ADITYA GAURAV

☎ +91-7845820047
✉ advit200@gmail.com
Github.com/Advit200
linkedin.com/in/advit200

Experienced Data Scientist with close to **9 years** of expertise in data analysis, ML/DL/Gen AI and transforming business objectives into actionable data science tasks and generating value through AI. Demonstrates strong analytical and problem-solving capabilities, meticulous attention to detail, and excellent teamwork skills.

PROFILE SYNPOSIS

- Adept at creating Enterprise grade Agentic LLM based application, fine tuning LLM/embedding model & RAG.
- Proficient in developing, enhancing and maintaining Time Series, Supervised Machine Learning and Deep Learning based Predictive and Statistical models.
- Experienced in MLOps, managing the complete AI lifecycle from ETL pipeline development to model deployment on cloud platforms – AWS and Azure.
- Possesses a solid grasp of Statistics, Hypothesis Testing and Mathematics behind AI algorithms.
- Proficient in writing optimized Python code with a strong understanding of DSA and OOPs concepts.
- Good knowledge of AWS and Azure cloud services, including Azure ML Studio, Azure Data Factory, Azure App services and Azure Kubernetes Service.
- Basic knowledge of Big Data ecosystem and Pyspark.

TECHNICAL SKILLS

Languages	Python, SAS, SQL
ML Domain	Regression, Classification, NLP, Clustering, Pricing, Time Series, Network Analysis, Gen AI/LLM
Databases	MySQL, MongoDB, Vector DB
Web Framework	Flask, Django, FastAPI
BI Tool	Power BI, Tableau
Cloud	AWS, Azure

CERTIFICATION

- Microsoft certified - Azure Fundamentals
- Applied Accelerated AI – *NPTEL (Nvidia)*
- Data Structure & Algorithm,Python– *NPTEL(IIT Madras)*
- Network Analysis and Network Optimization – *SAS*
- Generative AI with Large Language Model - *Coursera*

EDUCATION

College	Degree	CGPA	Year
BITS Pilani	MTech-AIML	8.49	2023-25
VIT University, Vellore	BTech-Mech	8.55	2012-16
DAV Public School	Class-XII	78%	2011
DAV Public School	Class-X	83%	2009

WORK EXPERIENCE

Sr. Data Scientist | SymphonyAI | Dec' 23-Ongoing

➢ **AI for Work:** Developed Agentic AI Studio, a low-code/no-code, easily customizable SaaS product aimed at improving employee productivity. The platform allows users to define a flow by dragging components onto a canvas, seamlessly connecting them to agents to trigger their agentic flow.

- Built custom components/MCP servers for HR, IT, coding assistance and employees general queries, which are available as out-of-the-box templates for new customers. Utilized Langflow, Langchain and Langfuse for development.

➢ **Enterprise Knowledge Management (EKM):** Developed a RAG-based application that integrates with SharePoint and Confluence to retrieve personalized knowledge articles based on user location, role, and permission, delivering precise response tailored to each user's needs.

- Used advanced techniques like re-ranking, multi-query, parent document retriever, follow-up questions and fine-tuned embedding model (used MRL).

➢ **Copilot:** Developed a copilot to automate 70% of ITSM tasks for analysts and end users, from creating service requests to resolving tickets. It also handles scheduling meetings, booking flights, applying for leave, checking company policies and many more, making it an all-in-one platform for various tasks. Used LangGraph for development and deployed on AKS.

➢ **Service Desk Intelligence (SDI):** Trained a Tensorflow based non-linear NN model using diverse data types to predict ticket escalation risk, change management success probability and resolution SLA meet probability.

- Developed a Catch-And-Dispatch optimization model to assign new tickets to the best-suited workgroup and analyst. Implemented MILP using pulp library, resulting in a measurable reduction in resolution time and improved analyst efficiency.

➢ **MLOps and ETL pipeline:** Automated data science life cycle by using Azure Data Factory, Databricks/PySpark and Azure ML to trigger model training on new data, perform hypermeter tuning, register the best model and deploy it to production while maintaining version control.

Data Scientist | Koch Business Solutions | May' 21-Dec' 23

➢ **Nqaire :** Designed a LLM based chatbot utilizing Azure OpenAI and Langchain, to analyse global events' financial, economic, and other impacts on Koch companies by amalgamating external internet data with internal data.

