AQUIB FARHAAN HUSSAIN

DATA SCIENTIST

Bangalore, India | +91 8638510516 | aquib.farhaan2@gmail.com
LinkedIn: aquibfarhaan | GitHub: aquib97

PROFESSIONAL SUMMARY

Experienced Data Science professional with 6+ years of expertise in developing AI products, solution-driven architectures, and automated pipelines. Skilled in Predictive Modeling, Data Analytics, Statistical Analysis, and delivering scalable, impact-driven solutions.

SKILLS

Programming Languages: Python, SQL
Database Management: MySQL, SQLite, Redis, MongoDB
Framework: FastAPI, Flask, RestAPI
OS: Windows, LINUX
Cloud Platform: Microsoft AZURE, Google Cloud Platform (GCP)
Field of Study: Machine Learning, Deep Learning, NLP, GenAI, Statistics, Time series
Tools: Numpy, Pandas, Scipy, Statsmodel, Scikit-Learn, NLTK, Tensorflow, Keras, MLflow, Databricks, LangChain
Visualization tools: Matplotlib, Seaborn, Plotly, PowerBI

WORK EXPERIENCE

ACCELEDGE | Pune, Maharashtra
DATA SCIENTIST

Apr 2023 - Present

VOICE-BOT PLATFORM

- Developed a scalable voicebot integrated with a dialer API to manage outbound call campaigns, tracking **10,000+** queued, active, and completed calls daily via MongoDB and temporary caching.
- Designed and implemented an end-to-end service for real-time audio interactions, integrating speech-to-text transcription, multilingual support, and dynamic conversational behavior using LLM-driven logic and intent classification with JSON-based dialogue mapping.
- Upgraded the system by incorporating GenAI (OpenAI + LangChain) to enable scalable and dynamic response generation, enhancing performance and expanding conversational capabilities.
- Integrated a robust audio response generation pipeline with session-aware memory management to simulate natural, voice-based conversations in diverse languages and user personas.

WEB-BOT PLATFORM

- Engineered a scalable Chabot platform for Web-based and WhatsApp based application product configured with a JSON-based architecture, setup for bot flows and deployment on Linux based server.
- Implemented database management with MongoDB for session management, RAG and fine tuning leveraging with LLM and hugging face model to generate language modeling.
- Integrated advanced functionalities into the chatbot platform, including API integration, sentiment analysis, translation mapping, intent classification, multilingual support, and GenAI-driven capabilities.

AUDIO PROCESSING AND NLP AUTOMATION

- Built FastAPI microservices for multilingual audio pipelines, incorporating language identification (SpeechBrain), transcription (Google STT, SpeechRecognition), neural translation, and domain-specific NER models, processing over **5,000+** audio files/month.
- Architected APIs with flexible input support (file/base64), token-based security, modular utility functions, and resilient fault handling to ensure scalable deployments and seamless cross-language NLP tasks.

DATA SCRAPING AND AUTOMATION

- Developed a Selenium-based web scraping tool to systematically navigate websites and extract updated data files on a daily basis. Automated the workflow using a scheduler to enhance operational efficiency.
- Designed post-processing pipelines for extracted data and established a daily logging system, seamlessly integrating with an SQL database for real-time mapping and visualization on the client dashboard.

NETWORTH CORP | Bangalore, Karnataka

Feb 2021 - Apr 2023

DATA SCIENTIST

TOTAL EXTRACT LOSS OPTIMIZATION

- Led root cause analysis to identify and address factors contributing to yield loss in the Lauter Tun process, employing regression techniques and Shap value-driven feature importance.
- Generated actionable insights from the analysis to define the recommended range of process control parameters for maximization of yield based on analysis of historical data & increased the yield by **0.65%**.
- Designed and implemented scalable data models and data pipelines to streamline the storage and analysis of high-volume process datasets.

SPACE OPTIMIZATION

- Built an optimization tool to create the best SKU combination for route assembling thereby taking the currently vacant space of truck and minimizing cost.
- Calculated benefits with **$300k+** (pesos) from overall monthly from the output files.

ELECTRICITY FORECAST IN BREWERIES

- Found patterns of electricity consumption on the MAZ region Brewery dataset, where the records of kilowatts consumed are in every 15 minutes.
- Demonstrated proficiency in time series forecasting techniques, achieving an RMSPE of **37%** for ARIMA models and **10%** for LSTM models on the validation set.

OTHER ACHIEVEMENTS

- Developed and deployed an automation tool for validating MRP controller user details and managing SAP responsibilities, with an optimization module to prioritize cases. Runs daily, auto-handling average of **20** incidents.
- Built an automation tool to monitor SLA breach incidents, send periodic alerts to assignees, and help save priority cases, contributing to an average cost saving of **$10M**.

TECH MAHINDRA | Noida

Feb 2019 - Dec 2020

PYTHON DEVELOPER

- Developed and maintained Python-based applications for data processing and automation, improving operational efficiency by **30%**.
- Implemented ETL processes using Python libraries (Pandas, NumPy) to transform and load data from various sources into SQL databases.
- Created RESTful APIs using Flask framework to facilitate seamless data exchange between different systems.
- Collaborated with cross-functional teams to identify requirements and deliver solutions that met business needs.
- Automated manual reporting processes, reducing report generation time by **75%** and eliminating human error.

PROJECTS

SENTIMENT ANALYSIS AND MODEL TRAINING

- Analyzed a Twitter dataset of **1.6M** records to discern the sentiment (positive or negative) of tweets.
- Trained various models, including GloVe+Stacked Bi-LSTM, ANN, and Logistic Regression, to assess their performance in sentiment analysis on the data.

AUTOMATION OF OIL WELL SHUTDOWN PROCESS (FREELANCE)

- Investigated the behavior of neighboring wells in the vicinity of a center well shutdown within a specified radius.
- Managed data from **60** wells, handling a substantial dataset of approximately **13 GB**.
- Conducted experiments in predictive analysis and visualization techniques, automating the process to enhance data representation.

EDUCATION

Bachelor of Science (Honors in Physics)   Jun 2015 - May 2018

Gauhati University | Guwahati, India
C.G.P.A: 7.5/10

CERTIFICATIONS

- **Machine Learning Masters** | iNeuron Intelligence | Nov 2020 - Jul 2021
- **Business Analytics Masters** | iNeuron Intelligence | Oct 2020 - Jun 2021

LANGUAGES

- English
- Hindi
- Assamese