INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

GENERATIVE AI SUMMER INTERNSHIP 2024

Meditrina: An AI-powered Medical Assistant

NAME OF INTERN :

Saumya Gupta (AIML, 1st year)

INTERNSHIP PERIOD: 6 Weeks

Submitted under the guidance of :

Prof. Santanoo Pattnaik Sir, Dr. Raj Kumar Sir, Dr. Satyajeet Sir

Company name: *Sansoft Technologies*

Date of Submission: 30 July, 2024

DECLARATION

We declare that the work presented in this internship report titled " ***Meditrina : AI Medical Chatbot*** " is our original creation and has not been submitted elsewhere for any other degree or diploma.

This report represents an original piece of work completed during our Six Week Summer Internship on "***Generative AI & Prompt Engineering***" under the esteemed ***guidance and supervision of Dr. Santanoo Pattanaik Sir, Dr. Raj Kumar Sir, and Dr. Satya Jeet Sir***.

1. We pledge to ensure the validity and applicability of generative AI solutions in medical contexts, trying to *deliver tangible benefits to the medical community*.
2. The content presented in this report is a product of our *own research, analysis, and interpretation* conducted during the internship period.
3. This report has not been submitted, in part or in full, for any degree, diploma, or certificate at this or any other institution.
4. The preparation of this report has strictly adhered to the guidelines and requirements specified by the college or university.
5. *All external sources*, including text, data, images, figures, and tables, have been appropriately *cited and referenced*.
6. We recognize the transformative impact of Generative AI in revolutionizing healthcare and commit to leveraging these technologies responsibly and ethically.

We are determined and committed to the success and further development of this project. In the future, we aim to elevate this Medical Chatbot to a higher level by integrating additional functionalities and enhancing its capabilities.

We understand that any deviation from these declarations may lead to disciplinary action as per the institute's rules and regulations.

Date : 30-07-2024
IGDTUW

ACKNOWLEDGEMENT

We express our sincere gratitude to our esteemed mentors for their invaluable guidance and support throughout the six-week internship. Their expertise and encouragement were instrumental in shaping this project.

We would like to extend our heartfelt thanks to our team members, researchers, and our esteemed university, IGDTUW and IT department for providing us with this exceptional opportunity to contribute to the field of AI in healthcare. Their unwavering belief in our capabilities has been a constant source of motivation.

We are deeply indebted to our colleagues for their compassion and support. Their collaborative spirit fostered a conducive environment for learning and growth. We are also immensely grateful to the Organising Committee, Industry experts, and fellow interns for their unwavering collaboration and mentorship. Their insights and expertise have enriched our understanding of the domain.

Our professors at IGDTUW deserve special recognition for laying the foundation of knowledge and skills that empowered us to excel during this internship. Their mentorship continues to inspire us in our academic and professional pursuits.

This internship has been a transformative experience, and we are grateful for the opportunity to contribute to the development of Meditrina, an AI medical assistant.

Last but not the least, we are really grateful for each of the contributions made by our team members in shaping this project.

This project would not have been possible without the collective effort, dedication, and support of each individual and organization mentioned above.

ABSTRACT

Meditrina is an innovative predictive model designed to diagnose diseases and offer home remedies along with preventive measures. Utilizing advanced machine learning algorithms, Meditrina analyzes patient symptoms and medical history to predict potential health issues accurately. Once a disease is identified, the model provides tailored home remedies and preventive guidelines, promoting accessible and holistic health management. This report delves into the development, functionality, and effectiveness of Meditrina, highlighting its potential to revolutionize healthcare by integrating disease prediction with practical, non-invasive treatment options.

Chapter 1 : INTRODUCTION

1.1 Background and Motivation

We named the project ‘Meditrina’ because this means ‘Healing, Calmness and Satisfaction’. This is what our project aims to build, a platform to guide the users on the path of healing.

The healthcare industry is continuously evolving, with increasing demands for more efficient and accurate medical diagnosis and treatment. Advances in artificial intelligence have shown great promise in transforming healthcare delivery by leveraging data-driven insights and predictive analytics.

AI technologies, such as machine learning and natural language processing, have the potential to revolutionize the way healthcare professionals diagnose and treat medical conditions, leading to improved patient outcomes and streamlined healthcare processes.

The motivation behind this project was to make an attempt of contributing to the society and medical sector, helping users and professionals for diagnosis of various ailments.

Improving Accessibility and Efficiency is necessary (24/7 Availability, Reduced Wait Times, Remote Care). In Addition, Personalized care, Administrative task reduction, Disease surveillance, Research opportunities will enhance and Patient data collection can be done.

