Nutritional and Natural Farming: Agriculturists' Sentiment Analysis

Manish Rao Ghatge[1]*, Snehlata Barde[2]

[1] Research Scholar, MSIT, MATS University, Raipur, Chhattisgarh, India.

[2] Professor, MSIT, MATS University, Raipur, Chhattisgarh, India.

Email- [1] ghatge.manish@gmail.com

ABSTRACT:

The popularity of nutritional and natural farming is increasing because it saves the environment and makes a significant impact on the social and economic factors of the country. Our country endowed with native technologies & growth potential for nutritional and natural farming. India has lagged far behind in adopting nutritional and natural farming for several reasons. In the context of nutritional and natural farming, the agriculturists already related to this are supporting this but lots of other agriculturists have negative sentiments about nutritional and natural farming. In this article, we predicted M/L (Machine Learning) methods for performing sentiment analysis build on obtained comments containing Hindi, English and Chhattisgarhi languages data. Furthermore, authors concentrated on the accuracy and performance of the agricultural dataset to predict the sentiment of the tested dataset.

Keywords: Nutritional, Natural, Farming, Sentiment Analysis, Agriculture, Chhattisgarhi, Hindi.

INTRODUCTION:

In the quest for safer food, health benefits and food safety concerns have increased the demand for organically grown food over the past decades. Nutritional and natural farming can be understood as growing in absence of chemicals, genetically modified or synthetic products. Now public believes that chemically grown foods have significant negative health effects because of high levels of chemicals residues, high levels of nitrates, pesticides and the presence of genetically modified organisms, so they are very much concerned with this topic of natural farming. They express their views on websites about it.

Websites' play a crucial role in gathering public opinion; these opinions play an important role in making decisions. An online survey on the topic of nutritional and natural farming is conducted and a lot of agriculturists shown keen interest in this topic.

Most of the researches builds upon their work and introduces machine learning in the Sentiment Analysis on Hindi language and taken data from blogs in various website or social media as input. For implementation of machine language approaches first of all one require dataset for that we have to collect data from various sources then after completion of the pre-processing one will be training their models using machine learning approaches [1] [2].

A Chhattisgarhi, Hindi and English mix language SA corpus is build, which collects agriculturist's thoughts on nutritional and natural farming. Second part of the paper is literature review. 3rd Section is methodology discussion. Next section is Experimental setup. Conclusion brings the paper to a close.

RELATED WORK:

Vrunda C. Joshi et al. presented the practical approach to analyze tweets and sort them in positive and negative. Twitter Data is taken for experiment. Support Vector Machine (SVM) performed very well in SA, which is a very popular choice of scientists for NLP. [3]

Ghatge M.R. et al. made use of the basic concepts of Sentiment analysis on Indian language text. Many M/L techniques applied on corpus. Due to unavailability of rich lexicon resources for Hindi the improved machine learning techniques has been applied. [4]

Fazil et al. Presented that the M/L algorithms to build classifier is very effective. To sort the sentiment of Twitter users in positive and negative they have used M/L algorithms [5]

Yaman Kumar et al. introduced the BHAAV from Hindi stories. He notes that in addition to studying the sentiments of Hindi texts, the presented corpus also enables research related to the analysis of Hindi literature in terms of identifying intrinsic emotional arcs. It also has the potential to facilitate research related to human text-to-speech systems aimed at enhancing automated storytelling. Preserve the order of the sentences as they appear in the original story. Through different observations we confirm BHAAV to be rich with emotion cues and point to the potential applications. [6]

Dashtipouret al. presented that the A lexicon-based approach has a dictionary of phrases with scores that indicate polarity. They differ in the source of these words, the size of the dictionary, and the method used to assign scores to them. The structure of such a lexicon is subjective, so all these data have very few duplicates. [7]

Aditya Mogadala et al. presented the sentence-level which is Subjective classification using language-independent feature weighting and consistent selection methods across languages. In experiments performed in five different languages, including English and the South Asian language Hindi, the entropy-based ECCD (Category Coverage Difference Criterion) feature selection method using a language-independent feature weighting method outperformed other subjective has been shown to outperform conventional classification approaches. [8]

Md Shad Akhtar et al. It provides benchmarking settings, creates high-quality annotated datasets, builds machine learning models for sentiment analysis that demonstrate effective use of datasets, and finally community resources for further use in research. Presented SA challenges in Hindi by providing progress. We use Conditional Random Fields (CRF) and Support Vector Machines (SVM) as classification algorithms for aspect term extraction and sentiment analysis. [9]

Kishorjit Nongmeikapam et al. represented the SA of the Manipuri article. Language is very agglutinative in nature. The file is a letter to the editor of a local daily newspaper. The text is processed for part-of-speech (POS) tagging using a Conditional Random Field (CRF). The verb lexicon is manually modified with mood polarity (positive or negative or neutral). Using the POS tagger, the verbs of each sentence are identified, and a modified verb dictionary is used to convey the emotional polarity of the sentence. [10]

METHODOLOGY:

In our experimentation, we developed a dataset of Hindi-English-Chhattisgarhi mixed data.

