Raunak Ghosh

Email : ghoshraunak48@gmail.com
Mobile : +91-9051029918
Address: Kolkata, India,700051
Github: https://github.com/raunak-29
LinkedIn: https://www.linkedin.com/in/raunak-ghosh-448b191b9/

SUMMARY

Results-driven Data Scientist with a strong foundation in Mathematics and Statistics and nearly 3 years of experience developing scalable ML, NLP, and Generative AI solutions. Proficient in Python, SQL, ML/DL frameworks, and cloud platforms (AWS, Azure, GCP). Passionate about solving real-world problems through Machine Learning and Generative AI.

EXPERIENCE

Virtusa Consulting Services Pvt. Ltd., Hyderabad
Role: Software Engineer | June 2022 – Present

- **AI-Based Suspect Identification System**:-Designed and deployed a full-cycle ML solution to detect suspect utility connections, including data extraction from PostgreSQL, anomaly detection, feature engineering, and training a Random Forest classifier; deployed the model using FastAPI, automated daily inference via cron jobs, and integrated results into an Angular-based dashboard for real-time monitoring within the client's secure server environment.
- **Chat with Documents**: Led the development of a user-centric application leveraging AWS Bedrock's Generative AI models (Titan, Claude, Jurassic, Cohere). Enabled users to upload PDF documents and interact with LLMs for summarization, custom instruction-based content generation, and keyword extraction. Designed a three-tab UI: one for document processing, one for AI agent interaction, and one for document-specific chatbot communication.
- **Contact Centre Automation**: Spearheaded the development of a GenAI solution using Azure OpenAI's GPT-3.5/4 to automate query handling. The system performs summarization, keyword extraction, sentiment analysis, intent recognition, and generates follow-up messages from customer interactions.
- **GenAssure**: Contributed as a Generative AI Developer in a project focused on automated test case generation. Utilized UI workflow recordings (videos and images) to generate test cases, plans, and robot scripts using Google's Gemini-Pro and Gemini-Pro-Vision models. Technologies used: Python, OpenCV, LangChain, Prompt Engineering.
- Worked in a team project titled "**Wine Quality Analysis**". The dataset had several input variables and task was to predict the quality score of wine (between 0 and 10). My task was to perform basic EDA, check for outliers , converting imbalanced dataset into balanced one using **SMOTE** and perform Feature Selection. Finally different classification models were fitted (Logistic Regression, Decision Tree, SVM, Random Forest) and checked for accuracy. **Random Forest** was the found to be one with highest accuracy(94%) was fitted for predictions.
- **Sentiment Analysis – IMDB Reviews**: Collaborated with a small team on a sentiment analysis project. Conducted extensive data preprocessing (stopword removal, HTML tag cleaning with BeautifulSoup, stemming, and lemmatization). Transformed text using BoW and TF-IDF, and implemented classifiers including Logistic Regression, Decision Tree, and SVM. Achieved ~86% accuracy with Logistic Regression.

EDUCATION

- **IIT(ISM) Dhanbad** — Dhanbad, Jharkhand
  Master of Science in Mathematical and Computing; CGPA: 9.14 — *August. 2020 – May. 2022*
- **Jadavpur University** — Jadavpur, Kolkata
  Bachelor of Science in Mathematics; CGPA: 9.26 — *August. 2017 – August. 2020*

SKILLS

Technical Skills

- **Programming Languages**: Python (Proficient), C (Proficient)
- **ML/NLP/GenAI Libraries**: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Plotly, NLTK, Langchain, Llama Index, Langgraph, TensorFlow
- **Databases**: MySQL, PostgreSQL
- **Tools & Platforms**: MS Excel, Power BI

Mathematical Skills

- Linear Algebra, Differential Equations, Probability & Descriptive Statistics, Inferential Statistics, Numerical Analysis

Machine Learning & AI

- **Supervised Learning**: Linear Regression, Logistic Regression, KNN, Naïve Bayes, Decision Trees, Random Forest
- **Unsupervised Learning**: K-Means Clustering, Hierarchical Clustering
- **Deep Learning**: ANN, RNN, CNN
- **Natural Language Processing**: Bag of Words , TF-IDF, Word2Vec, Encoder- Decoder Architectures
- **Generative AI**: Prompt Engineering, RAG, Agentic-AI systems

CERTIFICATIONS

- Virtusa Certified GenAI Assisted Engineer
- Oracle SQL Basics authorized by IBM(Coursera)
- Supervised Machine learning: Regression authorized by IBM(Coursera)
- Unsupervised Machine learning authorized by IBM(Coursera)
- Introduction to Natural Language Processing provided by Analytics Vidhya
- Kore.AI Experience Optimization Platform Developers Basic Training
- Kore.AI Experience Optimization Platform Developers Advanced Training

ACHIEVEMENTS

- JAM 2020 qualified- AIR 480