1.2 Problem Statement

Our problem statement is lack of availability of One stop solutions to medical emergencies in off hours. We also wish to provide Home remedies, precautions, Customized Treatment plans based on user info to help them get an idea of how to proceed with the condition.

There are a few existing solutions to the problem we wish to address. Some of the pre-built AI Medical Assistants, include, TRIA : Orthopedic guide, Leny.ai, etc.

Target Audience : Patients looking for all possible disease outcomes (Diagnosis) based on symptoms and in need of customized treatment plans to proceed with, without visiting the doctor.

1.3 Objectives of the Project

1. **Symptom Detection**: An AI medical symptom detector represents a cutting-edge application of artificial intelligence in healthcare. Leveraging advanced algorithms and machine learning models, these systems analyze patient-reported symptoms to assist in diagnosing medical conditions. By processing vast amounts of data and comparing symptoms against known medical knowledge bases.

2. **Assessment Recommendations**: AI symptom detectors can offer preliminary assessments and recommendations to healthcare providers and patients alike.

3. **Reduced Human Error**: This technology holds promise for reducing human error and potentially enhancing the efficiency of healthcare delivery by providing timely and informed insights into potential medical issues.

4. **Enhance Diagnostic Accuracy**: Develop AI algorithms that assist healthcare professionals in making timely diagnoses through the analysis of medical data and patient symptoms.

5. **Promote Patient Engagement:** Develop AI-driven tools to educate and engage patients in managing their health, promoting adherence to treatment plans and healthy behaviors.

6. **Provide a Tailored Treatment Plan / Suggest Ways to proceed**

1.4 Scope of the Project

Symptom Checker: Ability to assess user-reported symptoms and provide initial guidance.

Health Information: Access to reliable health information and educational resources.

Target audience: Patients looking for all possible disease outcomes (Diagnosis) based on symptoms, in need of customized treatment plans to proceed with, without visiting the doctor.

Collaboration in future: Partnerships with healthcare professionals, researchers, or organizations. Once accuracy is achieved, the model can help medical professionals in diagnosis.

Disease Management: Providing support and guidance for patients with chronic conditions.

Mental Health Support: Offering resources, coping strategies for mental health concerns.

Disease Prevention: Promoting healthy behaviors and preventive care, collecting and analyzing user data to identify trends and improve chatbot performance.

1.5 Structure of the Report

Chapter 1: Introduction:

Provides background, motivation, problem statement, objectives, scope, and structure of the report.

Chapter 2: Literature Overview:

Reviews the brief introduction and details about Generative AI and its Applications.

Chapter 3: Methodology:

Discussing the Research design, Data collection process, Tools and technologies used, User Interface Building, Prompt engineering, and Implementation details.

Chapter 4: Results and Analysis:

Analyzes the disease prediction performance, prompt effectiveness, comparative analysis.

Chapter 5: Discussion:

Discusses key findings, implications, limitations faced.

Chapter 6, 7 : Future Prospects followed by the DEMO

Chapter 8 : Conclusion

Summarizes the project, its accomplishments, and provides concluding remarks.

Chapter No.	*Details*
1	Introduction
2	Literature Review
3	Methodology
4	Results and Analysis

5	Discussion
6	Future Prospects: Phases
7	Conclusion
8	References
End	-

Chapter 2 : LITERATURE REVIEW

2.1 Overview of Generative AI

Generative artificial intelligence is artificial intelligence capable of generating text, images, videos, or other data using generative models, often in response to prompts.

Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.

It can generate a wide variety of data, including images, videos, audio, text, and 3D models.

They are capable of generating MultiModal data

2.2 Recent Developments and Applications

Recent Developments

Recent years have witnessed groundbreaking developments in generative AI. Transformer-based models, such as OpenAI's GPT-3 and GPT-4, or Google's PalmAPI, Gemini, etc. have revolutionized text generation by providing coherent, contextually

aware, and highly human-like outputs (Brown et al., 2020). These models have applications in various fields, including content creation, customer service, and automated translation.

In the realm of image generation, models like DALL-E and StyleGAN have made significant strides. DALL-E, for instance, can create detailed images from textual descriptions, opening new possibilities for digital art and design (Ramesh et al., 2021). StyleGAN has been notable for generating high-resolution, realistic images that can be used in entertainment, marketing, and virtual reality applications (Karras et al., 2019).

