Semantics Analysis of Agricultural Experts' Opinions for Crop Productivity through Machine Learning

Mehak Rehman, Abdul Razzaq, Irfan Ahmad Baig, Javeria Jabeen, Muhammad Hammad Nadeem Tahir, Umar Ijaz Ahmed, Adnan Altaf & Touqeer Abbas

Published online: 14 Dec 2021.

Submit your article to this journal ⤤

Article views: 1698

View related articles ⤤

View Crossmark data ⤤

Citing articles: 1 View citing articles ⤤

Semantics Analysis of Agricultural Experts' Opinions for Crop Productivity through Machine Learning

Mehak Rehman[a], Abdul Razzaq[a], Irfan Ahmad Baig[b], Javeria Jabeen[a], Muhammad Hammad Nadeem Tahir[c], Umar Ijaz Ahmed[b], Adnan Altaf[a], and Touqeer Abbas[a]

[a]Department of Computer Science, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan; [b]Department of Agribusiness and Applied Economics, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan; [c]Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan

ABSTRACT

Semantic analysis is a particular technique, which is an interesting area of research that associates with Natural Language Processing (NLP), artificial intelligence, opinion mining, text clustering, and classification. Numerous text processing techniques are being used to find out sentiments from the comments, such as social media tweets, hoax, fiction, nonfiction, novels, books, movies, health care, and stock exchange. Agrarian experts' opinions play a vital role in the agriculture sector that yields good crop productivity. This paper presents a descriptive analysis of agriculture experts' opinions through machine learning methods based on textual data collection. The data has been collected by surveying various academia, research institute, and industry of Punjab, Pakistan. The impact of various agricultural inputs such as seed quality, soil quality, soil-intensive tillage, climate changes, water shortage, synthetic fertilizer, and precision technologies on crop productivity have been collected through questionnaires. This research provides a descriptive analysis of collected agrarians experts opinions to increase the crop yield by providing awareness regarding current agriculture inputs to farmers by using machine learning. The current research provides a cohesive expert guideline for improving crop productivity, useful for agricultural policymaking, and conveys adequate farmers' knowledge. Consequently, the proposed method is an innovative way of discovering recommendations of agrarians through sentiment analysis in survey data using machine learning methods. Furthermore, to the best of our knowledge, agrarians experts opinions on enhancing crop productivity have been considered for the first time in Pakistan.

Introduction

Semantics extraction with unique degrees of the analyzed texts, consisting of phrases, sentences, and documents. Recently, researchers have concentrated on semantics which is the interaction between human

CONTACT Abdul Razzaq ✉ abdul.razzaq@mnsuam.edu.pk 🔵 Department of Computer Science, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan

language (Chowdhury 2003). Currently, a big collection of documents has been sent and saved electronically; however, there is a need to preprocess text to extract meaningful information. Semantics can define how different words have altered meanings for other people (Wolf 1991). Semantics analysis applications accomplished on social media such as professional networking services (LinkedIn), social networking sites (Facebook), media sharing networks (Instagram, Pinterest, YouTube), social blogging networks (Tumblr, Medium), discussion networks (Reddit, Quora,), and review networks (Yelp, Glassdoor). These enables people from all over the world to post and share images, videos, audios, and professional profile information through LinkedIn (Bontcheva and Rout 2014). Specifically, scholars performed sentiments analysis on several social media datasets, particularly healthcare datasets, Facebook comments, movies datasets, and tweets (Saif et al. 2013). Over the last decade, with the explosion of work for exploring various aspects of sentiment analysis: detecting subjective and objective sentences; classifying sentences as positive, negative, or neutral; detecting the person expressing the sentiment and the target of the sentiment analysis; detecting emotions such as joy, fear, and anger in the text. Surveys by (Liu and Zhang 2012) give a summary of various of these approaches. In today's living world contexts, documents are stored electronically in every domain, as text data are increased highly in industry, business, technology, and the agriculture sector. Agriculture has become innovative by using IOT (Farooq et al. 2020), Cloud Computing (Mekala and Viswanathan 2017), Artificial Intelligence (Smith 2018), Machine Learning (Benos et al. 2021), Deep Learning, and Data Science (Angiani et al. 2016). Generally, agricultural productivity depends on some essential factors like fertilizer, seed, soil, water, and climate change (Ahmad and Heng 2012). In Pakistan, Punjab is the main agriculture zone of major and minor crops that contributes 18.9% of GDP and 42.3% of the labor force (Elahi et al. 2020). In the agriculture field only, we have had a large amount of text data through diverse platforms: Tweeter, Facebook, and LinkedIn groups (Martini et al. 2011). The scientists used semantics analysis on agriculture datasets to judge similarities, sentiments, emotions, feelings, and thoughts regarding crop productivity.

