IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.12, December 2022 E-ISSN: 2347-2693

Smart Interactive Farmer-Bot Using Bagging & Natural Language Processing

Uddhav Patil[1], Nikhil Kinikar[2*], Madanu Amisha[3], Aishwarya Chouthe[4], Rajeshwari Goudar[5]

[1,2,3,4,5]School of Computer Engineering and Technology, MIT Academy of Engineering, Alandi, Pune, Maharashtra, India

**Corresponding Author: nmkinikar7@gmai.com*

DOI: https://doi.org/10.26438/ijcse/v10i12.813 | Available online at: www.ijcseonline.org

Received: 19/Sept/2022, Accepted: 21/Oct/2022, Published: 31/Dec/2022

Abstract- About 58% of Indians rely on agriculture as their main source of income, and this sector contributes significantly to the nation's Gross Domestic Product (GDP). As the agricultural process evolves, it is increasingly crucial to disseminate information about it when it becomes available. Improvements in the crop production and weather (rainfall) prediction can be increased with the right data and prediction responses obtained in relation to weather and regions. This can be accomplished with a farmer-chatbot, which is a conversational software that uses pre-programmed responses or artificial intelligence to answer the queries without the intervention of a human. The proposed Farmer-chatbot system, allows farmers to ask questions and receive precise replies in common language. The query is processed using the Ensemble Learning Algorithm and the Natural Language Processing to anticipate solutions to the users.

Keywords-Ensemble Learning, Bagging Method, Natural language processing, Data Cleaning, Prediction.

I. INTRODUCTION

The Indian economy is heavily reliant on agriculture, however Indian farmers are currently suffering a major problem in terms of their livelihood due to less knowledge about modern agricultural technologies, government initiatives, equipment, and related other issues. The majority of farmers in rural areas are unaware of technology and information about crops, soil qualities, and the up-to-date tools available. Crop production and livestock outcome will suffer as a result of a lack of knowledge about the latest agricultural technology and information about farming procedures [1].

Since Agriculture shows a growing adoption of new technologies[2], and an intensified use of electronic devices and business models, in order to improve product quality, increase production, reduce costs and the impact on the environment [3], one can address the difficulty of delivering qualitative solutions or information to farmers by gathering data about agriculture such as the climate conditions like rainfall, soil type, crop productions, and many more with the use of Artificial Intelligence Machine Learning And Data Science.[4][5].

The goal of a chatbot framework is to mimic human conversation. To do this, a chatbot's architecture combines a language model with computational calculations to mimic the casual dialogue between a human user and a computer using everyday language. The user answer is generated using AIML algorithms, and if the inquiry is not addressed or does not offer a response, it is forwarded to the concerned experts to get the other details.[6] The work completed by others that serves as the foundation for the background concept and created model is presented in section II. The proposed system details are covered in section III, along with the technology and methodology used. The result analysis of the system is discussed in section IV. The concussion and future scope is discussed in section V and VII respectively.

II. RELATED WORK

Talkbot, which one can say “a piece of regular language preparation” is developed for Extraction of data using Artificial Intelligence methods [7]. The model is prepared according to the human language to solve all the queries over application. The Horticulture area is derived by this application to fulfill the current situations. This model helps to overcome the issues made by existing AI methods. The model used in this application provides a lot of advantages over the previous solutions like it is more convenient than typing up the query. But the technologies used by this AI technology have some limitations and therefore raises some issues that lead to accuracy.

The paper proposed for the Productivity of farmers. This is mainly focused on the paper. It proposes a system with portability and having intelligence uses data mining and data mining techniques which assists farmers and users with different farming g related techniques and methods. Also helps farmers to decide the various production of crops based on the current weather, climate conditions as well as soil conditions [8]. The E-Agro chatbot creates a chat room and chatbot to help farmers make timely decisions on farming by utilizing peer discussion and subject matter experts. The Artificial Markup Language is used to train the model which will forecast the output

depending on the given query . Because the chatbot is deployed on a cloud platform, the client doesn't need more computing power to access this system. The technique makes use of WebSocket and socket.io, an event-based WebSocket emulation. The training model utilized a natural language processing technique and artificial intelligence.[9]

Artificial Intelligence and machine learning technology are taking the industry to new heights. This "The TalkBot" technology gives farmers a platform to get the information they want and to update in accordance with upcoming market trends and technologies with a very user-friendly interface.

