

# HEALTHCARE ASSISTANT CHATBOT IN NEPALI LANGUAGE



By

**Sujan Poudel**

Masters of Technology in Information Technology

Kathmandu University 2020

**A Project Proposal**

Submitted to the Department of Computer Science and Engineering

Of the

Kathmandu University

March, 2022

## ABSTRACT

The advancement of technology provides opportunities for growth for several domains. The use of Artificial Intelligence in almost all domains is increasing day by day because of its wide range of capabilities. The researchers are extending the use of AI in several domains for its improvement. Among them, the healthcare domain is such a domain where most of the research happens due to it being directly related to human life. Healthcare is the most precious thing in human life. However, there are still several problems, especially in developing countries like Nepal which can be addressed with the help of these advanced technologies. According to census 2021, one-third of the Nepalese population live in rural areas which means they have weak connections with hospitals and doctors. The main problem is that they have to walk long distances for basic health checkups. This project aims to build a chatbot that can respond after a user inputs symptoms which can address problems of not only people of rural areas but people everywhere for initial diagnosis of diseases. This chatbot will use a subset of AI called Natural Language Processing (NLP) and supports roman Nepali Unicode and the Nepali language.

**Keywords:** Healthcare, Natural Language Processing, Natural Language Understanding, Chatbot, TF-IDF, Cosine Similarity

## **TABLE OF CONTENTS**

<b>ABSTRACT</b>	<b>i</b>
<b>LIST OF FIGURES</b>	<b>iii</b>
<b>CHAPTER-1</b>	<b>1</b>
<b>INTRODUCTION</b>	<b>1</b>
1.1. Background	1
1.2. Motivations	2
1.3. Problem Statements	2
1.4. Project Objectives	3
<b>CHAPTER-2</b>	<b>4</b>
<b>LITERATURE REVIEW</b>	<b>4</b>
<b>CHAPTER-3</b>	<b>6</b>
<b>METHODOLOGY</b>	<b>6</b>
3.1. Natural Language Processing	6
3.2. Response Generation	7
3.3. Knowledge Base Creation	7
3.4. Dialogue Management	8
<b>CHAPTER-4</b>	<b>9</b>
<b>EXPECTED RESULTS</b>	<b>9</b>
<b>APPENDIX A: GANTT CHART</b>	<b>10</b>
<b>REFERENCES</b>	<b>11</b>

## LIST OF FIGURES

Figure 3.1: Chatbot Architecture	6
Figure 3.2: Output layer flow	8

# **CHAPTER-1**

## **INTRODUCTION**

### **1.1. Background**

The adoption of different technology is advancing in several domains as day passes. The Healthcare domain is such a domain where a variety of technologies are implemented to achieve convenient health support. In the domain of healthcare, health care assistants have huge scope. Health care assistants are built by implementing several technologies such as website and mobile application based assistants where the response is generated from servers. The chatbot is another emerging technology that is assisting in online health assistant services. The word chatbot is derived from two words 'chat' and 'robot' which is a computer program that simulates human-like conversations. The use of chatbots is exponentially increasing in several domains such as healthcare, e-commerce, finance, hospitality and so on. The main reason behind the growing adoption of chatbot technology in various fields is due to its effectiveness. The chatbot is an online human-computer dialogue system with natural language (Jia, 2003). The chatbot facilitates the use of natural language to perform tasks. Generally, there are two kinds of chatbots which are rule-based chatbots and NLP based chatbots. The rule-based chatbots are based on a series of predefined patterns and they make decisions based on the decision tree. However, NLP based chatbots are Artificial Intelligence (AI) powered chatbots that apply AI programs to help chatbots analyse and understand natural human language to communicate with users. So, the increasing popularity and capability of Natural Language Processing (NLP) unfold yet another possibility that is AI-powered chatbots. There are several health assistant chatbots available however most of them are available in the English language. The proposed chatbot will support conversation in the Nepali and English languages. So, the proposed chatbot will basically be helpful to initially diagnose diseases in rural areas of Nepal.

This chatbot is proposed in the Nepali language to provide health assistance facilities for the initial diagnosis of diseases according to symptoms provided by users. Those input symptoms will be taken as input and perform NLU operations to identify the intent of the users. As per intent, a response will be generated to provide results as output to the users. For response

generation, two different models will be experimented namely: Seq2Seq Model and Reinforcement Learning with Seq2Seq.