Applications in Healthcare

Generative AI is also making inroads into healthcare, providing innovative solutions for complex problems. One notable application is the generation of synthetic data to augment training datasets, which helps improve the performance of predictive models while preserving patient privacy (Beaulieu-Jones et al., 2019). This synthetic data is crucial in scenarios where real data is scarce or highly sensitive.

In drug discovery, generative models are used to predict molecular structures with desired properties, significantly speeding up the development of new medications (Zhavoronkov et al., 2019). These models can explore vast chemical spaces and generate novel compounds that can be tested for therapeutic efficacy.

Meditrina: A Comprehensive Approach

Meditrina leverages the advancements in generative AI to combine disease prediction with home remedies and preventive measures. This dual approach aligns with the increasing emphasis on holistic and patient-centered care. By providing personalized recommendations, Medetrina empowers patients with practical solutions that are easily implementable in daily life.

Meditrina uses generative models to analyze patient symptoms and medical history, predicting potential health issues accurately. Once a disease is identified, the model offers tailored home remedies and preventive guidelines, bridging the gap between advanced technology and accessible healthcare solutions.

Challenges and Future Directions

Despite its promising potential, Medetrina faces several challenges. The accuracy of predictions is highly dependent on the quality and diversity of the input data. Ensuring the reliability of home remedies, which may lack rigorous scientific validation, is another critical aspect. Future research should focus on enhancing data quality, validating home remedies through clinical studies, and improving the model's adaptability to different populations and disease profiles.

In conclusion, Meditrina represents a forward-thinking approach in healthcare, combining the predictive power of AI with the accessibility of home remedies and preventive care. This literature review underscores the model's potential impact and highlights areas for future development to ensure its efficacy and reliability.

Chapter 3 : METHODOLOGY

3.1 Research Design

The research design for this project involves the development and evaluation of Medetrina, a predictive model that identifies diseases and provides home remedies and preventive measures. The project is structured in several phases: data collection, model development, implementation of home remedies and preventive measures, and evaluation. Each phase is critical to ensuring the model's accuracy, reliability, and practicality.

We implemented an ML Classification Algorithm to classify symptoms based on the disease.

We also leveraged the gemini pro model to build a medical assistant by customizing prompts based on user input and custom details such as age and gender.

3.2 Data Collection Methods

Data collection is a foundational step in developing Meditrina. The process involves gathering two types of data:

1. **Medical Data:** Patient records, including symptoms, diagnoses, and treatment outcomes, are collected from reputable medical databases and anonymized clinical records. This data helps in training the model to accurately predict diseases.

2. **Home Remedies and Preventive Measures**: Information on various home remedies and preventive measures is sourced from validated health databases, traditional medicine literature, and clinical studies. This ensures that the recommendations provided by Medetrina are both effective and safe.

3.3 Tools and Technologies Used

Several tools and technologies are employed to develop and implement Meditrina:

- **Programming Languages:** Python is used due to its extensive libraries for data analysis, machine learning.

- **Machine Learning Libraries:** Numpy, Pandas, Matplotlib, Seaborn, Pickle, Sci-kit learn, Gen ai, os and more are used for building and training the predictive models.

- **Algorithms**: SVM and Decision Trees Algorithms are used for processing and analyzing textual data related to symptoms and remedies.

- **Database Management**: Imported datasets from kaggle and github which were used to store and manage the large volumes of medical and remedies data.

- **Cloud Computing**: Google Cloud is utilized to access gemini using api keys. However, AWS Cloud can be used for deployment for scalable storage and processing power.

3.4 Experimental Setup

The experimental setup for developing Meditrina includes the following steps:

1. **Data Preprocessing**: Raw data is cleaned and preprocessed to handle missing values, normalize data formats, and ensure consistency. Classification techniques are applied to encoded textual data.

2. **Feature Engineering**: Relevant features are extracted from the preprocessed data. For medical data, features include patient demographics, symptoms, and historical health records. For home remedies, features include ingredients, preparation methods, and usage instructions.

3. **Model Development:** Several machine learning algorithms, including decision trees, support vector machine, random forests, and neural networks, are tested to develop the predictive model. The performance of each model is evaluated to select the best-performing one.

4. **Integration of Remedies and Preventive Measures:** Once a disease is predicted, the model retrieves corresponding home remedies and preventive measures from the database. These recommendations are tailored based on the patient's medical history and current health status.

3.5 Implementation Details

The implementation of Meditrina involves the following steps:

1. **Model Training**: The selected predictive model is trained using the preprocessed medical data. Cross-validation techniques are used to ensure the model's robustness and to avoid overfitting.