Various techniques have been used for selecting and extracting features from text such as Frequency Features Term Frequency Inverse Document Frequency (TF-IDF), Count Vector, N-grams (Uni-Gram, Bi-Gram, Tri-Gram), and Bag of Words (BOW) (Mirończuk and Protasiewicz 2018). The Term Frequency and Inverse Document Frequency (TFIDF) have been widely used method for features extraction (Abualigah et al. 2017). Machine Learning has been currently

implemented on soil types for agriculture crops productivity and management system (Saikai, Patel, and Mitchell 2020; Dongare et al. 2020). The deep learning approach is also covering many problems related to agriculture (Hoang et al. 2013), bioinformatics, and computational biology of plants, (Muharam et al. 2021). Many of the scholars have presented different ICT applications in agriculture in remote sensing, ecosystem service, crop yield forecasting, land monitoring climate change, and online demand of agricultural products (Abd-Elmabod et al. 2020; Weiss, Jacob, and Duveiller 2020; Kantasa-ard et al. 2020). Independently, scientists are applying machine learning techniques in agricultural input for measuring their effects (Benos et al. 2021).

Semantics analysis has been used for the management of crop, soil, and water in the agriculture domain (Karthikeyan, et al., 2020). Many Agricultural applications like digital agriculture (Jayaraman et al. 2015) follow IOT infrastructure, which relates to the crop management system (Prathibha, Hongal, and Jyothi 2017). The focus of the proposed study was to apply a semantics analysis on agrarian opinion and providing their recommendations/guidelines for farmers that play a valuable role in crop growth and management. Therefore, an analysis of agricultural experts' opinions toward crop productivity is present in this study. Major Contributions of the proposed work are:

- Collection of descriptive opinions of the agricultural experts through questionnaires
- Analyze the descriptive opinions of the agricultural experts through machine learning techniques.
- Determine the significant factors that affect agricultural productivity through opinion mining that are helpful to farmers and policymakers.

The rest of the paper is arranged as: section 2 describes the proposed methodology and related materials. Section 3 illustrates the experiment results and discussion. The conclusion is drawn in section 4.

Materials and Methods

In this research text opinions, regarding agriculture productivity were collected from agrarians experts. After preprocessing, feature extraction techniques such as N-gram, BOW, TF, and TF-IDF were applied on corpus for informative features. KNN and Naïve Bayes algorithms were selected for training the model. In the end, a comparison between agrarians' responses was carried out using the cosine similarity. Figure 1 shows the flow of the system for the proposed study.
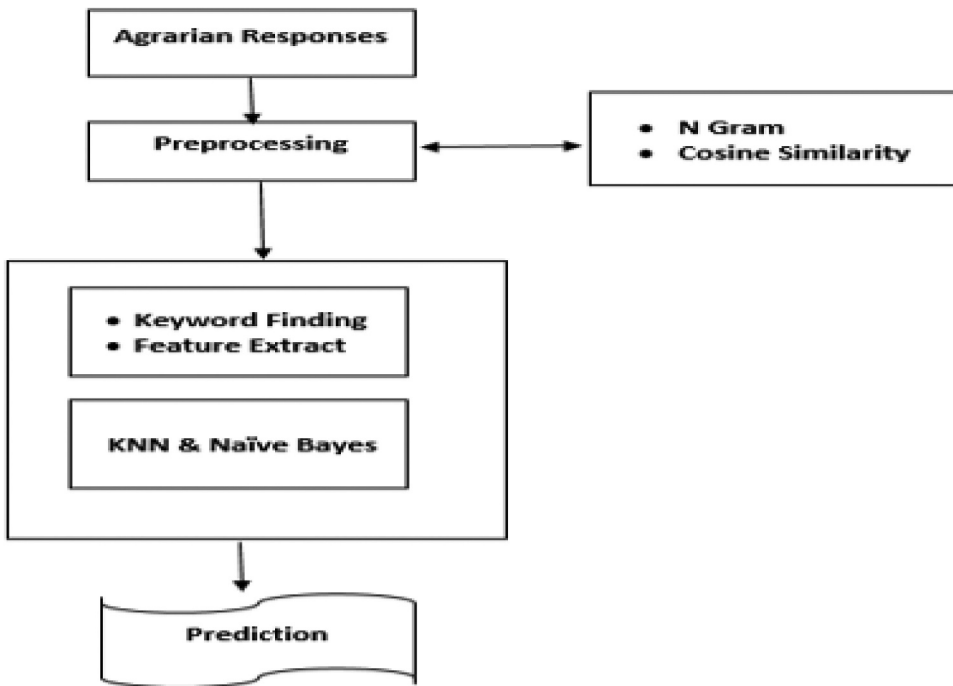
Figure 1. Flow of the system.

Data Collection

A questionnaire consisting of 10 questions was designed with the help of agricultural and social science experts regarding agriculture productivity: soil, seed, pest, insects, water shortage, climate change, precision agriculture, synthetic fertilizer, post-harvest, and government policies. Google Form and Word Document were created and shared with respondents using social media. Respondents were agrarian experts from different DAI/institutes, Academia, Industry, and Research institutes. The experts were specialists in Agronomy, Soil, and Environmental Sciences, Entomology, Plant Breeding and Genetics, Biotechnology, Seed Science and Technology, Forestry, Climate Change and Horticulture, Plant Pathology, Cotton, Crop Nutrients Tillage, Development of Transgenic Crops, Water Treatment, and Oxidation, Plant Growth Promoting Rhizobacteria, Plant Physiology, Tissue Culture, and Biochemistry. Proformas were disseminated at UAF (University of Agriculture Faisalabad), Bahauddin Zakariya University Multan (BZU), University of Gujrat (UAG), MNS University of Agriculture, Multan (MNSUAM), Cotton Center Research Institute Multan (CCRI), and National Institute for Biotechnology Genetic Engineering (NIBGE). Approximately one hundred responses were collected from agricultural experts, and the same is

Table 1. Questions regarding data collection.