## **1.2. Motivations**

Online healthcare assistants are becoming more famous and have the ability to address healthcare problems. In developing countries like Nepal, there is a huge problem of healthcare, especially in rural areas. 33.9% of the total population of Nepal are living in rural areas (*Key Highlights from the Census Report 2021, 2022*). The use of websites and mobile platform based healthcare assistants lies behind due to its less user-friendly environment and it takes comparatively more response time. The main reason behind people prefer to use chatbots rather than websites and mobile application based systems for healthcare is because of its improved user experience. There are other several advantages of using chatbots over those websites and applications systems for online healthcare assistants which are presented below:

- i) More than one-third of the population still live in rural areas hence they have poor access to hospitals and medical professionals. So, the platform which has the ability to perform basic disease diagnosis by providing symptoms and getting information about the disease will help to get further treatment.
- ii) Usually, users prefer chat because it is easier to use with a convenient environment and fast response time.

## **1.3. Problem Statements**

The key benefit of using chatbots in healthcare assistants using NLP is it allows the use of natural language. There are several built-in libraries available for the English language however the main challenge of adopting NLP based chatbots for healthcare assistants systems in Nepal is an implementation of the Nepali language. The problem statement of this project are pointed out below:

- i) There are no datasets in the Nepali language for healthcare for building chatbots till now.

- ii) Deep learning for response generation in the Nepali language.

### **1.4. Project Objectives**

The main objective of this project is to build a healthcare assistant chatbot that supports English and Nepali languages. The implementation of a chatbot for healthcare assistants can enhance user satisfaction and provide initial health information to people according to symptoms. Some of the other objectives of this project are:

- i) To create a Chatbot that can take input from users and also replies.
- ii) To develop a corpus for training the chatbot and with the learning feature.
- iii) Improved user experience by providing a human-like conversational feel on virtual healthcare assistants.

The structure of this proposal is organised as follows: section 2 consists of a literature review, section 3 discusses the methodology of the proposed work and section 3 discusses expected results.

## CHAPTER-2

### LITERATURE REVIEW

Behera (2016) proposed a semi-automatic NLP based chatbot called *Chappie* to address business requirements efficiently and automatically. First, it uses NLP to analyse chats and extract the intents of the users. Once the user provides input, the intent of the user is categorised by using Naive Bayes classifier on the basis of service offered by the company such as food, travel, utilities etc. and also generate sub-categories for user input. Then, information and Artificial Intelligence Markup Language (AIML) are used to make conversation with the users. Initially, Chappie is trained with a manually labelled corpus of 4000 words and then filtered unnecessary sentences and reduced corpus to 1500 sentences which help Chappie to improve accuracy. They implemented multiple Naive Bayes and Support Vector Machines (SVM) to train corpus. Although this chatbot is intended to provide effective business solutions, the semi-automatic nature of this chatbot raises the response time of the chatbot.

Athota et al.(2020) proposed a chatbot for medical using Artificial intelligence that can be used for disease diagnosis and provide basic details about the disease before consulting a doctor. First, clients can send questions to the chatbot through text. Secondly, this chatbot application performs text preprocessing operations such as tokenization, stop word removal and then n-gram, TF-IDF and cosine similarity are used for ranking and sentence similarity calculations which is also termed feature extraction. Finally, the answers to the query are provided to the users. This paper used a third party expert system to handle user queries that are not presented in the database however the details of how those third party expert systems work are not mentioned in the paper. Similarly, the datasets used for building chatbot are not presented in the paper and then it switches to manual mode.

Deepika et al. (2020) developed a generative based chatbot named *Jollity* in telegram using botfather and ngrok to assist depressed people. This chatbot is built using an open-source conversational AI framework called Rasa. Unlike previous work, this chatbot implements Natural Language Understanding (NLU) to fix spelling like mistakes and helps to provide exact results to the user. The experimental results show that this chatbot has 90% accuracy. This chatbot is like a virtual friend to depressed people where depressed people can get motivational