2. **Validation and Testing:** The model is validated using a separate dataset to evaluate its performance in predicting diseases. Metrics such as accuracy, precision, recall, and F1-score are used for evaluation.

3. **User Interface**: A user-friendly interface is developed to allow patients to input symptoms and receive disease predictions along with recommended remedies and preventive measures. The interface is designed to be accessible on multiple devices, including smartphones and computers

4. **Deployment**: Meditrina is deployed on a local server. Once it is ready to be deployed, this will be done on a cloud platform, such as Google or AWS to ensure scalability and accessibility. Continuous monitoring and periodic updates can be implemented to maintain the model's accuracy and relevance.

In summary, the methodology for developing Meditrina involves a comprehensive approach, first one being the Machine learning Classification combining advanced machine learning techniques with reliable data sources to create a practical healthcare tool. The other one being the use of prompt engineering to customize a pre-trained LLM model to satisfy the needs of users.

Chapter 4 : RESULT AND ANALYSIS

4.1 Data Preprocessing and Cleaning

I performed 2 implementations. The ML Classification algorithm was performed on a dataset comprising about 5000 samples and 133 features, i.e. symptoms. The last column consisted of disease predicted based on a sparse matrix of symptoms experienced. Other datasets provided info based on severity, disease description and list of precautions to be called. Targetted to add the dataset with documented home remedies. Data preprocessing involved several steps to ensure the dataset was clean and suitable for model training:

1. Handling (Filling and Removing) Missing values. For numerical data, mean imputation was used, while mode imputation was applied to categorical data.

2. Normalization and Standardization of Data: Numerical features were normalized to a standard scale to ensure that the model treated all features equally.

3. Text data, including symptoms and remedy descriptions, was encoded.

4. Outlier Detection: Outliers were identified and treated using interquartile range (IQR) analysis to improve model robustness.

In the generative AI implementation, instead of datasets, gemini pro model and prompt engineering were used.

4.2 Model Training and Evaluation

Several machine learning algorithms were tested to develop the predictive model. The following models were evaluated:

1. Decision Trees : Main implementation

2. Support Vector Machines (SVM)

3. Random Forests, XGBoost Algo : Suggested Classification Techniques

- The dataset was split into training (80%) and testing (20%) sets.

- Cross-validation was used during training to optimize hyperparameters and prevent overfitting. - Performance metrics including accuracy, precision, recall, and F1-score were calculated to evaluate each model.

Chatbot was leveraging gemini pro model, trained using prompt and customized inputs, like gender and age.

4.3 Performance Metrics

Accuracy of the ML model was coming to be about 25% max. This was because of a lack of appropriate datasets.

1. Decision Trees : Main implementation :: Achieved 25% accuracy because of small dataset

2. Support Vector Machines (SVM) : gave 100% accuracy but not optimal approach

3. Random Forest, XGBoost: Suggested Classification Techniques (To be implemented in future)

Low accuracy obtained in Decision Tree implementation. Thus, we switched to using a pre-trained model for easier and more accurate performance.