Sr.	Questions
1	How does soil health affect crop productivity?
2	What are the impacts of certified and uncertified seed in crop productivity?
3	How can insect pest problem be minimized to increase Agriculture Productivity without harming the Environment?
4	What are impact of plant disease and insects/pest on Crop Productivity?
5	How can we resolve water shortage issues and challenges in Agriculture?
6	How does climate change affect agriculture and farmers?
7	How can precision agriculture be helpful in improving Agriculture and Crop Productivity?
8	What are pros and cons of synthetic fertilizer in Agriculture Sector?
9	How can we improve the post-harvest losses of Crops/Fruits/Vegetables?
10	How Government policies affect in sustainable Agriculture Growth?

being continued for better results from the trained model. Table 1 shows the list of questions that were prepared for data collection from agricultural experts.

Preprocessing

Preprocessing is the process of scrubbing and preparing the text for classification. The text consists of implicit noise that needs to be removed using data cleaning techniques. In the present research, preprocessing techniques such as tokenization of words, stop words removal, stemming, lemmatization, and a bag of words (Singh and Kumari 2016) were applied. Tokenization was used to convert text into chunks. It is necessary to remove such words from the corpus; therefore, stop words have been used that are 'a,' 'an, 'the,' 'have,' 'has,' 'from,' 'we,' 'will,' 'they,' 'them,' and much more. Similar stemming, also called lemmatization, has been used. Lemmatization removed the suffix of a word entirely and obtained the basic word form (lemma) (Kowsari et al. 2019). Count vector defined by several occurrences of features a basic way to represent the text data numerically called one-hot encoding (count vectorization) (Vaghela, Jadav, and Scholar 2016). Word cloud is also called text cloud/tags clouds, generated from the source of textual data in which words are depicted in different sizes. Word clouds are an alternative way of analyzing text from online surveys and documents, which is much faster than coding Essentially, word clouds generators work by breaking the text down into component words and counting how frequently they appear in the context-based documents, as shown in Figure 2.

Figure 2. Word cloud.

Features Extraction

Text documents aim to select features from text to determine the most informative features that contain high dimensional informative (Abualigah et al. 2017). Researchers used divergent approaches to extract features from the corpus (Mirończuk and Protasiewicz 2018), Continuous Words (CBOW) with Skip-Grams, Term Frequency or Inverse Document Frequency (TF-IDF), Features Frequency (FF), N-grams, Word frequency and weight calculation intensity words, negation words, and overall sentence weight calculation to find poultry of words with negative or positive weights (Razzaq et al. 2019). Many methods exist that can be chosen according to the dataset requirements (Mirończuk and Protasiewicz 2018). In this current study, we have selected Bag of words (BOW), Term Frequency, N-Gram, TF-IDF for feature selection. BOW applies on the text document, and N-gram (Uni, Bi, and Tri-Grams) approach was used to find the impact of those words that repeat mostly with higher frequency from "Agrarian Experts."

Term Frequency Inverse Document Frequency, (TF-IDF), is a standard weight scheme, the primary aim of this weighting scheme is to find the most informative feature where it represents the intrinsic content of the document (Abualigah et al. 2017).

$$Wi.j = tf \times \log N/dfi \qquad (1)$$

TF = Number of Occurance of i and j, DF = Number of Documents containing I, N = Total number of Documentsz

TF-IDF technique has been used in the study,for finding features with high frequency, which is the mostly used method to a small dataset with the specific content-based domain with BOW.

Classification Algorithm

We used Naïve Bays (NB) and K Nearest Neighbors (KNN) models for classification. KNN is instance-based learning, also called lazy algorithm, but it is a versatile algorithm used for text classification (Soucy and Mineau 2001) and regression. KNN is a feature-dependent algorithm. Lim (boundary) proposed methodology improves KNN performance based on text classification using well-estimated parameters (K-value). This study chose K = 3 for prediction and reduced the trained model's error.

$$d(x, y) = \sum_{i=1}^{n} (xi - yi) \tag{2}$$

where; x = Total Dataset; y = Total no of Labels e.g x denote the total dataset such as questions regarding agriculture range is $x_i = (x_{i1}\ x_{i2}\ x_{i3}\ \ldots\ \ldots\ x_{in})$ and y denote the total number of labels like agrarians experts views range is $y_i = (y_{i1}\ y_{i2}\ y_{i3}\ \ldots\ldots\ldots\ y_{in})$.

Naïve Bayes is another simple classifier based on bayesian probability, assuming that strong independence exists between features probabilities (Hutto and Gilbert, 2014). One of the advantages of the NB classifier is that it requires a small amount of training data to calculate the parameters for prediction, that's why we have selected this approach for this dataset. (Tripathy, Agrawal, and Rath 2015). NB is a popular classifier for opinion mining, semantics, and sentiments studies. In the past, many scholars have used these two algorithms in their methodology due to their effectiveness and simplicity (Ikonomakis, Kotsiantis, and Tampakas 2005).

$$P(C|x) = \frac{P(X|C)P(C)}{P(X)} \tag{3}$$

Where $P(C|x)$ represent posterior probability and $P(X)$ predictor prior probability. We chosed Naive Bayes due to small dataset and KNN was selected because it performed better on semantics-based text classifications studies. (Vaghela, Jadav, and Scholar 2016).

