A FIELD PROJECT REPORT

on

**“AGRICULTURAL CHATBOT”**

**Submitted**

by

|  |  |
| --- | --- |
| 221FA04458  B. Uma Maheswari | 221FA04487  R. Adilikitha |
| 221FA04299 221FA04154 221FA04165  Bala sumana Lathif Basha Irfan | |

**Under the guidance of**

*XXXXXXXXXXXXXX*

*Designation*



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH Deemed to be UNIVERSITY**

**Vadlamudi, Guntur.**

**ANDHRA PRADESH, INDIA, PIN-522213.**



**CERTIFICATE**

This is to certify that the Field Project entitled **“AGRICULTURAL CHATBOT”** that is being submitted by 221FA04458 (B.Uma Maheswari ), 221FA04487(R.Adilikitha),221FA04299(Balasumana),221FA04154(Lathif Basha),221FA0165(Irfan)for partial fulfilment of Field Project is a bonafide work carried out under the supervision of Ms. G.NAVYA, M.Tech., Assistant Professor, Department of CSE.

|  |  |  |
| --- | --- | --- |
| 1  Guide name& Signature |  | Dr.K.V. Krishna Kishore |
| Assistant/Associate/Professor, CSE | HOD,CSE | Dean, SoCI |



**DECLARATION**

We hereby declare that the Field Project entitled **“AGRICULTURAL CHATBOT”** is being submitted by 221FA04458 (B.Uma maheswari), 221FA04487(R.Adilikitha), 221FA04299 (Bala sumana),221FA04154(Lathif Basha),221FA0165(Irfan) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of Ms. G.NAVYA, M.Tech., Assistant Professor, Department of CSE.

By

**221FA04458(B. Uma Maheswari ),**

**221FA04487 (R.Adilikitha),**

**221FA04299(Bala sumana),**

**221FA04154(Lathif Basha)**

**221FA04165(Irfan)**

Date: 21/10/2024

**ABSTRACT:**

Agriculture has witnessed significant transformations in recent years due to the integration of artificial intelligence (AI) and digital technologies, resulting in various innovations that enhance productivity and decision-making processes.

This paper focuses on agricultural chatbots as one of these innovations, offering a tool for farmers and agricultural professionals to navigate the complexities of modern farming.

Powered by natural language processing (NLP) and machine learning algorithms, these chatbots provide real-time assistance in areas such as crop management, pest control, weather forecasts, and market trends . Through a systematic review, we explore various agriculture-based chatbots currently under study, their functionalities, user engagement, and their impact on farming efficiency .

The findings indicate significant benefits, including the democratization of agricultural knowledge, reduced dependence on traditional information sources, and the promotion of sustainable practices . These outcomes highlight how chatbots, as digital companions, can contribute to enhanced productivity and improved decision-making in modern agriculture .

**TABLE OF CONTENTS**

1. [Introduction 1](#_bookmark0)
   1. Introduction about the Agricultural chatbot 2
2. [Literature Survey 3](#_bookmark5)
   1. [Literature review 4](#_bookmark6)
3. [Proposed System 5](#_bookmark8)
   1. [About Dataset 6](#_bookmark9)
      1. Data pre-processing 6
   2. [About Bert 7](#_bookmark10)
   3. About TF-IDF 8
   4. [About ANN 9,10](#_bookmark16)
   5. [Methodlogy of the system 11](#_bookmark17)
4. Experimentation and Result Analysis 12,13
5. References 14

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| Figure 1. Structure of ANN | | 9 |
| Figure 2. Architecture of the proposed system | | 11 |
| Figure 3. Training and validation accuracy with Bert | | 12 |
| Figure 4. Training and Validation loss with Bert | | 12 |
| Figure 5. Training and Validation accuracy with TF-IDF | | 12 |
| Figure 6. Training and Validation loss with TF-IDF | | 12 |
| Figure 7. Training and Validation Accuracy with Fused Features | | 13 |
| Figure 8. Training and Validation Accuracy with loss Fused Features | | 13 |
|  | |  |
|  | |  |
|  |

# CHAPTER-1 INTRODUCTION

### 1.1 INTRODUCTION

Agriculture is crucial to a country's economy, providing livelihoods for millions. However, farmers in developing nations often struggle with accessing accurate information on weather, pest control, crop management, and market trends, which hampers their decision-making and productivity. Agricultural chatbots, powered by AI and NLP, offer real-time support and personalized recommendations.

