Bella

CONVERSATIONAL FITNESS CHATBOT FOR DESK WORKERS

THE TEAM

1. Benta Irungu
2. Carolyne Kambura
3. Harriet Joseph
4. Kepha Atika
5. Kennedy Murithi

SUPERVISOR

Samwel Jane

Moringa School

Data Science Core

Cohort: DSF-PT02

1. Business Understanding

1.1 Overview

During recent decades, the penetration of information communication technologies and the popularization of internet-based services has diminished the necessity of physical movement (e.g., business travel, or a walk to a colleague's office) and made it possible for people to complete most of their work while seated. This operating mode improved the efficiency of work, but also substantially reduced physical exercise and increased sedentary behaviours. According to the WHO, 60% to 85% of people in the world live in sedentary lifestyles without sufficient physical exercise, making lack of physical activity one of the most serious health problems across different ages. Beyond that, task-oriented working norms and high-pressured working environments have exacerbated office workers' inactivity at work.

Recent studies have shown an increase in sedentary behaviours in the working environment. In most office environments, it is tacitly approved that office-based employees stay in sedentary conditions at most times during the working day. The International Labor Organization reported that working hours are generally regulated to 8 hours per working day and are mostly spent in sedentary conditions. Earlier research suggested that prolonged physical inactivity at work can dramatically increase the risks of developing many occupational diseases and injuries. For instance, low levels of physical activity are increasingly recognized as important contributors to a variety of health problems, including heart disease, hypertension, colorectal cancer, obesity, and osteoporosis. Sedentary lifestyles are also significantly associated with an increased incidence of psychosocial problems, including depression, stress, and loneliness. Additionally, sedentary behaviours can lead to low back pain, neck discomforts, chronic shoulder problems, and many other musculoskeletal injuries.

When focusing on determinants of physical exercise at the workplace, earlier research has shown that workplace physical exercises can improve the health conditions of office-based workers who sit a lot. Many previous studies have been focused on investigating exercise interventions for office workers using e.g., public policies, health programs, environmental change, social supports, motivational materials, digital tools, expert consultations, etc. There are many barriers to physical exercise at the office, such as lack of time to perform physical activities, work burden and performance concerns, workplace policies and norms. For instance, performing physical exercises in the workplace may be heavily influenced by colleagues' and superiors' behaviour and attitude. There are also several limitations in the

office environment, such as lack of public space and public facilities. The barriers indicated above are crucial to understanding office workers' exercise behaviour.

Bella, a conversational chatbot for desk bound workers fitness, will provide information on managing office fitness, exercises, stretches and will include suggestions for ergonomics and stress management. This will result in better health outcomes and higher quality of life for office workers. The fitness assistant will provide 24/7 fitness assistance, and immediate response to office fitness-related queries.

1.2 Problem Statement

Prolonged physical inactivity at work can dramatically increase the risks of developing many occupational diseases and injuries. For instance, low levels of physical activity are increasingly recognized as important contributors to a variety of health problems, including heart disease, hypertension, colorectal cancer, obesity, and osteoporosis. Sedentary lifestyles are also significantly associated with an increased incidence of psychosocial problems, including depression, stress, and loneliness. Additionally, sedentary behaviours can lead to low back pain, neck discomforts, chronic shoulder problems, and many other musculoskeletal injuries.

Our goal, therefore, is to develop a conversational chatbot that can provide individuals with information and support on a wide range of office fitness and exercise topics. This will lead to better health outcomes and quality of life for deskbound workers.

1.3 Objective

Develop a conversational chatbot capable of understanding natural language queries related to office exercises, stretches and general stress management.

2. Data Understanding

The data was sourced from reliable websites through web scraping from different sources. The websites the data was obtained from include:

https://www.metropolisindia.com/blog/lifestyle/your-guide-to-desk-exercise-at-workplace-and-staying-fit/

https://www.indeed.com/career-advice/career-development/exercise-at-desk

https://health.clevelandclinic.org/desk-exercises/

https://www.insider.com/guides/health/fitness/desk-exercises

https://www.nasa.gov/sites/default/files/atoms/files/hq_deskfit_booklet_6.10.2020.pdf

https://www.vantagefit.io/blog/desk-exercises/

https://snacknation.com/blog/office-exercises/

3. Data Preparation

3.1 Steps For Rasa Data Preparation

(i) Gather and organize the training data:

The data was collected and organized into intents, and entities, and used to train models.

