Selection Method to maximize crop yield rate using machine learning technique (2015). doi:10.1109/ICSTM.2015.7225403.
35. A. Kawtrakul, R. Amorntarant, H. Chanlekha, Development of an expert system for personalized crop planning, in: 7th International ACM Conference on Management of Computational and Collective Intelligence in Digital EcoSystems, MEDES 2015, Association for Computing Machinery, 2015, pp. 250–257. doi:10.1145/2857218.2857272.
36. N. Deepa, K. Ganesan, Multi-class classification using hybrid soft decision model for agriculture crop selection, *Neural Computing and Applications* 30 (4) (2016) 1025–1038. doi:10.1007/s00521-016-2749-y.
37. A. U. A. S. B. L. N. R. Sridhar, Fuzzy Logic Based Hybrid Recommender of Maximum Yield Crop Using Soil, Weather and Cost, *ICTACT Journal on Soft Computing* 6 (4) (2016) 1261–1269. doi:10.21917/ijsc.2016.0173.
38. S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, J. Nisha, Crop recommendation system for precision agriculture, *International Conference on Advanced Computing* (2017). doi:10.1109/ICoAC.2017.7951740.
39. M. J. Mokarrama, M. S. Arefin, RSF: A recommendation system for farmers, 2018, pp. 843–850. doi:10.1109/R10-HTC.2017.8289086.
40. R. E. V. Raja, S. Kanaga SubaRishi, Demand based crop recommender system for farmers, in: *Proceedings -2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development*, pp. 194–199. doi:10.1109/TIAR.2017.8273714.
41. A. Kapoor, A. K. Verma, Crop Selection Using Fuzzy Logic-Based Expert System, *Applications of Soft Computing for the Web* (2017). doi:10.1007/978-981-10-7098-3 8.
42. N. Deepa, K. Ganesan, Decision-making tool for crop selection for agriculture development, *Neural Computing and Applications* 31 (4) (2017) 1215–1225. doi:10.1007/s00521-017-3154-x.
43. R. Joshi, H. Fadewar, P. Bhalchandra, Fuzzy based intelligent system to predict most suitable crop, in: *Proceedings of the International Conference on Computing and Communications* 383. doi:10.2991/icccsp-16.2017.58.
44. C. Filippi, R. Mansini, E. Stevanato, Mixed integer linear programming models for optimal crop selection, *Computers and Operations Research* 81 (2017) 2639. doi:10.1016/j.cor.2016.12.004.
45. T. Islam, T. A. Chisty, A. Chakrabarty, A deep neural network approach for crop selection and yield prediction in Bangladesh, *IEEE Region 10 Humanitarian Technology Conference* (2019). doi:10.1109/R10-HTC.2018.8629828.
46. N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, N. K. Cauvery, Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique, *Proceedings 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions* (2018). doi:10.1109/CSITSS.2018.8768790.
47. Z. Doshi, S. Nadkarni, R. Agrawal, N. Shah, Agro-consultant: Intelligent crop recommendation system using machine learning algorithms, in: *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE, 2018, pp. 1–6. doi:10.1109/ICCUBEA.2018.8697349.
48. A. Arooj, M. Riaz, M. N. Akram, Evaluation of predictive data mining algorithms in soil data classification for optimized crop recommendation, *International Conference on Advancements in Computational Sciences* (2018). doi:10.1109/ICACS.2018.8697349.
49. M. Kuanr, B. Kesari Rath, S. Nandan Mohanty, Crop Recommender System for the Farmers using Mamdani Fuzzy Inference Model, *International Journal of Engineering & Technology* 7 (4.15) (2018). doi:10.14419/ijet.v7i4.15.23006.
50. M. B. Anley, T. B. Tesema, A Collaborative Approach to Build a KBS for Crop Selection: Combining Experts Knowledge and Machine Learning Knowledge Discovery, in: *Communications in Computer and Information Science*, Vol. 1026, Springer Verlag, 2019, pp. 80–92. doi:10.1007/978-3-030-26630-1 8.
51. S. Poongodi, M. Rajesh Babu, Analysis of crop suitability using clustering technique in Coimbatore region of Tamil Nadu, *Concurrency Computation* 31 (14) (2019) 1–13. doi:10.1002/cpe.5294.
52. N. Deepa, K. Ganesan, Hybrid rough fuzzy soft classifier based multi-class classification model for agriculture crop selection, *Soft Computing* 23 (21) (2019) 10793–10809. doi:10.1007/s00500-018-3633-8.
53. F.-H. Tseng, H.-H. Cho, H.-T. Wu, Applying Big Data for Intelligent Agriculture-Based Crop Selection Analysis, *IEEE Access* 7 (2019) 116965–116974. doi:10.1109/access.2019.2935564.
54. T. K. Fegade, B. V. Pawar, Network and Support Vector Machine (2020). doi:10.1007/978-981-13-9364-8 23.
55. Meeradevi, H. Salpekar, Design and Implementation of Mobile Application for Crop Yield Prediction using Machine Learning, in: *2019 Global Conference for Advancement in Technology (GCAT)*, IEEE, pp. 1–6. doi:10.1109/GCAT47503.2019.8978315.
56. T. Rizaldi, H. A. Putranto, H. Y. Riskiawan, D. P. S. Setyohadi, J. Riaviandy, Decision Support System for Land Selection to Increase Crops Productivity in Jember Regency Use Learning Vector Quantization (LVQ), *Proceedings - 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering* 1 82–85. doi:10.1109/ICOMITEE.2019.8921033.
57. C. O. Martinez-Ojeda, T. M. Amado, J. C. Dela Cruz, In Field Proximal Soil Sensing for Real Time Crop Recommendation Using Fuzzy Logic Model, in: *International Symposium on Multimedia and Communication Technology (IS-MAC)*, IEEE, 2019. doi:10.1109/ISMAC.2019.8836160.