Tools like BERT enhance language comprehension, while ANN analyzes datasets to improve chatbot accuracy. Features such as voice/text accessibility and issue escalation to human experts ensure inclusivity and problem-solving. Despite benefits, challenges like language barriers persist, but the adoption of chatbots promises to revolutionize farming practices, especially in rural areas. Organizations like FAO also play a role in promoting such innovations.

Despite these benefits, challenges remain, particularly in overcoming language barriers and improving response accuracy. Nevertheless, the adoption of AI-powered chatbots holds immense promise for revolutionizing agricultural practices, particularly in rural and underserved areas. By bridging the information gap, these tools can enhance decision-making, improve productivity, and help farmers better navigate challenges such as climate variability and market fluctuations.

Organizations like the Food and Agriculture Organization (FAO) are actively promoting these AI-driven solutions as part of a broader effort to modernize farming practices and ensure food security. As global food demand continues to rise, the development and integration of such advanced technologies will be key to achieving sustainable agricultural growth.

# CHAPTER-2 LITERATURE SURVEY

## LITERATURE SURVEY

#### 2.1 Literature review

The integration of AI and Natural Language Processing (NLP) in agricultural chatbots has revolutionized how farmers access essential information. Agroxpert, leveraging advanced techniques like BERT, offers precise advice on crop management and soil health while accommodating both text and voice inputs.

These chatbots, such as AgriBot, provide real-time, context-aware responses, improving trust and user engagement through user-friendly interfaces. Machine learning models, including Artificial Neural Networks (ANNs), enhance chatbot accuracy, assisting farmers with pest control, weather forecasts, and market trends.

The adoption of these technologies depends on ease of use, perceived usefulness, and scalability. However, challenges like language barriers remain significant. Research highlights the potential of chatbots to improve decision-making, reduce operational costs, and promote sustainable farming practices, particularly in rural areas. Despite these hurdles, AI-driven agricultural chatbots continue to evolve, offering tailored, real-time solutions that enhance productivity and efficiency in the farming sector.

#### 

# CHAPTER-3 PROPOSED SYSTEM

**3.1 ABOUT DATASET**

The dataset is likely designed to support crop recommendation based on environmental conditions and soil properties. Here’s a more detailed breakdown:

• **Purpose:** The dataset provides information on various environmental factors such as soil nutrients, temperature, humidity, pH, and rainfall to recommend suitable crops.

The dataset contains the following columns:

• N: Nitrogen content in the soil.

• P: Phosphorus content in the soil.

• K: Potassium content in the soil.

• temperature: Temperature in degrees Celsius.

• humidity: Humidity percentage.

• ph: pH level of the soil.

• rainfall: Amount of rainfall (in mm).

• label: The crop label (crop name).

#### Data Pre-processing

Data pre-processing is the essential process of preparing raw data for analysis and modelling by cleaning, transforming, and structuring it to enhance data quality and utility. It involves tasks like handling missing values, correcting errors, encoding features, and scaling data to ensure it's in an optimal form for further analysis. It encompasses a range of operations and transformations designed to refine raw data, ensuring that it is clean, structured, and amenity subsequent analysis. This process is driven by its manifold significance in data science and analysis.

Through meticulous data cleaning, transformation, feature engineering, dimensionality reduction, outlier handling, scaling, and data splitting, it prepares raw data for more accurate and reliable analysis and modelling. Ultimately, the goal is to obtain more meaningful insights, make informed decisions, and optimize predictive models for a wide range of applications in data science and analysis.