(ii) Annotate the data:

The data was annotated to identify intents and entities. This was done using the Rasa NLU Trainer tool.

(iii) Pre-processing: Tokenization and Lemmatization

These are common steps in natural language processing (NLP) that are used to prepare text data for further analysis. Pre-processing refers to the cleaning and normalization of text data. This can include tasks such as removing special characters, converting text to lowercase, and removing stop words.

Tokenization is the process of breaking down a piece of text into individual words or phrases. This is done so that the text can be more easily analyzed and processed by NLP algorithms.

Lemmatization is the process of reducing a word to its base form. This is done so that words that have the same meaning but different forms, such as "running" and "ran", are treated as the same word.

Featurization is the process of feature engineering in user messages. In RASA, we used pre-trained word embeddings and supervised embeddings.

Rasa is an open-source framework for building conversational AI that provides pre-processing, tokenization, and lemmatization functionalities. Rasa provided a python package rasa_nlu which contains the tools to do pre-processing, tokenization, and

lemmatization.

To use these functionalities in Rasa:

Rasa_nlu package was installed and then the necessary classes and functions were imported. Then, text data was passed through the pre-processing, tokenization, and lemmatization steps, and then the resulting data was used for further training.

(iv) Pipeline Configuration.

The NLU pipeline was configured in the Rasa config file to specify which components should be used to process the data.

(v) Training the NLU model:

Annotated data was used to train the NLU model. This was used to extract intents and entities from user input. This then determined how the Chatbot should respond to different user inputs.

4. Chatbot Architecture

4.1 Steps involved in chatbot architecture.

(i) User Input:

The user interacts with the chatbot through the user interface, which is a mobile app, or messaging platform. The user's input is sent to the Rasa chatbot in the form of a text message.

(ii) Rasa NLU

Rasa NLU is responsible for understanding the intents and entities in a user's message where it takes in the user's message as inputs and outputs the intent and entities in the message.

Intents refer to the goal or purpose of a user's message and entities refer to specific useful pieces of information that can be extracted from users' input.

Rasa NLU uses machine learning to train models on annotated training data which is used to predict the intent and entities in new and unseen messages.

(iii) Rasa Core

Rasa Core is responsible for managing the conversation flow and making decisions about what to do next in the conversation. It takes in the output of Rasa NLU as input and decides what action to take next.

Rasa Core uses a dialogue management model called a policy to make decisions about what to do next in the conversation. The policy is trained on examples of dialogs and uses machine learning to predict the next action.

(iv) Utter Response

Here the bot will respond to the user and if the user has any other questions or wants more clarification their new message will go through the same cycle.

Together, RASA NLU and RASA Core form a complete conversational AI system that can understand user inputs and make decisions about how to respond. The two components can be used separately or together, depending on the use case.

4.2 Evaluation Metrics

There are several ways we measured the accuracy of our bot including:

Intention accuracy: Bella was able to understand and fulfill the user's intent.

Response accuracy: Our bot’s response aligned with the expected or correct answer.

Task completion rate: Bella was able to carry out a large percentage of tasks successfully.

Error analysis: From tasks, we have been giving Bella we are in the process of identifying and analyzing errors made by the bot to understand where it struggles and improve upon it.

5. Conclusion and Recommendations

5.1 Conclusion of the project

A significant tool that can aid office workers stay fit and engaged was built. The chatbot can understand and respond to user input in a way that mimics speaking to a real expert thanks to its natural language processing abilities.

5.2 Recommendations

(i) Multi-language support

Increased accessibility: multi-language support allows chatbots to reach and serve a broader audience, particularly in multilingual countries or regions where people speak different languages.

Improved customer service: multi-language support allows chatbots to communicate with customers in their preferred language, which can improve the customer service experience and build trust and loyalty.

Increased efficiency: multi-language support allows chatbots to handle customer inquiries and requests in multiple languages, which can reduce the need for human intervention and increase efficiency.

Better data analysis: multi-language support allows chatbots to collect data from a wider range of users, which can improve the accuracy and effectiveness of data analysis and decision-making.

(ii) Integration with other systems

The chatbot can be integrated with other apps and services, such as nutrition trackers, to provide users with a more comprehensive view of their health.

(iii) Personalization

The chatbot can use machine learning to create personalized fitness plans for each user, based on their individual needs and goals.