58. S. B. Kamatchi, R. Parvathi, Improvement of Crop Production Using Recommender System by Weather Forecasts, *Procedia Computer Science* 165 (2019) 724–732. doi:10.1016/j.procs.2020.01.023.
59. S. A. Z. Rahman, K. C. Mitra, S. M. Islam. Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series, 21st International Conference of Computer and Information Technology (2019). doi:10.1109/ICCITECHN.2018.8631943.
60. A. Kumar, S. Sarkar, C. Pradhan, Recommendation system for crop identification and pest control technique in agriculture, in: *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing*, pp. 185–189. doi:10.1109/ICCSP.2019.8698099.
61. V. K. A. Chougule, D. Mukhopadhyay, Crop suitability and fertilizers recommendation using data mining techniques, in: *Advances in Intelligent Systems and Computing*, Vol. 714, Springer Verlag, 2019, pp. 205–213. doi:10.1007/978-981-13-0224-4 19.
62. B. Viviliya, V. Vaidhehi, The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms, *International Journal of Innovative Technology and Exploring Engineering* 9 (2) (2019) 430–431. doi:10.35940/ijitee.b7219.129219.
63. R. Aarthi, D. Sivakumar, Modeling the Hierarchical Fuzzy System for Suitable Crop Recommendation, in: *Lecture Notes in Electrical Engineering*, Vol. 686, Springer Science and Business Media Deutschland GmbH, 2020, pp. 199–209. doi:10.1007/978-981-15-7031-5 19.
64. R. M.-E. M. Cadenas, M. Carmen, Development of an application to make knowledge available to the farmer: Detection of the most suitable crops for a more sustainable agriculture, *Journal of Ambient Intelligence and Smart Environments* 12 (5) (2020) 419–432. doi:10.3233/AIS-200575.
65. A. M. Rajeswari, A. S. Anushiya, K. S. A. Fathima, S. S. Priya, N. Mathumithaa. Fuzzy Decision Support System for Recommendation of Crop Cultivation based on Soil Type, *Proceedings of the 4th International Conference on Trends in Electronics and Informatics* (2020). doi:10.1109/ICOEI48184.2020.9142899.
66. K. Patel, H. B. Patel, A state-of-the-art survey on recommendation system and prospective extensions, *Computers and Electronics in Agriculture* 178 (2020). doi:10.1016/j.compag.2020.105779.
67. S. Jain, D. Ramesh, Machine Learning convergence for weather-based crop selection, *IEEE International Students' Conference on Electrical, Electronics and Computer Science* (2020). doi:10.1109/SCEECS48394.2020.75.
68. A. Liu, T. Lu, B. Wang, C. Chen, Crop Recommendation via Clustering Center Optimized Algorithm for Imbalanced Soil Data, in: *5th International Conference on Control, Robotics and Cybernetics (CRC)*, IEEE, 2020, pp. 31–35. doi:10.1109/CRC51253.2020.9253457.
69. F. Sari, F. Koyuncu, Multi Criteria Decision Analysis to Determine the Suitability of Agricultural Crops for Land Consolidation Areas, *International Journal of Engineering and Geosciences* 6 (2) (2021) 64–73. doi:10.26833/ijeg.683754.
70. A. Karwande, M. Wyawahare, T. Kolhe, S. Kamble, R. Magar, L. Maheshwari, Prediction of the Most Productive Crop in a Geographical Area Using Machine Learning, *Lecture Notes in Networks and Systems* 141 (2021) 433–441. doi:10.1007/978-981-15-7106-0 43.
71. G. Banerjee, U. Sarkar, I. Ghosh, *A Fuzzy Logic-Based Crop Recommendation System*, Springer Singapore, 2021. doi:10.1007/978-981-15-7834-2 6.
72. J. Bennett, S. Lanning, The netflix prize, <http://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf> (Aug. 2007).
73. Y. Koren, The bellkor solution to the netflix grand prize, [https://netflixprize.com/assets/GrandPrize2009\\_BPC\\_BellKor.pdf](https://netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf) (2009).