Cosine Similarity

Cosine Similarity has been used for document comparisons based on counting the maximum number of common words in the document.

$$Sim(A, B) = COS(\theta) = \frac{A.B}{||A||||B||} \tag{4}$$

In previous studies, the consine similarites has also been used in document comparsion on agriculture datasets (Prajapati and Kathiriya 2016). The cosine similarity method has been appleid for finding similarites between doucments using Equation (4) According to Equation (4), A and B are two matrices.

$$A = \begin{bmatrix} 0.00814302 & 0. & 0. & \ldots 0. & 0. & 0 \\ 0.00644564 & 0.01657335 & 0. & \ldots 0. & 0. & 0 \\ 0. & 0. & 0. & \ldots 0. & 0.0160478 & 0 \\ & & & \ldots & & \\ 0. & 0. & 0.03093901 & \ldots 0. & 0154695 & 0. & 0 \\ 0. & 0. & 0. & \ldots 0. & 0. & 0 \\ 0.00845222 & 0. & 0. & \ldots 0. & 0.03259914 & 0.01278 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.00814302 & 0. & 0. & \ldots 0. & 0. & 0 \\ 0.00644564 & 0.01657335 & 0. & \ldots 0. & 0. & 0 \\ 0. & 0. & 0. & \ldots 0. & 0.0160478 & 0 \\ \ldots & & & & & \\ 0. & 0. & 0.03093901 & \ldots 0. & 0154695 & 0. & 0 \\ 0. & 0. & 0. & \ldots 0. & 0. & 0 \\ 0.00845222 & 0. & 0. & \ldots 0. & 0.03259914 & 0.0127826 \end{bmatrix}$$

We find similarities between agrarian's feedback associates from different Institute/DAI, Academia, and Research Center in Punjab, Pakistan.

Figure 3 illustrates a comparison of documents that show documents similarities based on agrarian views. Ten documents are Soil, Seed Type, Pest, Plant, Diseases, Irrigation, Climate, Precision, Fertilizer, Harvest, and Policy are considered. Each document is compared with all other documents, e.g., the first document, Soil (Figure 3a), and 9 documents. This shows that Soil document is similar to document No 2 that is **Seed and** document No 9 **Policy** (Figure 3a) with the same polarity ratio **of 0.15**. The second document is Seed (Figure 3b) is compared with all 9 documents. The seed document is similar to document No 3, Pest, and document No 5 Irrigation (Figure 3b) with the same polarity ratio **of 0.15**. Similarly, ratios are compared in documents number C, D, E, F, G, H, I, and J, respectively.
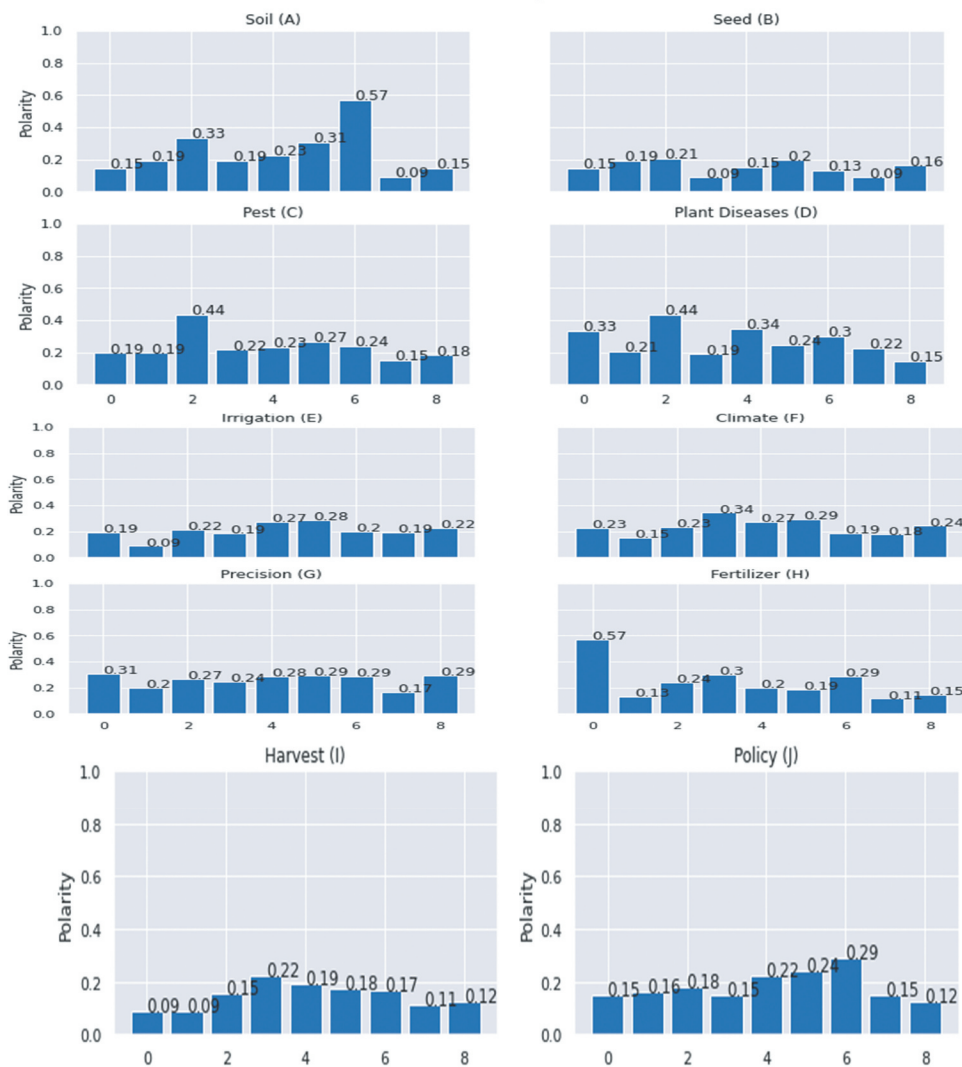
Figure 3. Document comparison.