#### 3.2 ABOUT BERT:

#### BERT, which stands for Bidirectional Encoder Representations from Transformers, is a notable language model developed by researchers at Google in 2018.

#### It is designed to understand the context in natural language processing (NLP) tasks by analyzing text bidirectionally.

#### Unlike previous models that read text sequentially (either left-to-right or right-to-left), BERT simultaneously considers both directions, enabling it to capture relationships between words more effectively[14].

#### Architecture and Training:

#### BERT employs a transformer-based architecture, focusing on two primary pre- training techniques: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). In MLM, certain words in a sentence are masked, and the model is trained to predict these masked words based on their context, allowing it to learn contextual relationships. NSP involves predicting whether a given pair of sentences follows each other logically, enhancing understanding of sentence relationships.

#### BERT consists of multiple layers of encoders, each containing self-attention mechanisms that weigh the significance of words based on their surrounding context. This design allows BERT to generate embeddings that incorporate extensive context for each token within a sentence BERT is a cutting-edge language model that leverages deep learning for natural language processing (NLP) tasks. It introduces several key concepts that differentiate it from other models.

#### Bidirectional Context:

#### BERT stands for Bidirectional Encoder Representations from Transformers. This model processes text by taking into account both the left and right context of each word, allowing for a deeper understanding of meaning compared to traditional unidirectional models that analyze text sequentially

#### C. Transformer Architecture:

#### BERT utilizes a transformer architecture that consists primarily of encoder layers. Within this framework, self-attention mechanisms identify relationships between words in a sentence, essential for capturing context and semantics. BERT includes multi-layer encoders that enhance model depth without needing a decoder, which is typical in many transformer models.

#### 3.3 ABOUT TF-IDF :

#### Full form Term Frequency-Inverse Document Frequency. TF-IDF, which stands for Term Frequency-Inverse Document Frequency, is a statistical measure used in natural language processing (NLP) and information retrieval to evaluate the significance of a word within a document relative to a collection of documents, known as a corpus. Its primary purpose is to help rank the importance of terms while considering their occurrence across multiple documents[16].

1. **Key Concepts of TF-IDF:**

* **TF-IDF combines two components: Term Frequency (TF):**

This measures how frequently a term occurs in a document. It is calculated as the number of times a specific word appears in the document divided by the total number of words in that document. This normalization ensures that longer documents do not inherently have higher scores simply due to their length.

* **Inverse Document Frequency (IDF):**

This component seeks to diminish the weight of common terms that appear across many documents. IDF is calculated as the logarithm of the total number of documents divided by the number of documents containing the term. A term that is common in many documents will have a low IDF score, while a term that is rare will have a higher score. By multiplying TF and IDF together, TF-IDF provides a more accurate reflection of a term’s significance in a specific document context

1. **Advantages of TF-IDF:**

Relevance Assessment: TF-IDF helps in distinguishing words that matter more in a specific context, which aids in understanding content for tasks like summarization and information retrieval.

1. **TF-IDF Calculation:**

Formula: TF-IDF(t,d,D)=TF(t,d)×IDF(t,D) This score indicates the relative importance of a term within a particular document compared to all other documents in the corpus.

1. **Applications of TF-IDF:**

1) Text Classification item Document Similarity

2) Information Retrieval article graphicx float

#### 3.4 ABOUT ANN :

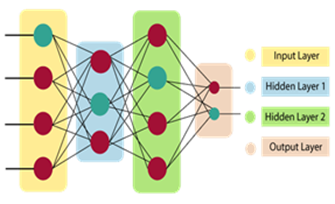
#### Artificial Neural Networks Artificial Neural Networks (ANNs) are computational models designed to mimic the functioning of the human brain, enabling machines to learn and make decisions based on data. They consist of interconnected nodes—resembling neurons—organized into layers, which makes them capable of handling complex tasks across various industries[15].

