AI-Powered Conversational Schemes Navigator using Natural Language processing for Tamil Nadu Government Schemes

Mrs.J.P.ASWINI,
Assistant Professor , Computer Science and Engineering,
St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,
Email Id -aswinianchu1997@gmail.com

MR.M. JAGADEESH,
Student, Computer Science and Engineering,
St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,
Email Id – jagadeeshdts@gmail.com

MR. S. VASANTHKUMAR,
Student, Computer Science and Engineering,
St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,
Email Id – smvasanth04@gmail.com

MR .G. DEEPAKKRISHNAN ,
Student, Computer Science and Engineering,
St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,
Email Id –tamil761978@gmail.com

Abstract - Government schemes encompass diverse areas such as education, healthcare, agriculture, social welfare, and infrastructure. However, a lack of awareness and difficulty in accessing accurate information often prevents individuals from availing of these benefits. To address these challenges, this project proposes the development of a Natural Language Processing (NLP)-based Chat Bot designed to provide seamless access to information about Tamil Nadu government schemes. The Chat Bot leverages advanced NLP frameworks, such as spaCy and Hugging Face Transformers, to process and interpret user queries, delivering precise and relevant responses. Comprehensive data on government schemes is collected, preprocessed, and used to train the model. Integrated into a user-friendly interface, the Chat Bot ensures effortless interaction, allowing users to inquire about various initiatives and obtain real-time information. Additional features include user authentication for personalized assistance and provisions for human support to handle complex queries. By offering an intelligent conversational interface, this project aims to enhance accessibility and engagement, empowering citizens to make informed decisions and utilize available government resources effectively.

I. INTRODUCTION

Government welfare schemes play a crucial role in improving the socio-economic conditions of

citizens, particularly in a diverse and populous state like Tamil Nadu. These schemes cover various domains, including education, healthcare, financial assistance, employment, and social security. However, despite the availability of numerous beneficial programs, many eligible beneficiaries fail to access them due to multiple challenges. Accessing government schemes in Tamil Nadu remains a significant challenge for many citizens due to several barriers. These include complex eligibility criteria, a lack of awareness about available programs, and difficulties in understanding official information due to language constraints. Many individuals, especially those from rural areas and marginalized communities, struggle to find schemes that align with their specific needs, leading to underutilization of essential welfare benefits. To address these challenges, this project introduces an AI-Powered Conversational Schemes Navigator, a system designed to simplify the process of discovering and understanding government schemes. By integrating Natural Language Processing (NLP) and machine learning, the system enables users to interact seamlessly in both Tamil and English, ensuring accessibility for a wider audience.

The chatbot-based interface allows users to ask questions in natural language, receive real-time responses, and explore relevant welfare programs based on their personal circumstances. It leverages structured databases and AI-driven recommendation algorithms to match users with the most suitable schemes. By providing instant, accurate, and personalized information, this system enhances transparency in public services and ensures that government benefits reach the people who need them the most. These schemes cover various domains, including education, healthcare, financial assistance, employment, and social security. However, despite the availability of numerous beneficial programs, many eligible beneficiaries fail to access them due to multiple challenges

This AI-driven approach not only reduces dependency on intermediaries but also empowers citizens by enabling self-service access to government welfare programs. The system serves as a bridge between the government and its beneficiaries, fostering inclusivity, efficiency, and ease of access in public service delivery.

II. BACKGROUND AND MOTIVATION

Tamil Nadu, a state with a population of over 70 million people, is known for its rich cultural diversity, and the majority of the population speaks Tamil. However, in addition to Tamil, various other languages such as English, Telugu, Hindi, and Kannada are also spoken across the state. This linguistic diversity presents a significant challenge in providing equal access to government welfare schemes and public services for all residents, particularly those in rural areas, who may not be familiar with more widely used languages like English.The Government of Tamil Nadu offers a wide variety of schemes aimed at improving the socio-economic status of citizens. These schemes span several sectors, including healthcare, education, agriculture, employment, social welfare, and housing. However, the awareness and uptake of these schemes have been hindered by a lack of easy-to-access information, especially for individuals who have limited proficiency in Tamil or English or those in remote areas with low literacy rates.While digital platforms have been launched to improve awareness, these are often difficult to navigate, particularly for people who are not comfortable with technology or do not have easy access to the internet. The key to bridging

this gap lies in simplifying the process of accessing scheme-related information and offering a more inclusive and interactive interface for users.