Results and Discussion

N-Grams

In this study, we have used the top twenty **(20)** features of words using N-Gram to see which word frequently appeared in the dataset from agrarians' opinion. In Uni-gram (also called unique word or single feature) agrarian experts primarily focused on "**Crop, Soil, and Water**" with frequencies **430, 235, and 215**, respectively. Word **Crop** is more dominant because of high frequency than others: Soil, Water, Plant, Seed, Insects, Agriculture, Production, Productivity and many more in the graph (Figure 4). Similarly, **71** times "**Crop Productivity, 64 Certified Seed, and 56 times Crop**
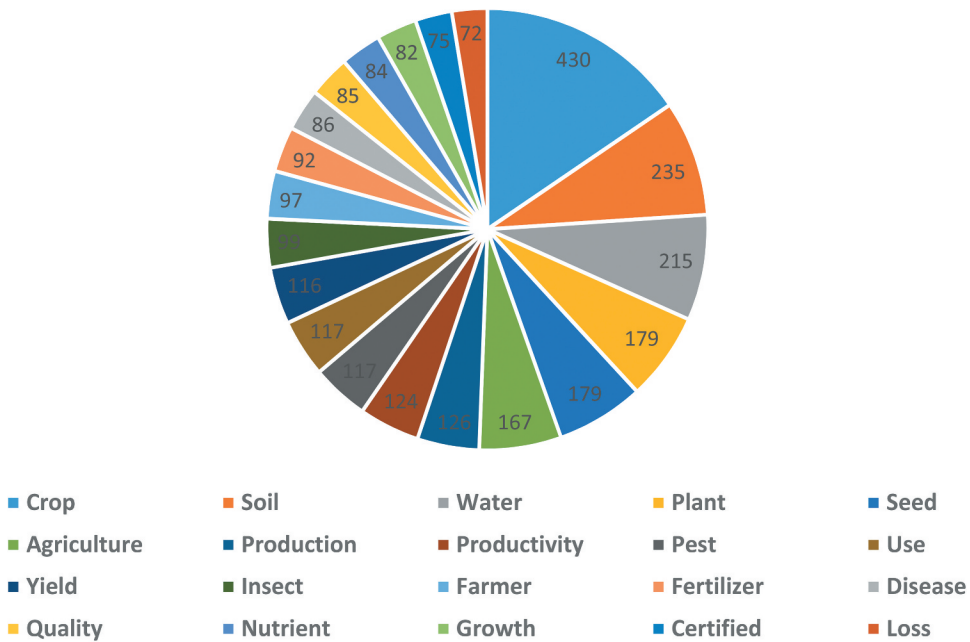
Figure 4. Uni-Gram.

Production" word was repeated from respondents' feedback in BI-gram as shown in graph (Figure 5). Insect's Pest, Precision, Soil Health, Post-Harvest, Climate Change, Synthetic Fertilizer, Uncertified Seed, and so on are presented on behalf of frequency. In Tri-gram

from agrarian experts most repeated words were **Post-Harvest Loss 20** times, **Diseases Insects Pest, 17** times, and **Plant Diseases Insect 14** times. Alike High-Efficiency Irrigation, Integrated Pest Management, Efficiency Irrigation System, Crop Productivity Soil, Use Synthetic Fertilizer are prominently based on frequency (Figure 6).