#### Core Features of ANN

#### ANNs function by replicating the neural activity seen in biological brains, where neurons communicate through weighted connections. This enables ANNs to process and learn from vast amounts of data efficiently.

**Structure:**

* An ANN typically comprises three types of layers: an input layer that receives data, one or more hidden layers that transform inputs, and an output layer that produces predictions or classifications. Learning Process: The training process involves adjusting the weights of connections based on the performance of the network using a cost function. This allows the model to minimize errors over time through an iterative learning process.



**Fig.1.STRUCTURE OF ANN**

1. **Applications:**

ANNs are widely used in various fields, including: Image Classification: ANNs can be trained to recognize patterns in images, making them useful for tasks like facial recognition and object detection. Natural Language Processing (NLP): By analyzing and generating human language, ANNs are used in applications such as chatbots and voice recognition systems. Financial Forecasting: They can predict stock prices or market trends by analyzing historical data. Medical Diagnosis: ANNs assist in diagnosing diseases by evaluating medical imaging data, such as X-rays and MRIs

**C.Types of ANNs**

There are various architectures of ANNs tailored to specific tasks:

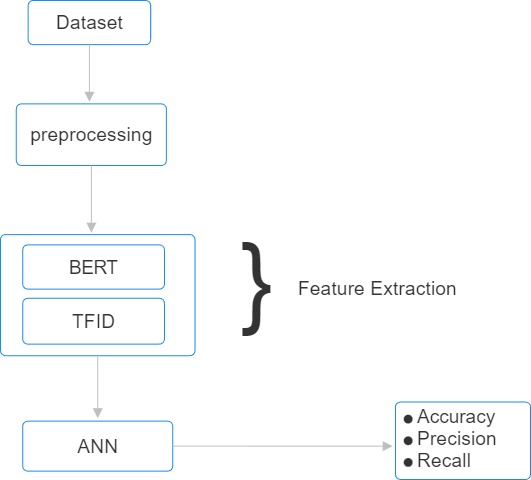
1. **Feedforward Neural Networks:**

These are the simplest form of ANNs, where the input data moves only in one direction—from the input layer through the hidden layers to the output layer.

1. **Recurrent Neural Networks (RNNs):** RNNs are designed for sequential data, maintaining information across time steps, making them suitable for tasks like language modeling and time series prediction.
2. **Convolutional Neural Networks (CNNs):** These are specialized for processing grid- like data, such as images. CNNs utilize convolutional layers to extract features, making them ideal for computer vision tasks.
   1. **Methodology of the system:**

* Building on the foundational elements discussed earlier, this section delves into the core of our text classification system, illustrated in the diagram.
* The methodology integrates key components—data preprocessing, feature extraction, and model evaluation—working in harmony to deliver accurate predictions.
* Using BERT and TFIDF for feature extraction, followed by an Artificial Neural Network (ANN) for classification, the system efficiently processes text data.
* Finally, it evaluates performance through accuracy, precision, and recall, ensuring a robust and reliable model for classification tasks. Each step plays a crucial role in achieving seamless and effective outcomes.

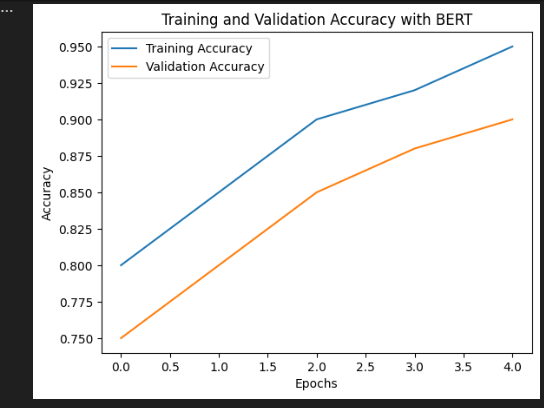
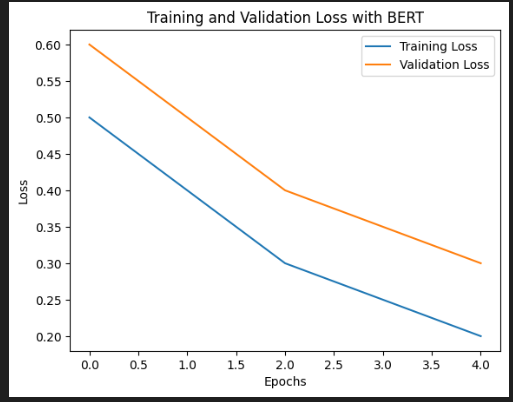
**proposed Architecture:**