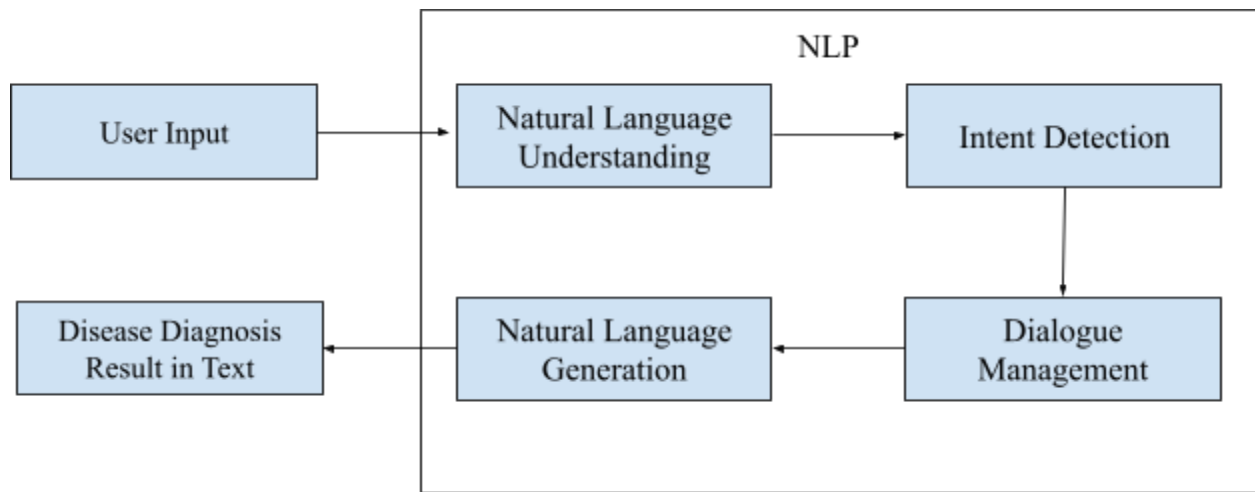
videos, articles and other materials to help them to get rid of stress. However, this chatbot repeatedly suggests resources and if that resource doesn't work in all cases.

Badlani et al. (2021) proposed a solution that describes multilingual healthcare ( i.e. English, Hindi and Gujarati) chatbot based system to help early diagnosis and treatment of diseases. User queries are responded to by calculating sentence similarity by using TF-IDF and Cosine Similarity. This chatbot also supports speech to text and text to speech conversion. The use SpeechRecognition library of python for speech to text conversion for English language and in the case of other languages Googletrans python library has been used for language translation. Similarly, Googletrans python library is used to convert the final response to user preferred language in text mode and for voice communication mode they employed Google Text to Speech (gTTS) and play sound libraries to produce output. They experimented with 4920 records of a dataset using 10-fold Cross-Validation in five different machine learning algorithms such as Random Forest Classifier, K-Nearest Neighbors (KNN), Support Vector Machine, Decision Tree and Multinomial Naive Bayes. Among them, they get higher accuracy i.e 98.43% from Random Forest Classifier. This chatbot consists of Automatic Speech Recognition (ASR) and Text-to-Speech features however those features are employed using third party libraries. Although this model gets high accuracy there is room for improvement using deep learning algorithms.

## CHAPTER-3

### METHODOLOGY

The chatbot for healthcare assistants can be useful to perform initial diagnosis of diseases from symptoms. The NLP based chatbot improves performance more than rule-based chatbot. The system design for the proposed work is shown in Figure 3.1.



**Figure 3.1:** Chatbot Architecture

Generally, chatbots are built for specific purposes in several domains. Each text-based chatbot performs based on four components: Natural Language Processing, Response Generation, Knowledge Base Creation and Dialogue Management (Cahn, 2017). This chatbot also performs those four general components. The job performed inside those components in this chatbot is explained below:

#### 3.1. Natural Language Processing

This is the key step that makes this chatbot different from a rule-based chatbot. The main goal of NLP is to take user input in the form of text and produce a structured representation of the text which is also called Natural Language Understanding (NLU). The result from NLU is used to

identify the intent of the user. The intent of the user can be categorised by performing the following steps:

**3.1.1 Tokenization:** The first step is to break down sentences into individual words and remove punctuation marks from the text. Then this tokenized data is analysed using techniques called Bag of Words to create a data structure which is discussed below:

*i. Bag of Words:* The structure, order and syntax of the sentences are ignored and count the number of occurrences of each word. The vector space model is created by removing stop words.

In this project, two statistical models will be used to perform information extraction namely: Support Vector Machine (SVM) and Deep NLP. The SVM is a supervised machine learning tool where the set of labelled training data are used and the algorithm generates the optimal hyperplane that classifies the sample into their proper models. Deep NLP denotes deep learning for natural language, unlike traditional neural networks it consists of more hidden layers and each layer handles more increasingly complex features.

## **3.2. Response Generation**

The central component of chatbot architecture in response generation accesses data corpus. There are various response generation techniques or models available for performing response generation tasks such as Rule-Based Models, Information Retrieval (IR)- Based Models, Statistical Machine Transaction Generative Models, Sequence to Sequence (Seq2Seq) Model, Reinforcement Learning with Seq2Seq bots and so on. Among them, Seq2Seq Model and Reinforcement Learning with Seq2Seq seems useful for the project. So, for this chatbot, those two models will be tested and the model with high accuracy will be selected in between them.