Model Training and Evaluation

In literature, various machine learning algorithms have been used for semantics analysis such as K-Nearest Neighbors (Hmeidi, Hawashin, and El-Qawasmeh 2008), Support Vector Machine (Cortes and Vapnik 1995), Neural Networks, Decision Tree, and Naive Bayes (Xia and Wang 2004). We applied K-Nearest Neighbors and Naïve Bayes algorithms for the classification of the text. Both algorithms performed better and obtained reasonable accuracy of the K-Nearst Neighbors and Naïve Bayes 84% and 87%, respectively tabulated in Table 2. The machine was trained using agrarians' opinions and to test the classifier predicated results. Furthermore, finally, the model has been tested on different agriculture
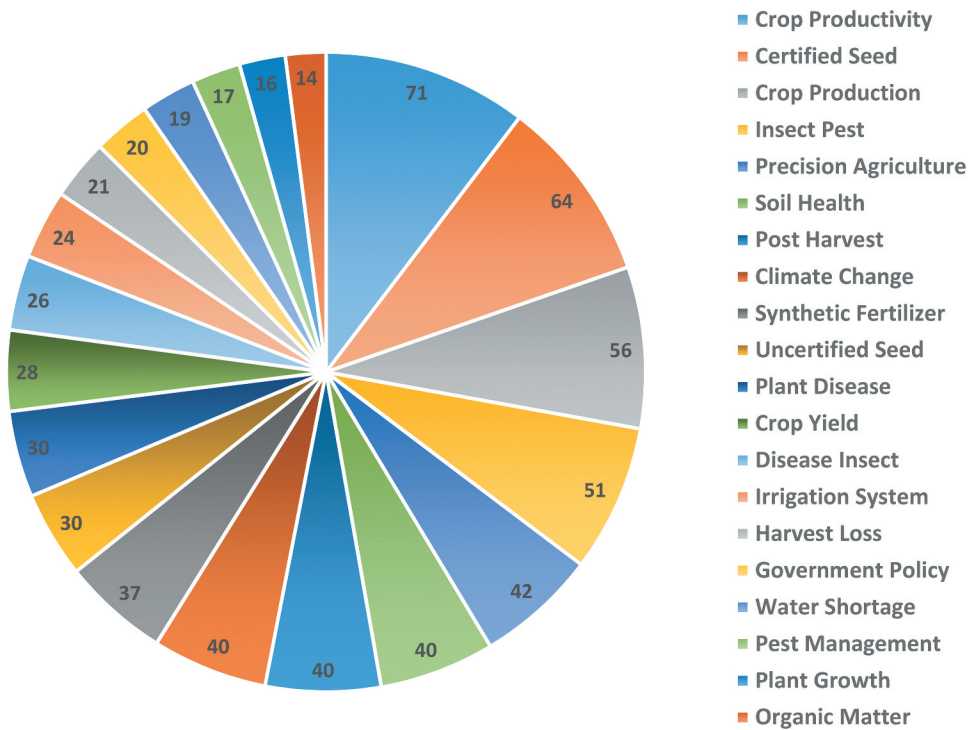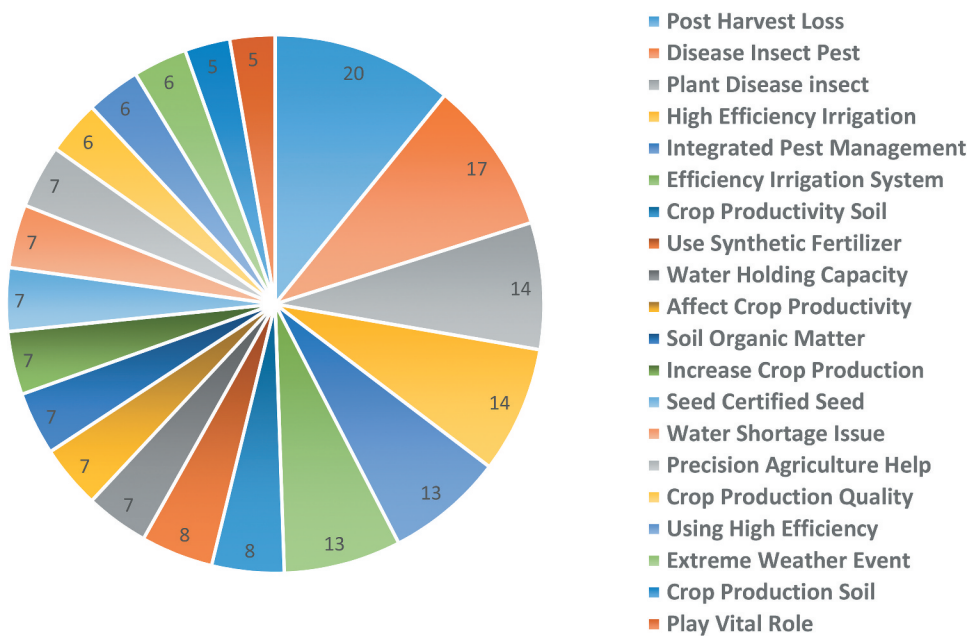
Figure 5. Bi-Gram.

Figure 6. Tri-Gram.

Table 2. Performance evaluation analysis.

Algortithm	Precision	Recall	F1 Score
KNN	85.6	85.8	84.4
Naïve Bayes	90.1	87.0	87.0

inputs: Soil, Seed Type, Pest, Plant, Diseases, Irrigation, Climate, Precision, Fertilizer, Harvest, and Policy and received relevant results of the queries. Moreover, model has evaluated by considering the accuracy, precision, recall, F1 Score (Wang and Li 2019).