## Tables

Tables 2, 4, 6, 8, and 10 are not available with this version.

## Figures

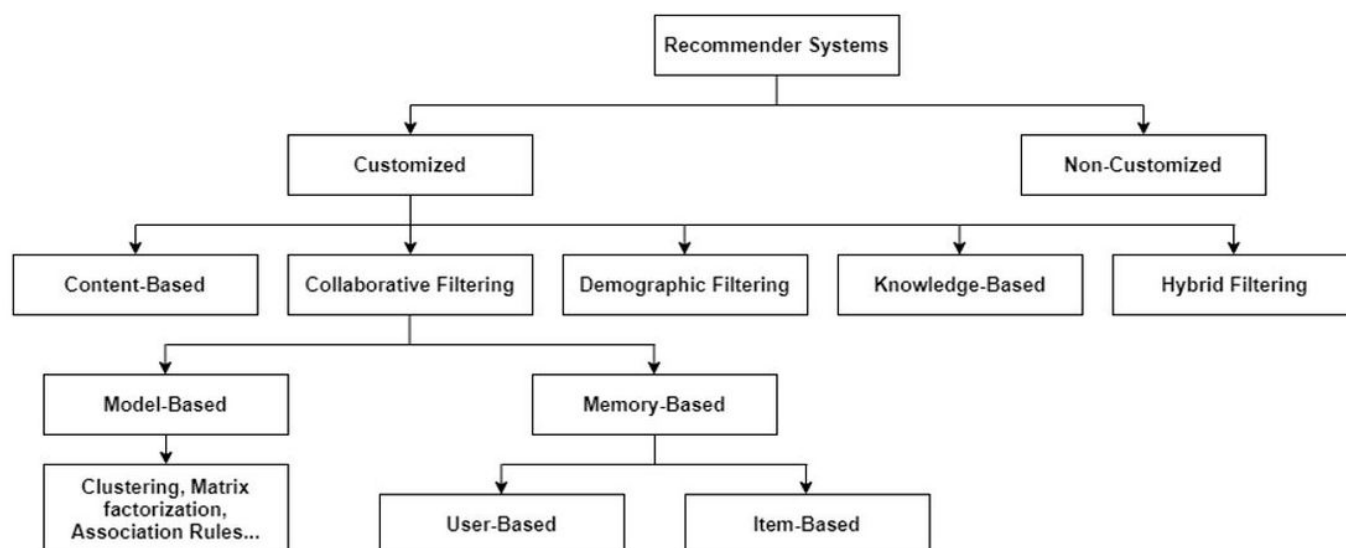


Figure 1

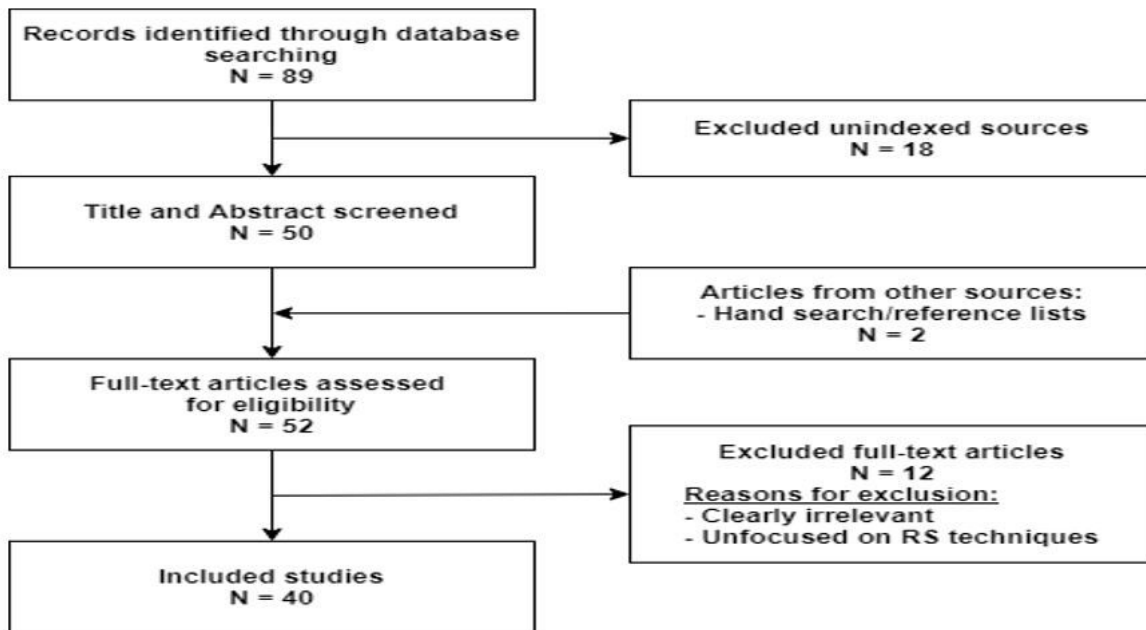
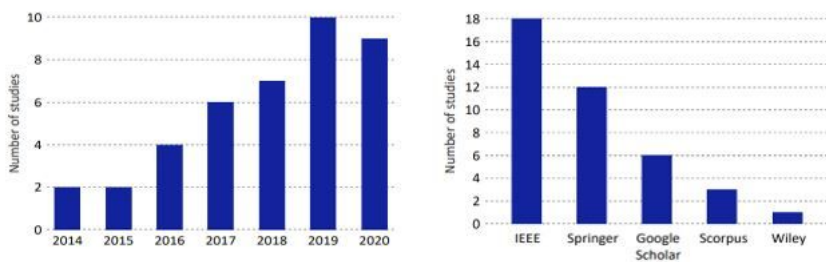
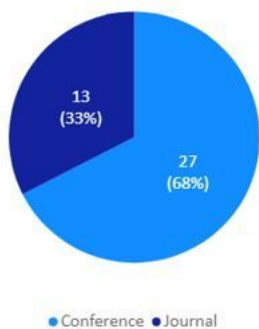


Figure 2

Flow chart of identification process of relevant CR articles.