Users of this chatbot system can converse with it as if it were a real person, acting as their virtual conversational assistant. [10]. The bot is developed in a more intellectual way that it can process queries with incorrect or incomplete sentences and incorrectly spelled words. The system employs Natural Language Processing to parse user queries, detect key terms, and provide accurate responses, which facilitates communication. In order to improve comprehension, categorization algorithms are used to generate the responses. Using text-to-speech algorithms, the bot also offers responses that are voice-oriented.

Farming-related technology have greatly advanced, and current market trends and farming practises are both evolving quickly. However, farmers must conduct numerous searches on online resources in order to find the information. The system "The Farmerbot" solves this issue and gives farmers a greater chance to get the information they need to keep up with market trends and technological advancements. This programme is essentially a chatbot that serves as the user's virtual conversational assistant and provides information in a very user-friendly manner.[11]. Using data from prior years' records, the system employs prediction algorithms to forecast future data, such as the price of crops for upcoming years. By using text to speech techniques the bot can provide voice oriented responses as well.

Without assistance from people, the automatic talkbot may also respond to user inquiries. To do this, several question kinds are used to teach the conversation bots. This talkbot employs Naive Bayes algorithms to choose the best responses from a collection of taught questions in order to respond to user inquiries. This bot is made to respond to farmer-related questions. The talk bot responds to these questions via voice and chat. [12]. If a farmer has any questions regarding agriculture, the bot will respond to them by voice or text. If the farmer needs help finding a route, the bot will display a Google Maps route. It also has a reminder feature in case a farmer forgets agriculture details and few important queries.

Agriculture monitoring systems using artificial intelligence helps farmers in making the best crop choices based on market conditions, understanding climatic conditions, and understanding which crops to choose based on the condition of the soil. It also helps farmers identify specific crop diseases. The technology stack contains Django framework, online API, Artificial intelligence algorithms and agent from scratch. System help farmers automate their farming, also encourages them to switch to precision farming for increased crop quality, growth while utilizing less resources[13]

III. METHODOLOGY

The system will gather the essential agricultural data in the form of datasets from various websites and government repositories, such as climate, crop data, etc. Similar to how chat messages are treated, all acquired data will be subsequently cleaned to eliminate any irrelevant information. The data from the datasets are thoroughly trained using machine learning techniques, such as the Bagging method and then the accuracy of the trained data is tested.

Natural Language Processing (NLP) is used to generate tokens by tokenizing user queries for required keywords, which are then compared to data in datasets for crop or weather prediction. Finally, NLP is used to generate required responses for user queries after all processing, including data gathering, pre-processing, cleaning, training, and testing has been accomplished.

Agricultural parameters

1. Rainfall – Rainfall has a significant impact on agricultural productivity because farming is primarily rain-fed.
2. Crop yields- To investigate the regularity and changes in agricultural yields for each season, past performance is incredibly helpful.

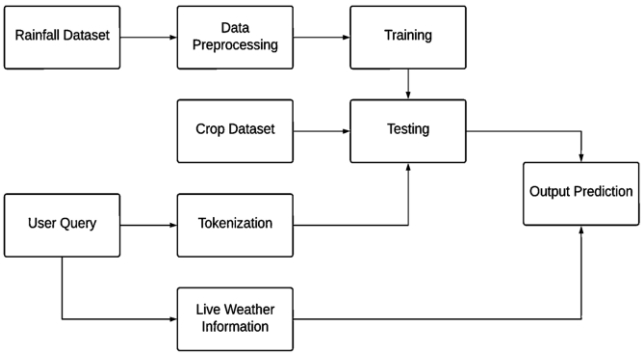

Figure 1. Block diagram of farmer-chatbot

A. DATA COLLECTION AND DATASETS

Collected data

1. Rainfall data – Data about rainfall in various regions.
2. Weather data – Data about the weather for various
3. Crop production data – Data related to crop production.

Preprocessing of Data

Before using data, raw data is transformed into desired and intelligible form via a process called data preprocessing. Data that is gathered in the real world has inaccuracies

frequently. It contains inconsistent data with missing values.

Data preprocessing steps:
1. Missing Values: In dataset missing values can be processed by putting average data value or by putting them depending on other functions.
2. Data normalization: It is required to prevent the situation when data is made of attributes of different scales as some features with higher range could impact the prediction more. For algorithms where prediction is based on measurements or distance it is helpful.