## **3.3. Knowledge Base Creation**

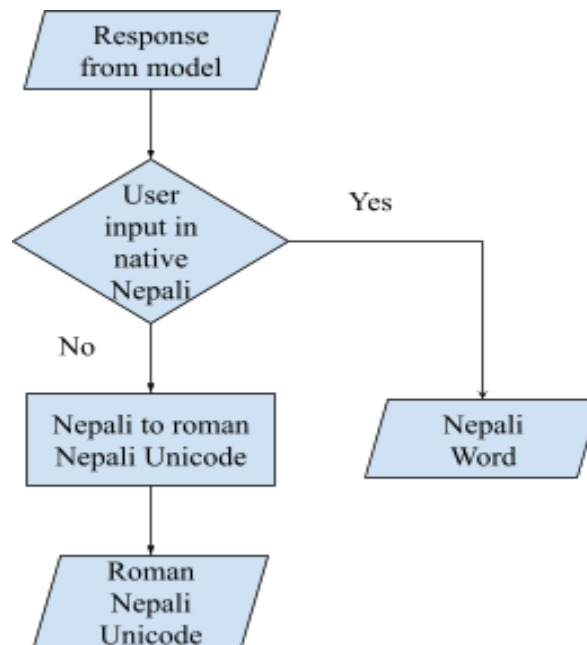
Chatbots become more intelligent as they have access to more knowledge. The corpora of data to train models of chatbot are retrieved by using information retrieval techniques such as scraping

by using a python library called BeautifulSoup and Scrappy. Furthermore, datasets will be prepared by taking the help of medical professionals.

### 3.4. Dialogue Management

Dialogue management in this chatbot is responsible for managing dialogue. This chatbot will have communication strategies such as in case a chatbot doesn't have an adequate response then it will ask the user to provide more information and so on.

The user will get a response from the model in a native Nepali text and in the output phase, the bot will check for in which mode a user provides input if a user sends input in native Nepali language then the bot will reply in native Nepali language otherwise the native will be converted into roman Nepali Unicode using built-in convertor and then display reply to the user.



**Figure 3.4:** Output layer flow

## **CHAPTER-4**

### **EXPECTED RESULTS**

This chatbot will build and embed in a web application as an interface to the users. The key expectation from this chatbot are listed below:

- i) This chatbot will provide an easy interface to the user to provide symptoms as input.
- ii) This chatbot will be trained with as many datasets as possible for basic disease diagnosis according to symptoms.
- iii) Generate a model that could have a higher accuracy rate for disease diagnosis which can be useful to provide useful information about diseases.
- iv) This chatbot is expected to be user friendly with a good user interface and user experience.

## APPENDIX A: GANTT CHART

The project will be five months long and the works are divided accordingly. The planned schedule for the project are illustrated in Gantt Chart below:

Tasks	February	March	April	May	June
Exploring related works					
Identify necessary technologies					
Build Datasets					
Development of Project					
Appraisal of project and make required changes					
Documentations					

## REFERENCES

- Athota, L., Shukla, V. K., Pandey, N., & Rana, A. (2020, June). Chatbot for healthcare system using artificial intelligence. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 619-622). IEEE.
- Badlani, S., Aditya, T., Dave, M., & Chaudhari, S. (2021). Multilingual Healthcare Chatbot Using Machine Learning. 2021 2Nd International Conference For Emerging Technology (INCET) (pp. 1-6). IEEE. <https://doi.org/10.1109/incet51464.2021.9456304>
- Behera, B. (2016). Chappie-a semi-automatic intelligent chatbot. cse. iitb. ac. In. [https://www.cse.iitb.ac.in/~bibek/WriteUP\\_2016.pdf](https://www.cse.iitb.ac.in/~bibek/WriteUP_2016.pdf)
- Cahn, J. (2017). CHATBOT: Architecture, design, & development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science.
- Deepika, K., Tilekya, V., Mamatha, J., & Subetha, T. (2020). Jollity chatbot- a contextual AI assistant. 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). <https://doi.org/10.1109/icssit48917.2020.9214076>
- Jia, J. (2003). The Study of the Application of a Keywords-based Chatbot System on the Teaching of Foreign Languages. ArXiv preprint cs/0310018. Retrieved from <https://arxiv.org/abs/cs/0310018>
- Key highlights from the Census Report 2021. (2022, February 15). <https://Nepaleconomicforum.org>. Retrieved February 25, 2022, from <https://nepaleconomicforum.org/key-highlights-from-the-census-report-2021/>