(a) Distribution of selected papers since 2014 (b) Distribution of papers based on selected database



(c) Distribution by type of publication. Figure 3: Distribution of selected papers

Figure 3

Distribution of selected papers

Figure 4

Timeline of articles by recommendation approaches.

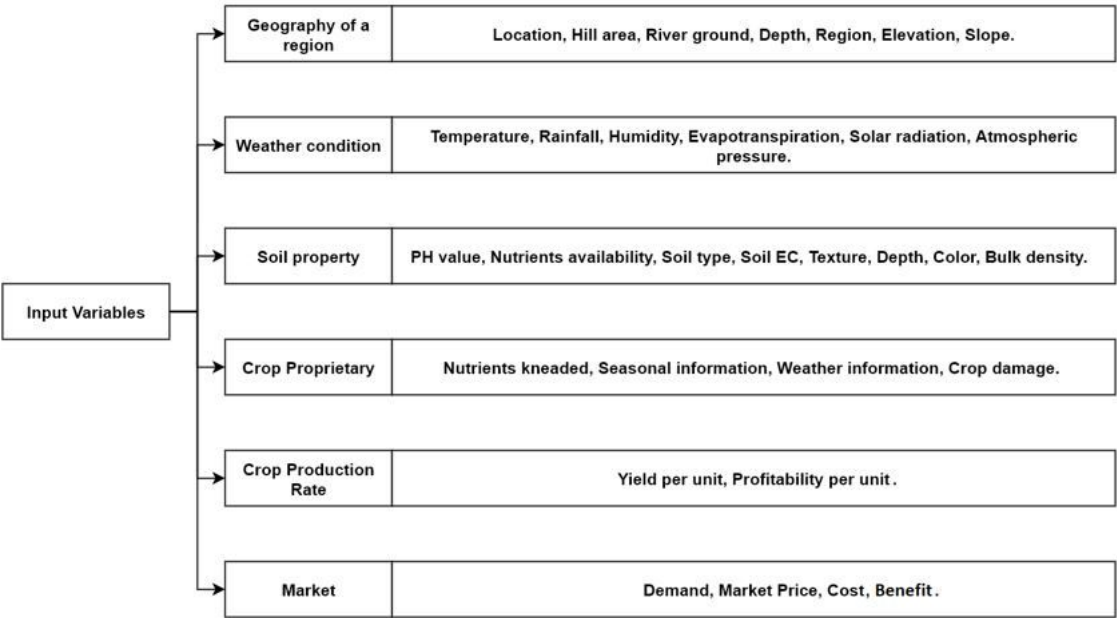


Figure 5

Input variables categories used for CR.

Figure 6

Articles distribution by country.