B. METHODOLOGY

Ensemble Learning

Ensemble learning is a general approach used in machine learning for better predictive performance by integrating the predictions obtained from the multiple models. Ensemble learning allows multiple ensembles for the predictive models. [4]
The two main classes of ensemble learning methods are bagging and boosting. As per the requirements, the bagging concept is used in the system.

Bagging

By lowering variance and preventing overfitting, bagging often referred to as Bootstrap aggregating, is used to increase accuracy and make the model more generalized. Various subsets of the training dataset are employed in this. A model using the same learning techniques for each subset is used to predict the outcomes for the same set of test data, such as decision trees, logistic regression, etc. In this each model is predicted individually and then uses a model averaging approach to get the final prediction outcome.

The important components of bagging method can be summarized as follows:
• Samples from the bootstrapped training dataset.
• Without pruning, decision trees were able to fit all samples.
• Voting or averaging for simple forecasts
Finally, bagging adds by modifying the training data used to fit each ensemble member, resulting in effective yet distinctive models.

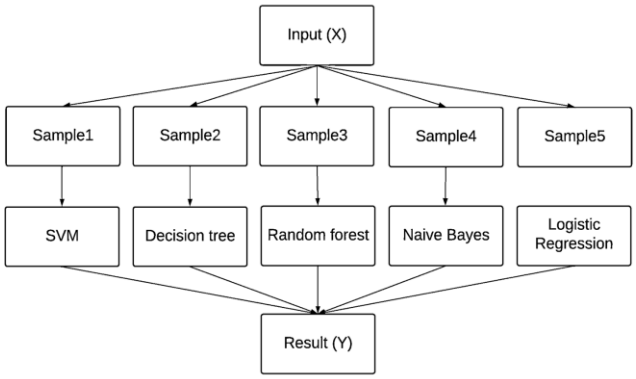

Figure 2. Bagging Method for farmer-chatbot

Decision Tree

A decision tree is a supervised machine learning method that may be applied to categorize or forecast data based on the answers to prior queries. The model is trained and tested using data sets that have the required categorization because supervised learning is by its very nature a supervised process.

The decision tree might not always present a simple choice or answer. Instead, it might present options to the data scientist, allowing them to make informed decisions on their own. Naturally, a decision tree looks like a tree. The tree begins at the root node. A series of decision nodes that show choices that should be made after the root node. The decision nodes give rise to leaf nodes that represent the choice's effects. Each decision node serves as a query or split point, and the leaf nodes that branch out from it show the potential answers.

Random Forest

Random forest is built using decision trees. This algorithm is used for modeling predictions and behavior analysis. In a random forest numerous decision trees are employed where each tree represents an instance of how the classification of input data is done in a random forest.

Steps for random forest algorithm:
1. N numbers of random records are taken.
2. For each sample individual decision trees are constructed.
3. Outputs of each decision tree are generated
4. With the help of majority voting final output is created.

Naive Bayes
Naïve Bayes employs a group of algorithms based on the Bayes Theorem. Naïve bayes are primarily used when a dataset has a feature matrix and the response vectors. As Naïve Bayes makes an assumption regarding each feature that it is an equal and independent feature. Usually the assumptions made by Naïve Bayes do not fit with real world incidents, moreover the independence assumptions are often incorrect although they work well in practice.[14]
Bayes' Theorem uses probability for the classification for that it uses mathematical equation

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where, A and B denotes to the events
Using this we find the probability of event A with event B which is true in nature and used as evidence. Also P(A) is the priori of A whereas P(A|B) are the posterior probability of B.

SVM
Support Vector Machine is a supervised learning algorithm. It generates dimensions of hyper planes using a variety of features.

Without many adjustments, SVM can uncover complex relationships between data. An operation that converts low-dimensional input space into higher-dimensional space is known as the SVM kernel. Kernel performs a variety of extremely complex data transformations before deciding how to separate the input in line with the desired labels or outputs.