Precision

The precision is also known as positive predicted values and it is the ratio of positive predicited value to the total predicted values and calcualted as (Haddi, Liu, and Shi 2013):

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Accuracy

Accuracy is a fraction of true prediction overall prediction formula is given below (Kowsari et al. 2019):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Recall

The Recall is a sensitivity and probability of detection i.e. (true positive rate). It is the ratio of correct positive prediction to the total positive (Haddi, Liu, and Shi 2013):

$$Recall = \frac{TP}{Tp + FN} \tag{7}$$

F1 Score

The F1 score is a measure of model accuracy on a dataset. The F1 score for the proposed classifier is calculated using (Equation 8).(Abualigah et al. 2017):

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recll} \tag{8}$$

- TP = True Positive, FP = False Positive, FN = False Negative, TN = True Negative

Discussion

Over the last decade, scientists were focused on crop inputs for finding their role in crop productivity such as soil, soil types, soil humidity (Dongare 2020), fertilizer management ((Saikai, Patel, and Mitchell 2020), crop management, seed, water temperature, climate changes, sustainability, chemical spry (Karthikeyan et al. Karthikeyan, et al., 2020). The present study has been conducted on "Agriculture Semantics Analysis" through machine learning model. The proposed study may also help in developing agricultural policies at the government level and a comparison is also made to find similarities between agrarians experts opinions. This research demonstrated that Naïve Bayes had better performed on experts opinions textual data. Supervised techniques like support vector machine, neural network, decision tree, random forest was not applied due to small dataset.The limitations of the proposed study are that the sample size needs to add more responses. There is a lack of previous research to compare and develop a real-time platform for disseminating findings to farmer communities.

Conclusion

In this research, we have presented a novel approach for collecting and analyzing the descriptive opinions of agricultural experts. The study has shown that **Crop, Certified Seed, and Post-Harvest Loss** are the significantly contribute to agricultural productivity. Similarities between agrarian's responses that belonged to different Institute/DAI, Academia have determined through cosine similarity and document comparison method.. This study was carried out using machine learning algorithm such as Naïve Byes and KNN algorithm and obtained 87% and 84% accuracy respectively. This study demonstrated that Naïve Byes has better performed better on text dataset of agriculture experts opinions.

Future Work

In future, agrarian opinions will be recorded in their voices for increasing good crop productivity. Their opinions and gestures will be analyzed and translated into multi-languages by using different deep leanring approaches. It will enhance the study and provide more valuable results due to enormous response level from agraian expets.

References

Abd-Elmabod, S. K., M. Muñoz-Rojas, A. Jordán, M. Anaya-Romero, J. D. Phillips, J. Laurence, Z. Zhang, P. Pereira, L. Fleskens, and M. van der Ploeg. 2020. Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. *Geoderma* 374: 114453. doi:10.1016/j.geoderma.2020.114453.

Abualigah, L. M., A. T. Khader, M. A. Al-Betar, and O. A. Alomari. 2017. Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. *Expert Systems with Applications* 84: 24–36. doi:10.1016/j.eswa.2017.05.002.

Ahmad, K., and A. C. T. Heng. 2012. Determinants of agriculture productivity growth in Pakistan. *International Research Journal of Finance and Economics* 95: 163–173.

Angiani, G., L. Ferrari, T. Fontanini, P. Fornacciari, E. Iotti, F. Magliani, and S. Manicardi. 2016. A comparison between preprocessing techniques for sentiment analysis in Twitter. *KDWeb* 7 (2): 37–56.

Benos, L., A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis. 2021. Machine Learning in agriculture: A comprehensive updated review. *Sensors* 21: 3758. doi:10.3390/s21113758.

Bontcheva, K., and D. Rout. 2014. Making sense of social media streams through semantics: A survey. *Semantic Web* 5: 373–403. doi:10.3233/SW-130110.

Chowdhury, G. G. 2003. Natural language processing. *Annual Review of Information Science and Technology* 37: 51–89. doi:10.1002/aris.1440370103.

Cortes, C., and V. Vapnik. 1995. Support-vector networks. *Machine learning* 20: 273–297.

Dai, X., Z. Zhuang, and P. X. Zhao. 2011. Computational analysis of miRNA targets in plants: Current status and challenges. *Briefings in Bioinformatics* 12: 115–21. doi:10.1093/bib/bbq065.

Dongare, M. 2020. Smart E-agriculture system using IoT and machine learning. *Studies in Indian Place Names* 40: 2486–93.

Elahi, E., Z. Khalid, C. Weijun, and H. Zhang. 2020. The public policy of agricultural land allotment to agrarians and its impact on crop productivity in Punjab province of Pakistan. *Land Use Policy* 90: 104324. doi:10.1016/j.landusepol.2019.104324.

Farooq, M. S., S. Riaz, A. Abid, T. Umer, and Y. B. Zikria. 2020. Role of IoT technology in agriculture: A systematic literature review. *Electronics* 9: 319. doi:10.3390/electronics9020319.

Haddi, E., X. Liu, and Y. Shi. 2013. The role of text pre-processing in sentiment analysis. *Procedia Computer Science* 17: 26–32. doi:10.1016/j.procs.2013.05.005.

Hmeidi, I., B. Hawashin, and E. El-Qawasmeh. 2008. Performance of KNN and SVM classifiers on full word Arabic articles. *Advanced Engineering Informatics* 22: 106–111. doi:10.1016/j.aei.2007.12.001.

Hoang, T.-A., W. W. Cohen, E.-P. Lim, D. Pierce, and D. P. Redlawsk. 2013. Politics, sharing and emotion in microblog. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Niagara Falls, ON, Canada, 282–89.

Hutto, C. J., and E. Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*, Italy, 216–225.