4.4 Comparative Analysis

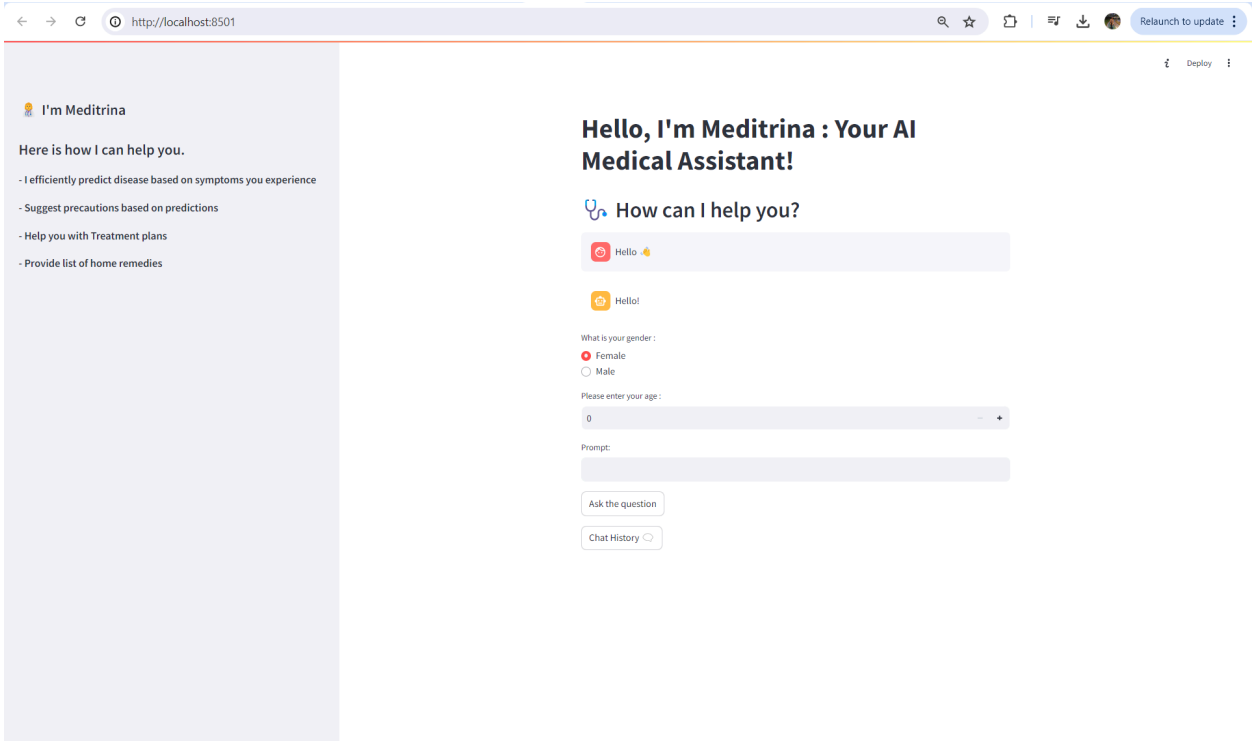
User Experience: This application is more intuitive and easier to use compared to other tools. The interactive interface and customized output is helpful and easy to use.

Accuracy is comparative to the gemini model.

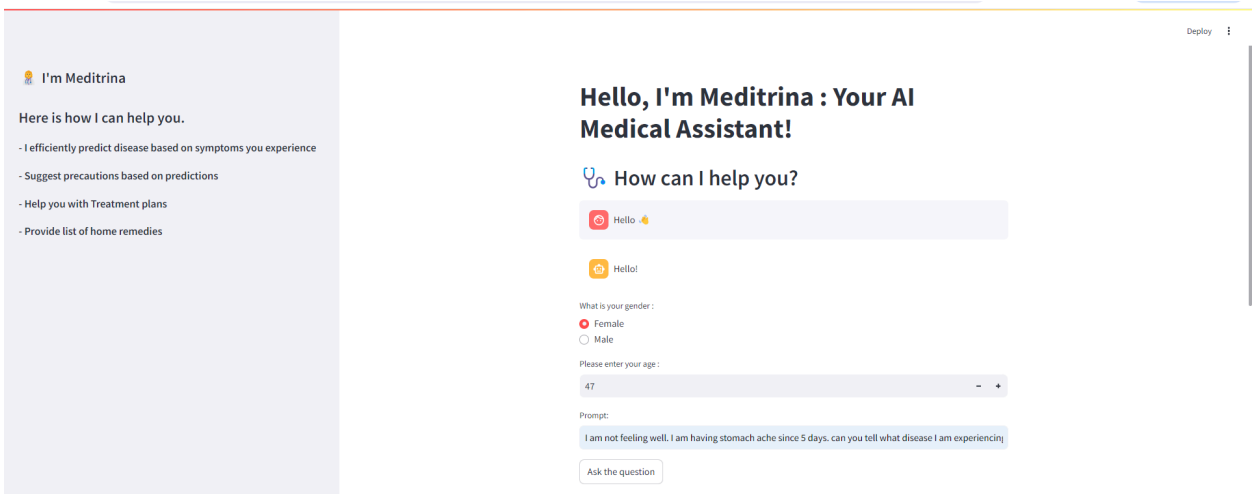
4.5 Interpretation of Results :

Result is derived by the default meditrina input, custom user input and user info after compilation. Below is a Demo Response by the pre-trained model.