Natural Language Processing (NLP) and AI-powered conversational agents can provide a powerful solution to this problem. By combining AI and NLP, an intuitive, conversational scheme navigator can help users discover, understand, and apply for government schemes by interacting in their preferred language—whether it be Tamil, English, or other regional languages

a. Language Barriers

Tamil Nadu is a multilingual state, and a significant portion of the population does not speak Tamil fluently. Many individuals, especially in rural areas, may speak only their local dialects and may not have the literacy skills to navigate complex government websites, which are often only available in Tamil or English. An AI-powered conversational interface would allow users to ask questions in their own language and receive answers in real-time, helping to overcome the language divide.

b. Lack of Awareness of Government Schemes

Despite the **abundance of government schemes** aimed at improving the lives of citizens, many people remain unaware of their eligibility or even the existence of these schemes. This lack of awareness often leads to under-utilization of available resources and missed opportunities for citizens, particularly those who need them the most.

III. NATURAL LANGUAGE PROCESSING FOR MULTILINGUAL SUPPORT

Natural Language Processing (NLP) for multilingual support is a critical aspect of building systems that can understand, process, and generate human language across different languages. In the context of government schemes in Tamil Nadu (or any region), NLP for multilingual support ensures that the system can serve a diverse set of users who speak different languages.

a. Multilingual Data Collection and Preprocessing

Corpus Creation: To effectively support multiple languages, a diverse and representative corpus of text in all supported languages is necessary. This includes government scheme details, user queries, FAQs, and other relevant documents in Tamil, English, and other local languages.

Text Preprocessing: Each language has its own preprocessing requirements, including:

Tokenization: Breaking down sentences into words or meaningful units (tokens). For Tamil, this means dealing with word boundaries and compound words.

Stopword Removal: Common words (like "the", "is", "and") that do not contribute significant meaning must be removed.

Normalization: Handling issues like different spellings, abbreviations, and slang used in informal language.

b. Language Detection and Switching

A multilingual NLP system should first detect the user's language to route the query to the appropriate processing pipeline. This is often accomplished with language detection models trained on a wide variety of text samples in each language.

If a user switches languages during an interaction, the system must dynamically switch between language models to interpret and respond accurately. For example, if a user starts in Tamil and switches to English, the system must be capable of handling both languages simultaneously.

c. Machine Translation (MT)

For effective multilingual NLP, a machine translation model is often used to translate text from one language to another. For instance, if a user asks a question in Tamil, the system could translate the query into English to leverage a more robust NLP model, then translate the answer back into Tamil.

Popular neural machine translation (NMT) models like Google Translate and OpenNMT can be trained to handle specific regional languages, like Tamil. The translation models must be highly accurate, as any errors could result in misunderstandings of government schemes and eligibility criteria.

d. Cross-Lingual Information Retrieval

The system may retrieve information about schemes in one language but present it in another. For example, government schemes might be detailed in English, but a user queries in Tamil. NLP models must handle cross-lingual retrieval, where the query is processed in Tamil, translated into English for the search, and then translated back into Tamil for the user's response.

IV. PERSONALIZED SCHEME RECOMMENDATION USING MACHINE LEARNING

3.1 User Profiling for Recommendations

The chatbot collects basic user details such as:

(Age, gender, occupation, and income level.,Past queries and interactions.,Preferences (e.g., education, healthcare, employment)).Using machine learning algorithms like Decision Trees, Random Forest, and Neural Networks, the system personalizes scheme recommendations.

3.2 Context-Aware Recommendations

If a user asks about agricultural subsidies, the chatbot can link related schemes like irrigation loans and crop insurance, ensuring comprehensive information.

2. Components of Personalized Scheme Recommendation System

To build an AI-powered personalized scheme recommendation system, the following components are essential:

a. Data Collection

The system would require various forms of user data and scheme data:User Data:Demographics (age, gender, income, occupation, region, etc.)

Personal Interests (e.g., agriculture, education, healthcare, entrepreneurship)

Past Interactions (previously used schemes, application history)

Behavioral Data (preferences, searches, clicks, etc.)