Ikonomakis, M., S. Kotsiantis, and V. Tampakas. 2005. Text classification using machine learning techniques. *WSEAS Transactions on Computers* 4: 966–974.

Jayaraman, P. P., D. Palmer, A. Zaslavsky, and D. Georgakopoulos. 2015. Do-it-yourself digital agriculture applications with semantically enhanced IoT platform. In *IEEE tenth international conference on intelligent sensors, sensor networks and information processing (IP)*, 1–6 IEEE, Singapore.

Kantasa-ard, A., M. Nouiri, A. Bekrar, A. Ait El Cadi, and Y. Sallez. 2020. Machine learning for demand forecasting in the physical internet: A case study of agricultural products in Thailand. *International Journal of Production Research* 1–25. doi:10.1080/00207543.2020.1844332.

Karthikeyan, P., K. Velswamy, P. Harshavardhanan, R. Rajagopal, V. JeyaKrishnan, and S. Velliangiri. 2020. Machine learning techniques application: Social media, agriculture, and scheduling in distributed systems. In *Handbook of research on applications and implementations of machine learning techniques*, 380–401. IGI Global, CMR Institute of Technology, India.

Kowsari, K., K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown. 2019. Text classification algorithms: A survey. *Information* 10: 150. doi:10.3390/info10040150.

Li, H., H. Xiao, T. Qiu, and P. Zhou. 2013. Food safety warning research based on internet public opinion monitoring and tracing. In *Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 481–84. IEEE, Fairfax, VA, USA.

Liu, B, and L. Zhang. 2012. A survey of opinion mining and sentiment analysis mining text data. Springer 415–463.

Martini, D., E. Mietzsch, M. Schmitz, and M. Kunisch. 2011. *The agriXchange platform as a means for coordination and support on data exchange in agriculture*, ed. E. Gelb, and K. Charvt, vol. 11. EFITA/WCCA, Darmstadt, Germany.

Mekala, M. S., and P. Viswanathan. 2017. A survey: Smart agriculture IoT with cloud computing. In *International conference on microelectronic devices, circuits and systems (ICMDCS)*, 1–7. IEEE, Vellore, India.

Mirończuk, M. M., and J. Protasiewicz. 2018. A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications* 106: 36–54. doi:10.1016/j.eswa.2018.03.058.

Muharam, F. M., K. Nurulhuda, Z. Zulkafli, M. A. Tarmizi, A. N. H. Abdullah, M. F. Che Hashim, S. N. Mohd Zad, D. Radhwane, and M. R. Ismail. 2021. UAV-and Random-Forest-AdaBoost (RFA)-based estimation of rice plant traits. *Agronomy* 11: 915. doi:10.3390/agronomy11050915.

Prajapati, B. P., and D. R. Kathiriya. 2016. Evaluation of effectiveness of k-Means cluster based fast k-nearest neighbor classification applied on agriculture dataset. *International Journal of Computer Science and Information Security* 14: 800.

Prathibha, S., A. Hongal, and M. Jyothi. 2017. IoT based monitoring system in smart agriculture. In *International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, 81–84. IEEE, Bangalore, India.

Razzaq, A., M. Asim, Z. Ali, S. Qadri, I. Mumtaz, D. M. Khan, and Q. Niaz. 2019. Text sentiment analysis using frequency-based vigorous features. *China Communications* 16: 145–153. doi:10.23919/JCC.2019.12.011.

Saif, H., M. Fernandez, Y. He, and H. Alani. 2013. 1st Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013), 3 Dec 2013, Turin, Italy.

Saikai, Y., V. Patel, and P. D. Mitchell. 2020. Machine learning for optimizing complex site-specific management. *Computers and Electronics in Agriculture* 174: 105381. doi:10.1016/j.compag.2020.105381.

Singh, T., and M. Kumari. 2016. Role of text pre-processing in twitter sentiment analysis. *Procedia Computer Science* 89: 549–54. doi:10.1016/j.procs.2016.06.095.

Smith, M. J. 2018. Getting value from artificial intelligence in agriculture. *Animal Production Scienc* 60: 46–54. doi:10.1071/AN18522.

Soucy, P., and G. W. Mineau. 2001. A simple KNN algorithm for text categorization. In *Proceedings International Conference on Data Mining*, 647–48. IEEE, San Jose, CA, USA.

Tripathy, A., A. Agrawal, and S. K. Rath. 2015. Classification of sentimental reviews using machine learning techniques. *Procedia Computer Science* 57: 821–829. doi:10.1016/j.procs.2015.07.523.

Vaghela, V. B., B. M. Jadav, and M. Scholar. 2016. Analysis of various sentiment classification techniques. *International Journal of Computer Applications* 140: 0975–8887.

Wang, R., and J. Li. 2019. Bayes test of precision, recall, and F1 measure for comparison of two natural language processing models. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4135–45, Florence, Italy.

Weiss, M., F. Jacob, and G. Duveiller. 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment* 236: 111402.

Wolf, G. M. B. 1991. *The beginnings of semantics. Essays, lectures and reviews*. Duckworth London, Stanford University Press.

Xia, Y., and J. Wang. 2004. A one-layer recurrent neural network for support vector machine learning. Transactions on Systems, Man, and Cybernetics, Instt. Eelect. Elect. Eng. 34: 1261–1269.