Logistic regression

Logistic regression is primarily a supervised classification algorithm. It is used to build a regression model which predicts the probability to answer if given entry data belongs to a certain category. To do so it employs sigmoid function which is an mathematical function that takes real value to map it between 0-1.

$$Y = 1 / 1+e^{-z}$$

When the value of z reaches positive infinity it makes a predicted value of y 1 and if it reaches negative infinity it makes predicted value of y 0. So if the Y>0.5 we classify it as positive if it is Y<0.5 we classify it as a negative label. [15][16]

Natural Language Processing

Natural Language Processing is the machine language manipulation, analysis, understanding, and interpretation technology. It makes it possible for the system to comprehend human input, even when it contains errors in grammar or incomplete sentences. This improves the categorization algorithm's effectiveness as well. [17]

The steps involved in the NLP-

1. Tokenization: Tokenization divides the message into a few words in the first procedure. The filtering method is then used to remove stop words or words that are not necessary [18]. Input query will be converted into tokens.
2. Noise Removal and Stop Words: Stop words (noise) such as "is, the, was, are, be, will, etc." are deleted. Noise removal also involves deleting words that aren't necessary for understanding the context.
3. Lexicon Normalization: Lexicon normalization is the process of combining all of the input representations into a one single representation.
4. Bag of Words or Vector Space model: Bag of word assigns weight to the tokens in the form of 1 or 0 to generate vector space.

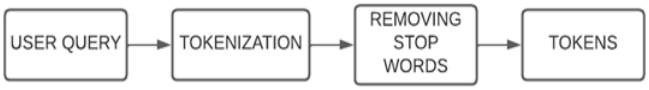

Figure 3. NLP Process

IV. RESULTS AND DISCUSSION

The results of every model used in this methodology have been calculated. The different accuracy parameters are considered for calculating the result for each model. The description of each parameter used in the model is given in Table 1. Every machine learning algorithm should be tested through some accuracy parameters. As the proposed methodology based on the classifications problems the parameters used are Mean Absolute Percentage error, , Recall, Precision, Confusion Matrix and F1 Score.

Table 1. Accuracy Parameters

Mean Absolute Percentage Error	Mean of the absolute percentage errors of forecasts.
Precision	Ratio of relevant information among retrieved.
Recall	Ratio of relevant information that were retrieved
F1 Score	The Mean value of recall and precision
ConfusionMatrix	Describe complete performance of model

The results are closely monitored with the two important parameters recall and precision. Figure 3 illustrates the impact of recall and precision.

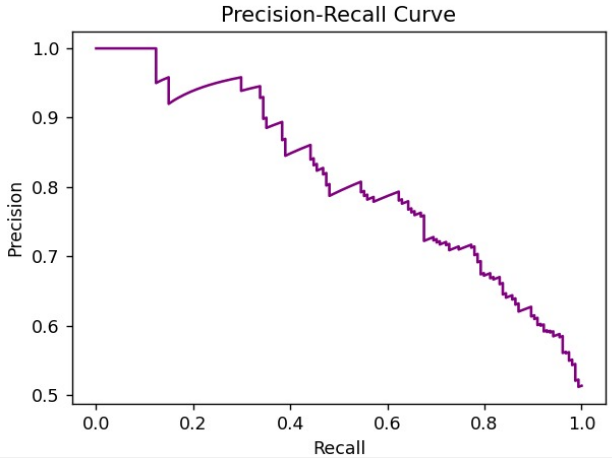

Figure 3. Precision-Recall value curve for Ensemble Model

The confusion matrix provides the curve that determines true positives and true negatives along with false positives and false negatives. The False positive rate according to the True positive rate is illustrated in Figure 4. It shows the values of the ensemble model with respect to the no skill that means zero learning.

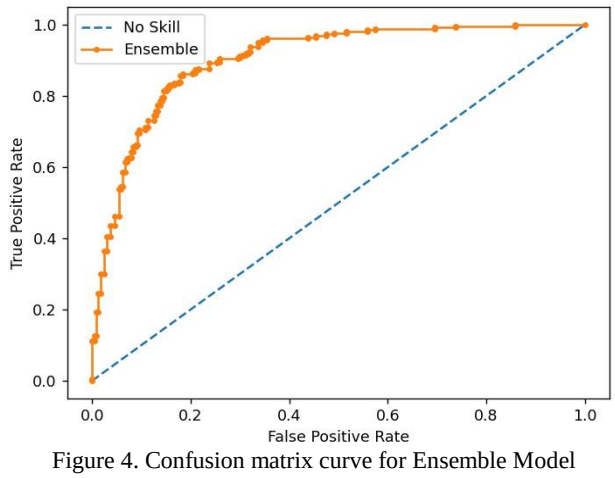

Figure 4. Confusion matrix curve for Ensemble Model

The proposed model has obtained all the precise accuracy metrics. As ensemble learning is combining or aggregating multiple models, the more the models will increase the accuracy. Figure 5 illustrates the effect on accuracy metrics when a new model is added to an existing model. First bar graph shows the accuray values with a combination of decision tree and random forest. Rest of the graphs show the values as per given in figure 5. This provides the proper results for the existing model.