Scheme Data:

Information about various government schemes available in Tamil Nadu, including:(Scheme Name Eligibility Criteria, Benefits, Application Process ,Duration, Target Audience (e.g., farmers, students, women, senior citizens) This data can be extracted from government portals or databases and stored for easy querying.

b. Preprocessing Data

Once the data is collected, it needs to be preprocessed: User Data Normalization: Ensuring that all user data (e.g., income range, region) is presented in a uniform format.

Scheme Data Standardization: Making sure that the information about schemes (e.g., eligibility criteria) is in a machine-readable format, possibly tagging or categorizing data into buckets like “income range”, “target group”, “region”, etc.

c. Feature Engineering

For effective machine learning models, it's important to extract features that help in prediction. Some features could be:User Demographics: Age, income, region, etc.

Scheme Eligibility Criteria: Specific conditions such as age range, income range, target audience (e.g., schemes for farmers, schemes for students).Interaction Data: User's past interactions with the system (e.g., schemes they’ve inquired about or applied for before).

V. INFORMATION RETRIEVAL AND KNOWLEDGE GRAPH INTEGRATION

1. Information Retrieval in the Context of Government Schemes

Information Retrieval (IR) refers to the process of obtaining relevant data from a large collection based on a user’s query. In the context of an AI-powered conversational schemes

navigator for Tamil Nadu Government schemes, information retrieval plays a pivotal role in quickly and accurately retrieving the most relevant scheme details based on user queries.

To make an AI-powered conversational system effective, it must be able to answer questions in natural language about the wide variety of government schemes, their eligibility criteria, application processes, benefits, and documentation. These queries may range from simple questions (e.g., "What are the health schemes for senior citizens?") to more complex queries (e.g., "Can I apply for the farmer's subsidy scheme if my land is located in a flood-prone area?").

Information Retrieval Techniques

Several techniques can be utilized to improve information retrieval in this system:

1. Keyword-Based Search: Traditional keyword-based search engines match user queries to documents containing relevant keywords. However, this approach may fail to consider semantic meaning, synonyms, or variations in query phrasing.
2. Semantic Search: To improve accuracy, semantic search methods can be employed. These methods focus on the meaning of words in the context of a query, not just the keywords. For example, if a user queries, "What schemes support students for higher education?" the system should be able to recognize that "higher education" may refer to both undergraduate and postgraduate courses, and thus, provide results for both.
3. Natural Language Processing (NLP) Techniques: NLP models can be used to enhance information retrieval by extracting the key entities and relationships from both the user's query and the documents containing scheme information. For example, named entity recognition (NER) can identify specific terms like "students," "farmers," or "elderly" in the user's query, enabling the system to focus on the most relevant information.
4. Contextual Search: Leveraging NLP-based models like BERT or GPT can improve query interpretation by considering the context of the user's request, leading to more accurate responses even if the user phrased their question in an ambiguous or informal manner.
5. Ranking and Relevance: Once relevant documents or pieces of information are retrieved, an additional layer of ranking is required to prioritize the most relevant results based on the user's query intent. This ensures that the top results are the ones that most closely match what the user is asking for.

VI. VOICE-ENABLED QUERY SYSTEM FOR ILLERATE USERS

A significant portion of Tamil Nadu's rural population struggles with literacy, making it difficult for them to access online government schemes. Traditional text-based searches are ineffective for such users, creating a digital divide that prevents them from benefiting from available welfare programs.

To address this, the AI-powered chatbot is equipped with a voice-enabled query system that allows users to speak their queries in Tamil instead of typing. The system then processes the spoken input using Speech-to-Text (STT) technology, retrieves relevant scheme information, and responds through Text-to-Speech (TTS) in Tamil. This ensures seamless accessibility for users with limited reading and writing abilities.

6.1 Speech Recognition and Processing

The voice-enabled query system involves the following key components:

1.Speech-to-Text (STT) Conversion

The user speaks a query into the chatbot (e.g., *"What are the schemes for farmers?"* in Tamil).The system uses Google Speech API, Mozilla DeepSpeech, or Wav2Vec to convert the spoken Tamil words into text.The converted text is processed using Natural Language Processing (NLP) techniques to understand user intent.