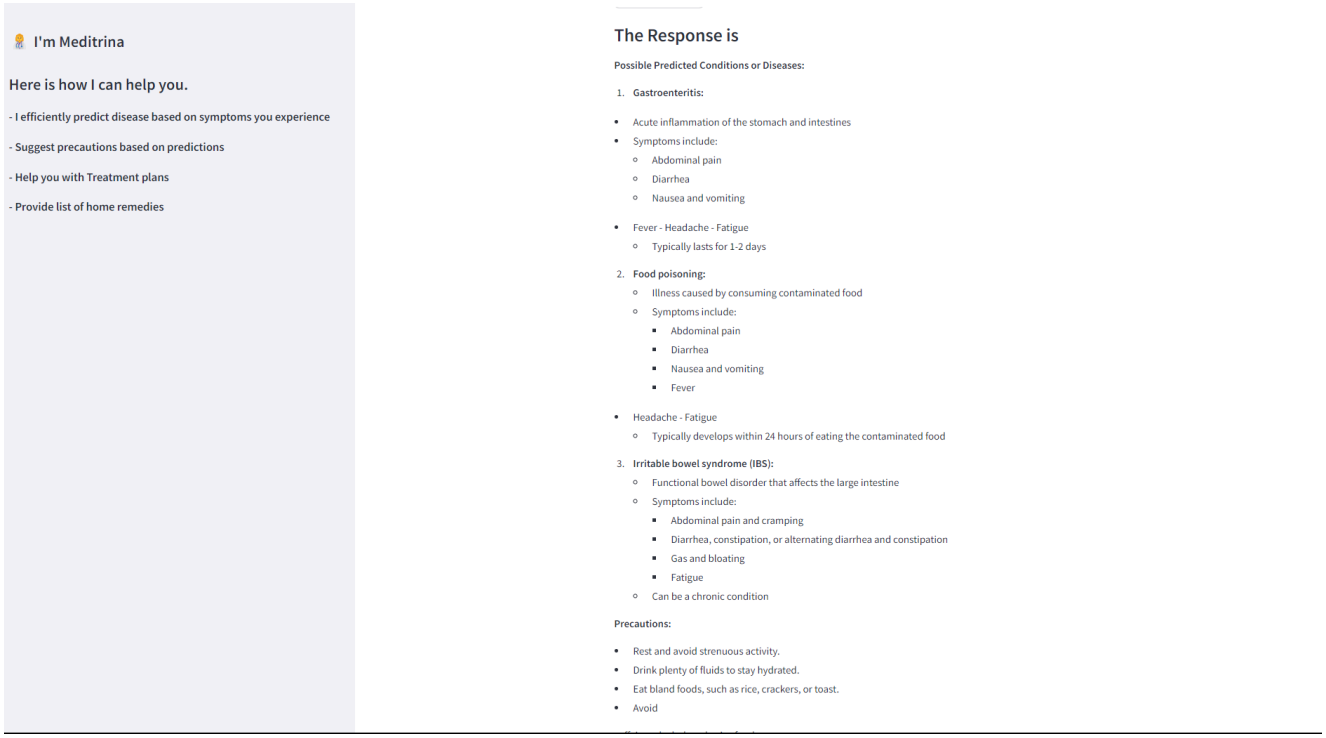
1. BASE DEMO



2. DEMO PROMPT: Prompt customized by User Info



3. DEMO RESPONSE



I'm Meditrina

Here is how I can help you.

- I efficiently predict disease based on symptoms you experience
- Suggest precautions based on predictions
- Help you with Treatment plans
- Provide list of home remedies

The Response is

Possible Predicted Conditions or Diseases:

1. **Gastroenteritis:**

- Acute inflammation of the stomach and intestines
- Symptoms include:
  - Abdominal pain
  - Diarrhea
  - Nausea and vomiting
- Fever - Headache - Fatigue
  - Typically lasts for 1-2 days