Table 1: Language mixed data.

index	label	text
1	1	जैविक खेती से मय ह बहुत खुश हव $$$$$$ वैरी हैप्पी $$$$
2	1	बहुत बढिया सभी किसान सहमत हैं जैविक खेती के लिए #किसान_जिंदाबाद
3	0	चलो भागो इन्हा से हुर्र रररररररररर
4	0	हमसे न हो पायेगा;
5	0	हमर पैसा हराम" के हरे का खर्चीला है
6	1	बहुत अच्छा #किसान एकता जिंदाबाद
7	1	मै आज बहुत खुश हूँ
8	0	यहां का किसान बिना खाद के मर गया है;
9	1	किसान ने हँसकर कहा हाँ वहीं तो जा रहा हूँ

1) Data pre-processing

To get better results of the proposed approach, a cleaning process is performed for data collection, dictionary generation.

i) Data cleaning

After data collection, data cleansing plays an important role. This step is used to extract relevant data. Data always contains unnecessary information. Therefore, it is important to pre-process the text so that unnecessary data can be removed. It also uses a pre-processing step to remove unnecessary tokens such as # tags, punctuation, repeated characters, URLs, spaces between words, and sentiment. In the proposed system, the following pre-processing steps are applied to remove special symbols.

Table 2: Language mixed data after cleaning.

index	label	text
1	1	जैविक खेती से मय ह बहुत खुश हववैरी हैप्पी
2	1	बहुत बढिया सभी किसान सहमत हैं जैविक खेती के लिए किसान_जिंदाबाद
3	0	चलो भागो इन्हा से हुर्र ररर
4	0	हमसेनहीं हो पायेगा
5	0	हमर पैसा हरामके हरे क्याखर्चीला है
6	1	बहुत अच्छा किसान एकता जिंदाबाद
7	1	मै आज बहुत खुश हूँ
8	0	यहां का किसान बिना खाद के मर गया है
9	1	किसान ने हँसकर कहा हाँ वहीं तो जा रहा हूँ

i) Tokenization

Tokenization is the process of breaking down words so that a computer can convert text into words. It divides a portion of text data into smaller meaningful units called tokens. Spaces and punctuation marks can be used to separate individual tokens of a sentence. In the proposed system, mixed-language specifications of tokens are given as input for identifiers. Consider the statement:-"मै आज बहुत खुश हूँ"will be tokenized as ['मै', 'आज', 'बहुत', 'खुश', 'हूँ'].

2) NAIVE BAYES CLASSIFIER

Naive Bayes is a conditional probability model, based on Bayes' theorem, which states that the conditional probability is given by -

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

For each of the K possible outcomes. Using a chain rule for iteratively applying the conditional probability definition:

$$p(C_k, x_1, \ldots, x_n) = p(x_1, \ldots, x_n, C_k)$$

$$= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2, \ldots, x_n, C_k)$$
$$= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k)\, p(x_3, \ldots, x_n, C_k)$$
$$= \cdots$$

$$= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k) \cdots p(x_{n-1} \mid x_n, C_k)\, p(x_n \mid C_k)\, p(C_k)$$

Under the naive assumption that all the features x are mutually independent, the probability now becomes:

$$p(C_k \mid x_1, \ldots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

This formula is the base behind the Multinomial Naive Bayes classifier which we used, which deals with the occurrences of a word in a single document.

EXPERIMENTAL SETUP

We have collected Chhattisgarhi Mixed data in Hindi and English. After pre-processing the data, a cleaned dataset is prepared for better results of the proposed approach.

Table 3. Statistics of Dataset

Properties	Size
Positive reviews	6701
Negative reviews	8400
Total reviews	15101

For the experiment, multilingual text data in Hindi, English and Chhattisgarh were collected from an online survey on food and natural farming. A total of 15101 comments were collected and the data are saved in text files with (.csv) and (.txt) extensions.

User Reviews are divided into two sections. Training covers 90% of user reviews and testing accounts for 10%. The machine learning model experiments used the default parameters.