2.Query Interpretation and Scheme Retrieval

The chatbot analyzes the query using Named Entity Recognition (NER) to identify keywords (e.g., *"farmer," "loan," "subsidy"*) and fetches relevant scheme details from the database.

3.Text-to-Speech (TTS) Response Generation

Instead of displaying text, the chatbot converts the response into speech using TTS models like Tacotron 2, Festival Speech Synthesis, or Google Cloud TTS.The system speaks out the scheme details in Tamil, enabling users to understand without needing to read.

VII. CONCLUSION

In conclusion, the project has successfully developed a comprehensive platform for accessing Tamil Nadu government scheme information, catering to the needs of both administrators and users. By leveraging technologies such as Python, Flask, MySQL, TensorFlow, and Bootstrap, the system offers an intuitive and efficient interface for users to explore and benefit from various government initiatives. The TN Schemes Web App provides easy access to scheme details through a user-friendly interface, allowing users to search, filter, and understand schemes based on keywords or eligibility criteria. Secure authentication ensures personalized access, while feedback channels and notifications facilitate user engagement and keep users informed about updates. The integration of the TN Schemes Bot Chat Window enables real-time interaction with the chatbot, further enhancing accessibility and user experience. Users can input queries related to government schemes and receive accurate responses promptly, ensuring efficient information retrieval. The development of the Scheme Net Model and its integration with the system enables personalized scheme recommendations based on user

preferences and eligibility criteria, enhancing the utility and effectiveness of the platform. Overall, the project has successfully addressed the need for a centralized platform to access Tamil Nadu government scheme information, providing valuable insights, enhancing accessibility, and empowering citizens to benefit from government initiatives effectively.

VIII. REFERENCE

1.Zhang, Y., & Wallace, B. (2017). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820.

2.Goyal, P., Gupta, R., & Goyal, L. M. (2020). A review of chatbot and natural language processing. International Journal of Advanced Research in Computer Science, 11(4), 69-75.

3.Rashid, S. M., Abdullah, A. H., & Ahmed, M. A. (2019). Development of a chatbot using natural language processing for customer service. International Journal of Computer Science and Information Security (IJCSIS), 17(5), 167.

4.Lowe, R., & Pow, N. (2017). The rise of the conversational interface: A new kid on the block. Computer, 50(8), 58-63.

5.Rajabi, A., Asgarian, A., & Ebrahimi, M. (2018). A comparative study of machine learning algorithms for automated response selection in chatbot systems. In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 45-52).

6.Singh, A., & Sharma, M. (2020). AI Chatbot: A review of literature. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 23-28). IEEE.

7.Saini, V., & Singh, S. (2019). A review on chatbots in customer service industry. In 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 313-317). IEEE.

8.Hernandez-Mendez, A., Perez-Meana, H., & Sucar, L. E. (2018). Natural language processing and chatbots: A survey of current research and future possibilities. Journal of Computing and Information Technology, 26(1), 1-18.

9.Debnath, B., Chakraborty, D., & Mandal, S. K. (2019). Chatbot for e-learning: A review. In Proceedings of the 2nd International Conference on Inventive Research in Computing Applications (pp. 186-190). IEEE.

10.Gao, W., & Huang, H. (2019). An intelligent chatbot system for online customer service. In Proceedings of the 2019 2nd International Conference on Education and Multimedia Technology (pp. 208-211). ACM.

11.Sarker, S., & Rana, S. (2020). AI based chatbot for customer service: A review. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1774-1778). IEEE.

12.Muduli, S., & Sharma, S. (2021). Implementation of a conversational chatbot system for e-commerce. In Intelligent Computing, Information and Control Systems (pp. 753-760). Springer.

13.Ahmad, M., Kamal, A., & Shahzad, W. (2019). A review of chatbots in customer service. In 2019 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE.

14.H. Jin and H. Kim, "Developing a Chatbot Service Model for Customer Support," in International Journal of Human-Computer Interaction, vol. 36, no. 12, pp. 1188-1195, 2020.

15.J. R. Lloyd and C. A. Boyd, "The Application of Chatbots in Learning Environments: A Review of Recent Research," in Journal of Educational Technology Development and Exchange, vol. 13, no. 1, pp. 1-14, 2020.