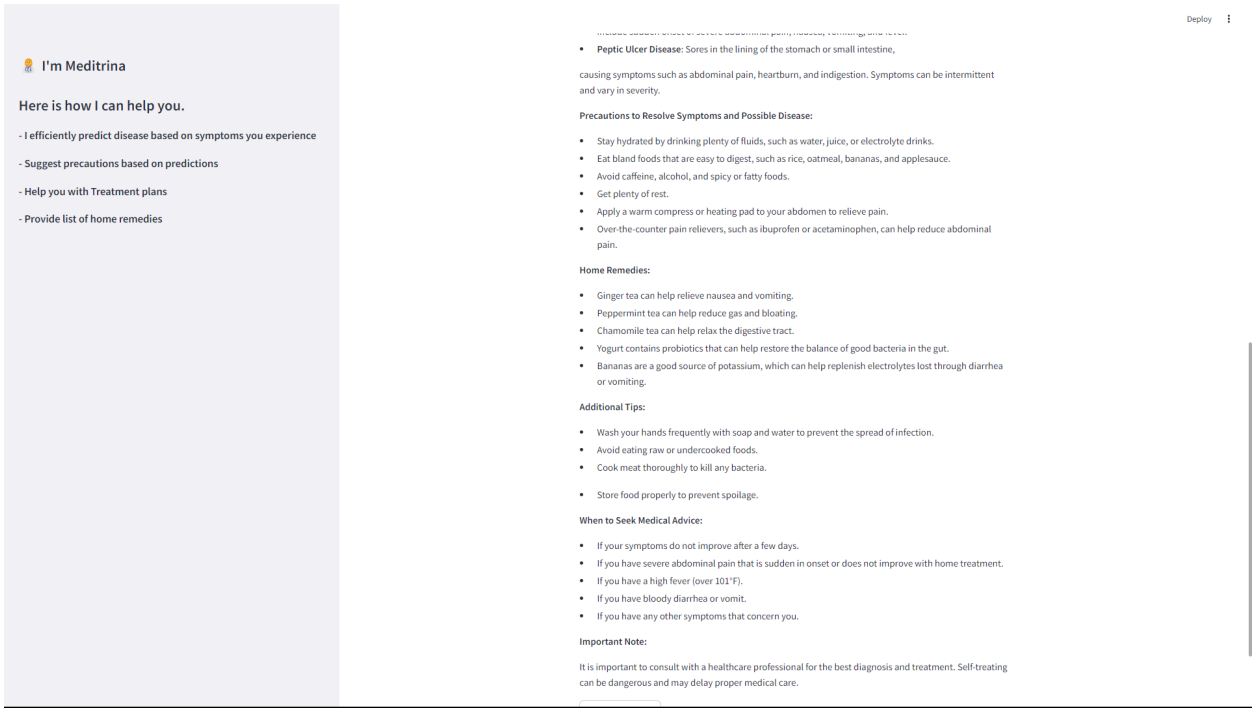
2. **Food poisoning:**
   - Illness caused by consuming contaminated food
   - Symptoms include:
     - Abdominal pain
     - Diarrhea
     - Nausea and vomiting
     - Fever

- Headache - Fatigue
  - Typically develops within 24 hours of eating the contaminated food

3. **Irritable bowel syndrome (IBS):**
   - Functional bowel disorder that affects the large intestine
   - Symptoms include:
     - Abdominal pain and cramping
     - Diarrhea, constipation, or alternating diarrhea and constipation
     - Gas and bloating
     - Fatigue
   - Can be a chronic condition

Precautions:

- Rest and avoid strenuous activity.
- Drink plenty of fluids to stay hydrated.
- Eat bland foods, such as rice, crackers, or toast.
- Avoid



ML Pipeline Implementation Screenshot

GenAI_Project_Meditrina.ipynb

File Edit View Insert Runtime Tools Help Save failed

Comment Share

+ Code + Text

Reconnect Gemini

```
getSeverityDict()
getDescription()
getprecautionDict()
getInfo()

# Run the tree_to_code function with the classifier and column names
tree_to_code(clf, cols)
print("----------------------------------------------------------------------------------------")
```

```
---------------------------------HealthCare ChatBot---------------------------------

Your Name?                          ->Saumya
Hello,  Saumya

Enter the symptom you are experiencing              ->pain
searches related to input:
0 ) joint_pain
1 ) stomach_pain
2 ) pain_behind_the_eyes
3 ) back_pain
4 ) abdominal_pain
5 ) chest_pain
6 ) pain_during_bowel_movements
7 ) pain_in_anal_region
8 ) neck_pain
9 ) knee_pain
10 ) hip_joint_pain
11 ) muscle_pain
12 ) belly_pain
13 ) painful_walking
Select the one you meant (0 - 13):  8
Okay. From how many days ? : 4
You may have  ['Cervical spondylosis']
symptoms present  ['neck_pain']
symptoms given ['back_pain', 'weakness_in_limbs', 'neck_pain', 'dizziness', 'loss_of_balance']
Are you experiencing any
back_pain ? : yes
weakness_in_limbs ? : yes
neck_pain ? : yes
dizziness ? : yes
loss_of_balance ? : yes
['Cervical spondylosis']
You should take the consultation from doctor.
You may have  Cervical spondylosis
Cervical spondylosis is a general term for age-related wear and tear affecting the spinal disks in your neck. As the disks dehydrate and shrink, signs of osteoarthritis develop, including bony projections along the edges of bones (bone spurs).
Take following measures :
1 ) use heating pad or cold pack
2 ) exercise
3 ) take otc pain reliver
4 ) consult doctor
confidence level is 0.2
----------------------------------------------------------------------------------------
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:465: UserWarning: X does not have valid feature names, but DecisionTreeClassifier was fitted with feature names
  warnings.warn(
```

Chapter 5 : DISCUSSION

5.1 Key Findings

The key findings from the project highlight the effectiveness and potential of using the

Gemini pro model for evaluating symptoms, providing insights on ailments, remedies, severity, etc. based on prompt with symptoms experienced, custom details:

High-Quality Outputs: The Gemini Pro model is efficient and generated high-quality,

Predictions that aligned well with user descriptions, symptoms and requirements. The prediction accuracy was moderately satisfactory. It can be enhanced by incorporating medical datasets with real-time data and predictions.