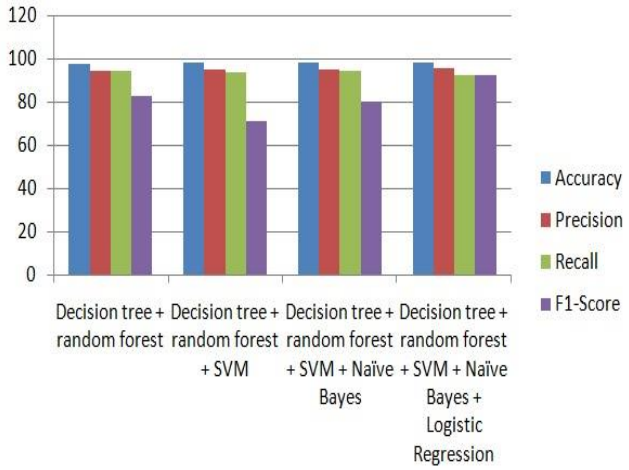

Figure 5. Overall comparison with all models

V. CONCLUSION

This work provided an overview of concepts, goods, and platforms from both the past and the present works in addition to introducing the principles of chatbots. Here, there was a thorough discussion of the current fascination with chatbots, their impact on agriculture, their benefits, and their drawbacks. Through the development of a sample chatbot, various aspects of chatbot implementation and its applications have been discussed .AI in agriculture will assist farmers in automating their operations as well as making the switch to precision farming, which will increase crop yields and improve crop quality while consuming less resources. AI based agriculture helps farmers find more methodical ways to safeguard their crop against pests and weeds as well as better ways to cultivate, harvest and forecast weather data The farmer chatbot consists of three modules: NLP for text extraction and bagging, with a solid UI for ease of comprehension,in which the chatbot would communicate with farmers and answer questions.

VI. FUTURE SCOPE

This chatbot has some limitations like real world scenario based answering, limited query responses which leads to down quality of the chatbot. Future improvements could include responding in regional languages and developing an IoT- based system for detecting temperature anomalies in agricultural fields. Future study includes how IOT components can be integrated into ensemble learning. To find IoT abnormalities using SDN, DeL-IoT uses deep ensemble learning, improving device uptime, detection effectiveness, switch-level dynamic flow control, and device status prediction.

REFERENCES

[1] Yashaswini. D. K, Hemalatha. R, Niveditha. G, “Smart Chatbot for Agriculture”, International Journal of Engineering Science and Computing, Vol.**9**, Issue.**5**, **2019.**

[2] Prashant Y. Niranjan, Vijay S. Rajpurohit, Rasika Malgi, “Development of Agriculture Chatbot using Machine Learning Techniques”, International Journal of Innovative Technology and Exploring Engineering, Vol.**9**, Issue.**2S**, pp.**24-28, 2019.**

[3] Gustavo Marques Mostaco, Icaro Ramires Costa Desouza, Leonaordo Barreto Campos, Carlos Edurardo, “AgronomoBot: a smart answering Chatbot applied to agricultural sensor networks”, Proceedings of the 14th International Conference on Precision Agriculture **2018.**

[4] R. Sandrilla, M. Savitha Devi, “A ROBUST TECHNIQUE OF FAKE NEWS IDENTIFICATION USING ENSEMBLE FEATURE SELECTION”,Indian Journal of Computer Science and Engineering (IJCSE), Vol.**12**,No.**6, 2021.**

[5] P. Sreelakshmi, Gaggara Harika, Kavya Karat, R. Madhumitha, K. Vijith, “Automated Agrobot”, Indian Journal of Science and Technology, Vol.**9,** Issue.**30, 2016.**