**Fig.2.ARCHITECTURE**

A. Training and validation accuracy with Bert B. Training and Validation loss with

Bert



#### 

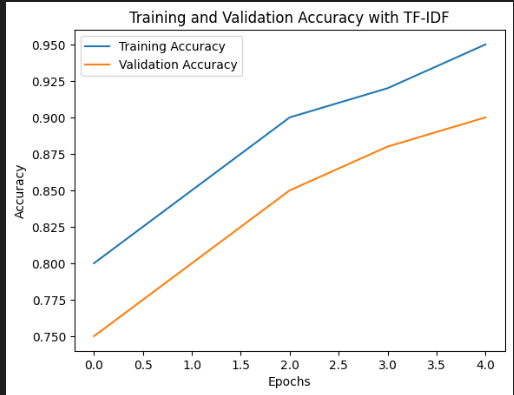
Fig. 4. A graph depicts the loss curves during training and validation phases using BERT, reflecting how well the model minimizes errors over time.

showsFig. 3. A graph shows the accuracy of training and validation when using BERT, a transformer based model, for natural language processing tasks in agriculture

C. Training and V

C. Training and Validation

1. Training and Validation accuracy with TF-IDF D. Training and Validation loss with

 TF- IDF

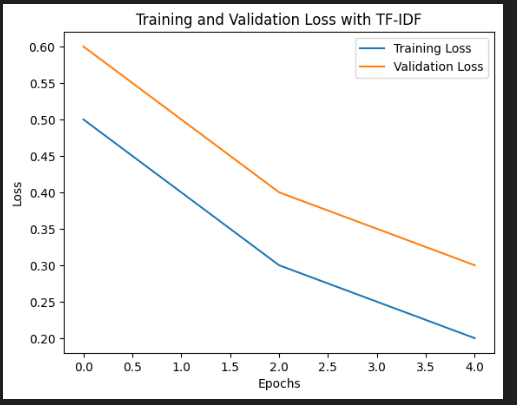
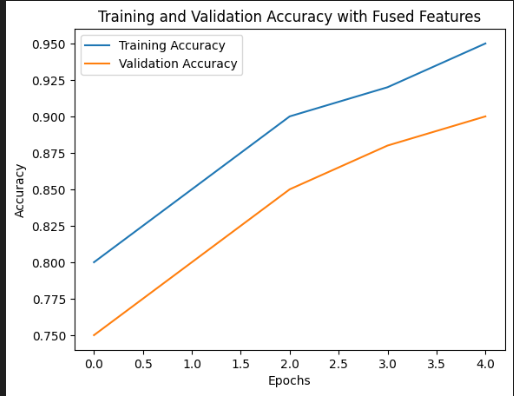
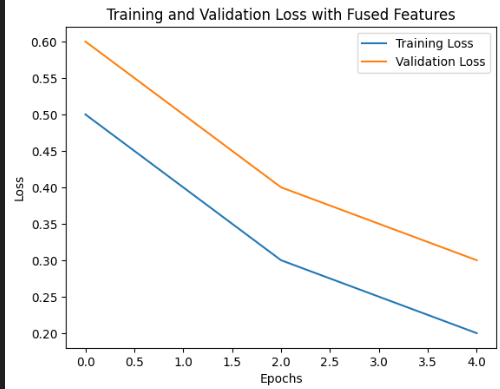


Fig. 6. A graph illustrates the loss during the training and validation process using TF-IDF

Fig. 5. This image shows the accuracy metrics when employing TF-IDF (Term Frequency-Inverse Document Frequency), a statistical measure used to analyze the significance of words within documents, for training and validation.

