Data analytics chatbot

This project is aimed at developing an AI powered data analytics chatbot to help students learn and apply data analytics techniques

This chatbot is built using several powerful libraries such as pandas scipy statistics scikit-learn matplotlib seaborn plotly streamlit and many more

This provides an llm integration using the module Gemini

This bot will assist students in generating insights analytics performing statistical analysis visualizations and major understanding core ML functions

**Software use for this project**

Inorder to perform this project jupyter notebook cannot be used coz of this limited integrations instead tools like VS code as well as pycharm can be used to perform this project since it comes with various integrations

In this project we will be moving forward with the VS code application

**Objectives:**

Interactive learning – It enables students to interact with various tools and receive real time codes as well as visualizations

ML support – Assist in applying basic ML techniques or ML algorithms and help in evaluating models

Visualizations – It can generate highly detailed visualizations and provide instant plots which can be used to understand the datasets

Code snippets and explanation – The chatbot should provide various code snippets and explanations as per the users requirement in performing EDA statistical analysis as well as ML models

NLP – With the use of Gemini llm the users can utilize the gpt to help with questions and detailed explanations

Modules used for this project

**Backend:**

* Python fundamentals
* Numpy
* Pandas
* Matplotlib
* Seaborn
* Scipy
* Statistics
* Requests
* Scikit-learn (ML orientations)
* Plotly

**Frontend:**

* Streamlit module
* NLP

Gemini module (API) - For integrating GPT to handle more complex queries for better understanding

Regex parsing – RE module – IT is used for reading user queries, extracting relevant data from datasets with complex queries using specific functions.

**Key features**

There are several key features which will be implemented using the modules and the software

Some of those features are

Data upload and handling – The users can upload datasets or files and the bot will display the first few rows as a preview and generate insights based on the query provided

EDA – Users can request statistics visualizations and several other ML tools or functions

Statistical Analysis – Users can fetch descriptive statistics as well as testing statistics

ML Assistance – The user can build or utilize basic ML models such as supervised models or unsupervised models

And also it provides and utilizes model evaluation techniques such as accuracy classification report and many more

Gemini integration – If the user have some queries the gemini api can be utilized as gemini’s api can be used to generate human like responses

Query parsing with regex – This bot can recognize patterns structures provided by the user such as several data queries as well as statistical queries

**Future improvements of the project**

The chatbot is limited to csv files but in the future excel json file formats and many other files can be integrated with this chatbot

Advanced ML support include many ML algorithms such as random forest support vector and many more

Enhanced NLP Implement advanced NLP techniques for queries

**Project workflow**

User interaction

Users will interact with the chatbot with the help of streamlit interface by typing queries

Query parsing – Now before generating results the query is recognised by basic keyword matching

If the query is too advanced gemini’s api is invoke to provide the results according to the query and dataset

Task execution – If the query relates to data handling relevant functions from pandas matplotlib seaborn and scipy are executed but if the query requires statistical analysis modules such as numpy scipy and scikit-learn will be utilized

Response – The result of the query will be provided and displayed using the streamlit module inside the frontend part

Inorder to establish a python environment or virtual environment inside the terminal this syntax is used

Python -m venv venv

Gemini has limited usage

**Conclusion for the project**

This data analytics chatbot is created or developed using python fundamentals streamlit and most importantly gemini

It offers a comprehensive learning experience for students who are in the domain of data analytics as well as data science

With the help of several python modules the user has integrated visualizations statistics ML and NLP with the help of regex module

This chatbot serves as an important educational tool which is dynamic and engaging

This chatbot is powered by gemini 1.5 flash model llm to answer questions related to datasets

This project contains all the fundamental concepts as well as some advance concepts which have been learnt and implemented over the course of time

**References**

* Python documentation
* Streamlit documentation
* Numpy documentation
* Pandas documentation
* Scikit-learn documentation
* Matplotlib Documentation
* Chatgpt model
* Gemini model

**Project folder structre**

* Data Analytics Chatbot
* .venv
* .streamlit – secrets.toml
* app.py
* chatbotlogic.py
* gemini handler.py
* requirements.txt
* utils.py
* .assets
* \_pycache