[6] Munira Ansari, Mohammed Saad Parbulkar, Saali Shaikh, Talha Khan, Anupam Singh, “Intelligent Chatbot”, International Journal of Engineering Research & Technology (IJERT), Special Issue – **79-82, 2021.**

[7] Brahmananda Reddy D, 2 Gurubasava “Agriculture Chatbot Application Using Python” Vol.**5**, Issue.**1, 2020.**

[8] D. Sawant, A. Jaiswal, J. Singh and P. Shah, "AgriBot - An intelligent interactive interface to assist farmers in agricultural activities," 2019 IEEE Bombay Section Signature Conference (IBSSC), pp.**1-6, 2019.**

[9] Ekanayake, J. and Saputhanthri, L. “E-AGRO: Intelligent Chat-Bot. IoT and Artificial Intelligence to Enhance Farming Industry", AGRIS on-line Papers in Economics and Informatics, Vol.**12**, No.**1**, pp.**15-21, 2020.**

[10] J. Vijayalakshmi, K. PandiMeena, “Agriculture TalkBot Using AI”, Vol.**8**, Issue-**2S5**, **2019.**

[11] L. Kannagi, Ramya C., Shreya R., Sowmiya R. “Virtual Conversational Assistant–The FARMBOT”-International Journal of Engineering Technology Science and Research- Vol.**5**, **2018.**

[12] P. Kaviya , M. Bhavyashree , M. Deepak Krishnan , M. Sugacini, “Artificial Intelligence Based Farmer Assistant Chatbot”, International Journal of Research in Engineering, Science and Management, Vol.**4**, Issue.**4**, pp.**26-29, 2021.**

[13] H Manoj T Gadiyar, Dr. Thyagaraju G S, Soubhagya K- “Online based Agriculture Monitoring System Using AI” - ICRADL – **2021.**

[14] Tran Ngoc Viet,Hoang Le Minh,Le Cong Hieu, Tong Hung Anh "THE NAÏVE BAYES ALGORITHM FOR LEARNING DATA ANALYTICS"- International Journal of Computer Science Engineering Vol.**12**, No.**4**, **2021.**

[15] Tina Elizabeth Mathew, K S Anil Kumar "A LOGISTIC REGRESSION BASED HYBRID MODEL FOR BREAST CANCER CLASSIFICATION,"International Journal of Computer Science Engineering Vol.**11** No.**6, 2020.**

[16] A. Bhattacharjee , J. Kharade, "Cluster-Then-Predict and Predictive Algorithms (Logistic Regression) "International Journal of Computer Science Engineering Vol.**6,** Issue.**2, 2018.**

[17] Dr.S. Geetha, Dr. S. Balaji, Santhiya. A, Subashri. C, and Subicsha. S - “Farm’s SmartBOT”, Vol.**12**, No.**10, 2021.**

[18] R. Gunawan, I. Taufik, E. Mulyana, O. T. Kurahman, M.A. Ramdhani and Mahmud, "Chatbot Application on Internet Of Things (IoT) to Support Smart Urban Agriculture," IEEE 5th International Conference on Wireless and Telematics (ICWT), pp.**1-6, 2019.**

AUTHORS PROFILE

Dr. Rajeshwari M Goudar is a Professor in the School of Computer Engineering department of MIT Academy of Engineering Pune. She has teaching experience of 27 years. Her research interests include areas of computer networking, network security and operating systems.

Mr. Uddhav Patil pursuing Bachelor of Technology from MIT Academy of Engineering, Pune. He has completed diploma in 2020 from the government polytechnic college. He has done various research and contributions in technology. He is going to contribute to the latest technologies like cloud, Big Data, Hadoop etc.

Mr. Nikhil Kinikar is pursuing Bachelor of Technology in Computer Engineering from MIT Academy of Engineering, Pune. Prior to that he has done a Diploma in Computer Engineering. His main research work focuses on Machine learning and Data Science fields.

Ms. Madanu Amisha pursued Diploma in Computer Engineering in 2020, currently a student and pursuing Bachelor's of Technology in the School of Computer Engineering and Technology in MIT Academy of Engineering, Pune-India. Her research interest focuses on Data Science, Artificial Intelligence and Machine Learning and Big Data

Ms. Aishwarya Chouthe has done diploma in Computer Engineering in 2020 and is currently pursuing Bachelor of Technology in Computer Engineering from MIT Academy of Engineering.