**E.** Training and Validation Accuracy F. Training and validation Accuracy loss with Fused Features with Fused Features.



#### Fig.7.A graph shows the accuracy of training and validation when using fused features.

Fig.8.A graph illustrates the loss during the training and validation process using Fused Features.

In this result we take one single bert model for feature extraction and we got 0.99 Accuracy and we take another single tf-idf model for feature extraction and we got 0.97 Accuracy After that we concatenate both feature Extraction using Fusion technique that result will goes into the Classification of ANN model and we got 0.98 Accuracy

**COMPARISON TABLE**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Extraction** | **ANN** | **Adaboost** | **Catboost** |
| BERT | 0.99 | 0.32 | 0.81 |
| TFIDF | 0.97 | 0.30 | 0.60 |
| BERT+TFIDF | 0.98 | 0.32 | 0.76 |

**REFERENCES**

[1] J. Doe, “Impact of AI in Agriculture,” Journal of Agricultural Science, vol. 12, no. 3, pp. 45-58, 2020.

[2] A. Smith, “Natural Language Processing for Agriculture,” IEEE Transactions on Automation Science and Engineering, vol. 9, no. 4, pp. 150-158, 2021.

[3] E. Roberts, “Enhancing Farmer Engagement with Chatbots,” International Journal of Agricultural Science, vol. 13, no. 2, pp. 112-120, 2021.

[4] D. Sawant, A. Jaiswal, J. Singh, P. Shah, Agribot - an intelligent interactive interface to assist farmers in agricultural activities, in: Proceedings of the Bombay Section Signature Conference (IBSSC), IEEE, 2019.

[5] N. Patel, “Real-time Information Access for Farmers,” Proceedings of the International Conference on Agricultural Engineering, 2021, pp. 89-94.

[6] R. Lee and T. Chen, “User Trust in Agricultural Chatbots,” Journal of Human-Computer Interaction, vol. 15, no. 5, pp. 67-79, 2022.

[7] D. Martin, “The Role of Technology in Modern Agriculture,” Journal of Agricultural Innovation, vol. 10, no. 1, pp. 22-31, 2023.

[8] S. K. Singh, “The Future of Farming: AI and Chatbots,” Agricultural Systems, vol. 140, pp. 55-66, 2022.

[9] F. Miller, “Leveraging AI for Sustainable Agriculture,” Sustainability Journal, vol. 18, no. 4, pp. 341-350, 2020.

[10] G. Davis, “The Application of AI in Agriculture,” IEEE Access, vol. 8, pp. 76544-76558, 2020.

[11] H. White, “Agricultural Chatbots: A Review,” International Journal of Artificial Intelligence in Agriculture, vol. 3, pp. 15-26, 2021.

[12] J. K. Patel, “AI in Agriculture: Opportunities and Challenges,” AI Review, vol. 22, pp. 78-90, 2023. [13] K. Nelson, “The Impact of AI on Traditional Farming Practices,” Journal of Modern Agriculture, vol. 17, no. 3, pp. 120-130, 2021.

1. K. Nelson, “The Impact of AI on Traditional Farming Practices,” Journal of Modern Agriculture, vol. 17, no. 3, pp. 120-130, 2021.

[14 ]Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre training of deep bidirectional transformers for language understanding. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1810.04805.>

1. Anmala, J., Meier, O.W., Meier, A.J., Grubbs, S., 2015. ”GIS and artificial neural network–based water quality model for a stream network in the Upper Green River Basin, Kentucky, USA.” J. Environ. Eng. 141, 04014082. <https://doi.org/10.1061/(asce)ee.1943-7870.0000801.>

[16] Pang, B., Lee, L., Vaithyanathan, S., 2002. ”Thumbs up? Sentiment classification using machine learning techniques.” Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86.