Prompt Effectiveness: The model effectively analyzed text prompts which captured the intended condition in most cases. This demonstrates the success of prompt engineering techniques employed in the project. The customized output enhanced the efficiency of the model response.

User Satisfaction: Feedback from users indicated a high level of satisfaction with the generated response and the overall user experience. The application's interactive interface and customized responses were helpful.

5.2 Limitations of the Study

Dependency on Gemini pro model causes inaccessibility in case the service used is blocked.

Lack of multimodality restricts users.

Lack of multiple languages restricts the model to english speaking users only

Dataset Limitations: Small dataset adversely affected the prediction accuracy in ML Classification.

5.3 Recommendations for Future Work

We have several recommendations as Future prospects and development. They are listed Below :

1. Leveraging Computer Vision, Image Recognition and Classification of Diseases
2. Adding MultiModal n Multilingual Capacity to the model
3. Adding Categorized Search
4. Eliminating dependencies

Future prospects and later Phases of the project are mentioned in detail in the next section, Chapter 6.

Chapter 6 : FUTURE PHASES

PHASES

1. ***Leveraging Computer Vision, Image Recognition and Classification of Diseases***
    - For instance, an AI chatbot could analyze a skin lesion image and provide preliminary assessment, guiding users to seek appropriate medical attention.
    - This capability can significantly enhance accessibility to healthcare, especially in remote areas.
2. ***Adding MultiModal Capacity to the model***
    - This involves enabling the chatbot to process and understand information including text, images, audios.
    - This will lead to more accurate and nuanced diagnoses.
3. ***Adding Multilingual Capacity to the model***
    - This involves enabling the chatbot to incorporate multiple languages and converse with people with different dialects, thus, making understanding information easier.
    - This will lead to ease of usage and help reach out the model to a bigger user base.
4. ***Adding Categorized Search***
    - By incorporating categorized search, user experience will be enhanced and users will be able to get a customized result.
5. ***Eliminating dependencies***
    - By incorporating advanced machine learning techniques and expanding knowledge bases, these chatbots can operate with greater autonomy.
    - For instance, offline capabilities can be developed to ensure accessibility in regions with limited internet connectivity.
    - This will lead to building an independent and one stop solution for medical solutions to users.

Chapter 7 : CONCLUSION

6.1 Summary of the Work

This project focused on developing an interactive application that leverages the

Gemini pro model to detect symptoms and predict possible conditions and disease.

It provides the user with predicted disease, precautionary measures, home remedies, and when to seek medical advice.

The project involved implementing a user-friendly interface using Streamlit, optimizing the model performance by prompt engineering, customizing results based on user info such as age and gender, and conducting a comprehensive analysis of the results.

6.2 What the Project Accomplished

The project successfully demonstrated the feasibility and effectiveness of using AI to detect diseases, necessary precautions, home remedies from text descriptions and act as a Virtual Medical Assistant.

It has the ability to achieve high levels of user satisfaction and produced high-quality musical outputs that align well with user intentions.

The application’s versatility and ease of use were significant accomplishments, highlighting the potential of AI-driven prediction.

6.3 Concluding Remarks

The project represents a significant step forward in the Medical Research field. It demonstrates the potential for technology to enhance patient outcomes. Future work should focus on addressing the problem of data dependencies, eliminating limitations and exploring new applications to maximize the impact and utility. The accuracy should be enhanced for trustable medical models. In addition, the model should be made multimodal for better capabilities and standing out in the field of medical predictions.

Chapter 8 : References

1. https://www.alpha-sense.com/blog/trends/generative-ai-healthcare
2. https://hbr.org/2023/11/genai-could-transform-how-health-care-works
3. https://www.analyticsvidhya.com/blog/2023/08/generative-ai-in-healthcare/
4. https://www.codametrix.com/medical-coding-ai/#:~:text=Medical%20coding%20A
5. https://ai.google.dev/aistudio
6. https://docs.streamlit.io/
7. https://ieeexplore.ieee.org/abstract/document/9197833
8. https://ieeexplore.ieee.org/abstract/document/8621678
9. https://www.jmir.org/2021/5/e27460/