- The chatbot exhibits memory, possesses the ability to query internal DB - by generating SQL queries, search Wikipedia/Google News, formulate probing questions to enhance analysis, suggest better prompts and can present output in the form of charts, tables, or summarized text, detailing impacts.

➢ **Cell/CTB Forecast :** Conducted mill-level production volume forecast for product - Cellulose and Containerboard to enhance inventory management, production optimization, and customer satisfaction.

- Employed SARIMAX on four years of monthly training data to forecast 12 months ahead. Integrated planned outages and business forecast as exogenous variables in the models.

➢ **Sheet Feeders Pricing :** Led a pricing strategy initiative to enhance and standardize pricing for Sheet Feeders products by leveraging historical transaction data to improve profitability.

- Proposed model focused on adjusting "Material Margin as % of Net Sales (MMP)" to enforce a 20% profitability floor by shifting low-margin transactions below the MMP median.

➢ **BitterSweet**: The project's objective was to construct an aggregator featuring integrated text analytics and insights for Koch Industries by collecting and analysing reviews from diverse social media platforms.

- Conducted Sentiment Analysis, Aspect-based sentiment analysis employing a NLI model and Topic Modelling on the review text. The resulting analysis provided valuable insights for organizational leaders to evaluate the company's reputation and identify potential issues within their respective units.

➢ **Voice of customer**: VOC project was aimed at identifying patterns in data and performing text analytics on customer's feedback to find key factors influencing customer experience during different stages of business and suggest possible areas of improvement.

- Employed diverse NLP techniques, including keyword extraction (KeyBERT), NER (Spacy), Topic modelling (BERTopic). Analysis of negative feedback revealed that delayed equipment delivery, fabrication issues, and response time were prominent factors contributing to customer dissatisfaction.

➢ **Tissue Softness Prediction**: Project's goal was to create a real-time alert system for operators, facilitating prompt quality checks and minimizing waste due to tissue softness degradation. Developed a predictive model using streaming sensor data to predict tissue softness at 10-minute intervals.

- Applied Genetic Algorithm for feature selection and built a unified XGBoost model across 10 plants. Deployed the model on SAS ESP. Utilized SHAPLEY values for cause-effect analysis, aiding plant operators in decision-making.

Sr. Data Analyst | Accenture | Jan' 17 – April'21

➢ **OneReport** : This project centers on the development of a web application for data owners to upload data files, with automatic data retrieval from the database, leading to the creation of a consolidated Data Lake that facilitates effortless Power BI refreshing.

- Engineered the web application's backend using Django with MongoDB as the database, streamlining the data preprocessing pipeline. Implemented features for data validation, checks, and PII or sensitive information encryption before database insertion.
- Enabled Row-Level Security (RLS) and Scheduled Refresh for Power BI using Data Gateway. Additionally, developed a monthly email and Teams notification system through Flask API.

➢ **AQI Tool** : The project's primary goal was to develop a tool for the comparative analysis of business documents (Excel, CSV, PDF, PPT) against a standardized template, based on 14 metrics, ultimately generating an Asset Quality Index (AQI).

- Designed a graphical user interface (GUI) using tkinter, enabling users to select multiple files, apply filters/search functionality, and generate detailed reports along with the AQI. Implemented OCR for extracting text from images and employed TF-IDF with cosine similarity to identify duplicate images, sections, and tables within the documents.

PERSONAL PROJECTS

- Developed a sentiment classification model, which involved training a custom BERT model with PyTorch. API deployment was managed through Nginx, Docker, and Kubernetes on AWS-EKS.
- Constructed a supervised machine translation model utilizing the Locality-Sensitive Hashing (LSH) based k-Nearest Neighbours (k-NN) technique to translate English words into French.
- Created a chess game between human-computer using the Minimax and Alpha-Beta Pruning algorithms for strategic move selection.
- Developed a rabbit maze game employing A* and Hill climbing algorithm to navigate the maze efficiently and escape with minimal cost.

ACHIEVEMENTS

- Worked as Python & Data Analytics Instructor/Mentor for a batch of 10 freshers at KTC.
- Received Spot Bonus 5 times in two years at KTC.
- Gold Badge in Python and SQL – HackerRank.
- Secured GATE score (Mechanical) of 647 (95 percentile).