In our experiments, the data was sorted by labels and we shuffled them before splitting so that the same shuffling was done each run of the code. We trained the model by giving it more labels.
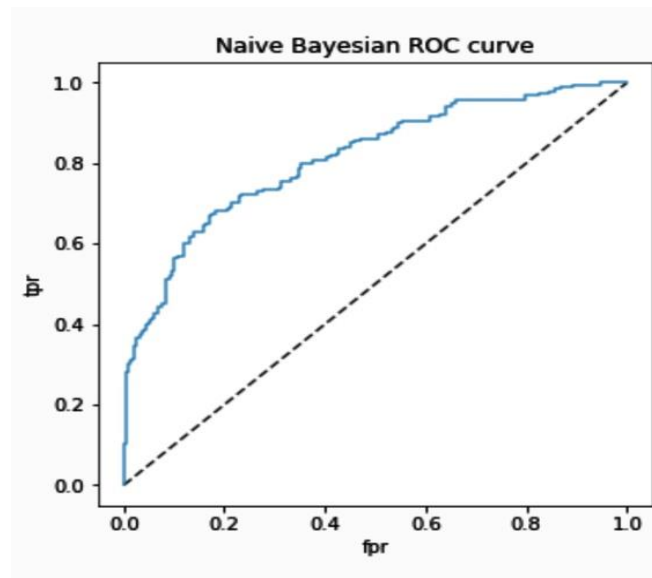
CONCLUSION:

In this article, we extracted Hindi and English Chhattisgarh text data on nutritional and natural farming from an online survey and analyzed sentiment using the Naive Bayes classifier.

Data related to the agricultural domain were collected from websites, cleaned and pre-processed. Hindi-English-Chhattisgarhi text data polarity experiments have shown accuracy up to 78%. In this work, the authors analysed that improved results were obtained after pre-treatment. Further, we will plan to explore other sentiment analysis tools and methods. By incorporating deep learning, there is still room for improvement.

ACKNOWLEDGMENTS:

REFERENCES:

1. Ghatge, M.R., Barde, S. (2023). The Sentiment Analysis for Hindi Language Using Convolution Neural Network and Deep Belief Network. In: , et al. Information Systems and Management Science. ISMS 2021. Lecture Notes in Networks and Systems, vol 521. Springer, Cham. https://doi.org/10.1007/978-3-031-13150-9_1

2. Joshi, A,et al.(2010). A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study. In: ,ICON 2010, 8th International Conference on Natural Language Processing.

3. Vrunda C. Joshi et al. (2017). An Approach to Sentiment Analysis on Gujarati Tweets. In: , Advances in Computational Sciences and Technology, ISSN 0973-6107, Vol.10, No.5 (2017), pp. 1487-1493.

4. Ghatge, M.R. and Barde, S., (2022). Comparison of CNN-LSTM in Sentiment Analysis for Hindi Mix Language. In: , 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 453-456).IEEE. https://doi.org/10.1109/ICTACS56270.2022.9988531

5. Fazil, M., &Abulaish, M. (2018). A Hybrid Approach for Detecting Automated Spammers in Twitter. IEEE Transactions on Information Forensics and Security, 13(11), 2707–2719. https://doi.org/10.1109/TIFS.2018.2825958

6. Yaman Kumar, DebanjanMahata, SagarAggarwal, AnmolChugh, RajatMaheshwari, Rajiv Ratn Shah, "BHAAV (भाव) - A Text Corpus for Emotion Analysis from Hindi Stories", IIIT-Delhi, India.

7. Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y. A., Gelbukh, A., & Zhou, Q. (2016). Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. Cognitive Computation, 8(4), 757–771. https://doi.org/10.1007/s12559-016-9415-7

8. AdityaMogadala, VasudevaVarma, "Language Independent Sentence-Level Subjectivity Analysis with Feature Selection", Copyright 2012, 26th Pacific Asia Conference on Language,Information and Computation, pp: 171–180.

9. Md Shad Akhtar, AsifEkbal and Pushpak Bhattacharyya, "Aspect based Sentiment Analysis in Hindi: Resource Creation and Evaluation", http://WWW.howws.com.

10. KishorjitNongmeikapam, DilipkumarKhangembam, WangkheimayumHemkumar, ShinghajitKhuraijam and SivajiBandyopadhyay,"Verb Based Manipuri Sentiment Analysis", International Journal on Natural Language Computing (IJNLC), Vol.3, No.3, June 2014.

11. Majumdar, J., Naraseeyappa, S., &Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. Journal of Big Data, 4(1). https://doi.org/10.1186/s40537-017-0077-4