Dibyajit Bag

Bachelor of Technology
Computer Science And Engineering
Netaji Subhash Engineering Colleger

+91-9073404947
bag.dibyajit@gmail.com
GitHub Profile
LinkedIn Profile

Education

•Bachelor of Technology in Computer Science and Engineering(Data Scientist) *2019-23*
Netaji Subhash Engineering College CGPA: 9.02

•Higher Secondary Education *2018-19*
Baranagar Narendranath Vidyamandir GRADE: A

•Secondary Education *2016-17*
Hindu School PERCENTAGE: 91.14

Projects

•Laptop Price Prediction Using Machine Learning

A ML based regression problem, implemented using advanced data preprocessing, feature engineering, and diverse ML algorithms.

- The project involves Exploratory data analysis(EDA) with visualizations, comprehensive data preprocessing including one-hot encoding and scaling, and employs KNeighborsRegressor and RandomForestRegressor for laptop price prediction.
- Technology Used: Python, Pandas, Matplotlib, Seaborn, Plotly, scikit-learn, XGBoost, Linear Regression, KNN.
  GitHub:Github Code Link

•Hotel Booking Analysis and Predictions

To analyze and understand the patterns in hotel bookings and predict cancellations.

- The project begins with comprehensive data visualization using libraries like Matplotlib, Seaborn, and Plotly to analyze hotel bookings. Advanced data preprocessing techniques, such as iterative imputation, are employed to handle missing data.
- Technology Used: Pandas, Matplotlib, Seaborn, Plotly, Scikit-learn, IterativeImputer, SimpleImputer, StandardScaler.
  GitHub:Github Code Link

•GitHub Repository Complexity Analysis

A tool designed to fetch, preprocess, and evaluate the technical complexity of GitHub repositories.

- The tool fetches user repositories from a provided GitHub URL and processes each repository. For Jupyter Notebooks, it converts them to Python code and segments large code snippets. For other file types, it handles both small and large files, ensuring that data is read in chunks when necessary. The tool then evaluates the technical complexity of each repository using a GPT-2 based model.
- Technology Used: Python, requests, OS, nbconvert, transformers, GPT2LMHeadModel,NLP, GPT2Tokenizer.
  GitHub:Github Code Link

•Interactive Assistant using OpenAI's Large language model(LLM) (NLP project)

To facilitate real-time interaction with OpenAI's language model, handling rate limits and token constraints.

- The script integrates with the OpenAI API to communicate with a language model and retrieve responses. It has built-in mechanisms to manage rate limits and token constraints, ensuring smooth interaction. The conversation history is maintained for context, and a user can interact with the assistant in a command-line interface.
- Technology Used : Python , Large language model(LLM), GPT-3.5-turbo, NLP, API.
  GitHub:Github Code Link

Awards & Achievements

- **Winning Mercor Competition** (July)
  - Code link: GitHub
  - Document: Drive Link
- **Publication In FOSET magazine**
  - Link: Drive Link

Technical Skills and Interests

Languages: C/C++, Python, PyTorch, TensorFlow
Libraries : C++ STL, Python Libraries
Data Analytic Skills: Excel, Google Sheets, SQL, Google Colab, PowerBI, Developing Dashboards, Exploratory Data Analysis(EDA)
Cloud/Databases: Structured Query Language(SQL)
Relevant Coursework: Operating Systems, Object Oriented Programming, Database Management System, Software Engineering.
Areas of Interest: Machine Learning, AI, Deep Learning, Data Science, Data Analyst.
Soft Skills: Problem Solving, Self-learning, Presentation, Adaptability

Certifications

•TensorFlow Developer Certificate in 2022:Zero to Mastery (Udemy) Certificate Link

•Introduction to Machine Learning (Coursera) Certificate Link

•AI For Everyone (Coursera) Certificate Link