Anurag Mishra

Email: anurag.mishra899n@gmail.com
Medium: anuragmishra_27746
Mobile: +91-9456243057
Github: github.com/anurag-mishra899
LinkedIn: Anurag Mishra

Skills Summary

- **Experience**: **6.5+ YoE** in designing and developing **Machine Learning, Deep Learning, Generative AI, and NLP** solutions for various industries, including retail, banking, telecommunications, and chemicals
- **Frameworks**: PyTorch, Langchain, LlamaIndex, Fastapi, Scikit, NLTK, SpaCy, TensorFlow, Keras
- **Tools**: Kubernetes, Docker, GIT, PostgreSQL, MySQL, SQLite
- **Platforms**: Azure Cloud, AWS, GCP, Linux
- **Certified in**: **Azure Fundamental, Azure Data Scientist, Databricks ML Professional**
- **Publications**: Article 1, Article 2, Article 3, GenAI Conference

Experience

- **EY GDS** Hybrid
  Associate Manager (Full-time) *Feb 2022 - Present*
  - **Deployed GenAI Chat-bot Solution in Production**:
    - Led **a team of 6 engineers** in designing and developing **Chat-bot solution for 2-3k users across 40 Branches** for Largest bank in Philippines.
    - Built solution reduces time-spent manually from **5-10 min/query to <30 seconds.**
  - **Developed LLM Production Pipeline**:
    - Implemented **custom prepossessing pipeline and chunking strategy** to handle **complex tables, pdf formatting & flow-charts etc.**
    - Build scalable deployment pipeline for using Fastapi, Langchain & Azure API Gateway, Kubernetes to **support throttling, fallbacks, load balancing to handle rate-limit, exceeded quota issues**.
  - **LLMOps & Evaluation Pipeline**:
    - Implemented custom RAG evaluation framework for **automated assessment for E2E RAG and individual components (routing, retrieval, generation etc)** to improve performance and optimize the solution.
    - Developed **monitoring pipeline using Langfuse** to observability and logging. Build Feedback Capturing mechanism to continuous enhancement of solution.
  - **Smart Call Centre Solution for US Based Automotive Client**:
    - Developed LLM & data ingestion pipeline integrating Azure DevOps, Azure App Services, OpenAI models, Langchain & integration with ServiceNow to manage Post-Call Actions, including call summarization, Ticket Creation & Updates, categorization and Quality Assurance.
    - Designed pipeline to handle 300-400 calls/centre which increase agent's productivity by 60% in hourly-time spent.
  - **Fine-tuning of Transformer-Based Models (BERT, RoBerta etc)**:
    - Perform fine-tuning of transformer based models using Pytorch, hugging-face for downstream tasks i.e. NER & Text Classification to build document extraction solution.
    - Deployed deep learning models as batch endpoints utilizing Azure Web Apps, Databricks Jobs, and MLflow for streamlined integration and efficient model management.
  - **Go-to-Market Strategy on Generative AI**: Involved in "Go to Market" strategy on Generative AI for various client and developed prototype on different industry use-cases using LLM models i.e. OpenAI, LlaMa, MPT, Dolly etc and took client awareness session on "Industry Implementation and Risks of GenAI"

- **Deloitte (Office of US)** Remote
  Consultant - AI (Full-time) *Feb 2021 - Feb 2022*
  - **Fixture Optimization for Inventory Management**: Devised optimization solution to allocate optimum number of fixtures for each product and brands for one of biggest global retail chain stores. The solution has provided $5 million reduction in cost and expense.
  - **Custom Clustering Approach with Constraints**: Implemented clustering and optimization algorithms to assign the optimum fixtures, while keeping up the various geographically and logistically business constraints.
  - **Sales Forecasting for Retail Client**: Performed Sales Forecasting model for each department per store to assess the demand. Utilized various Tree based algorithms e.g.- Gradient Boosting, XGBoost & LightGBM and regression-based model e.g.- ARIMA, & SARIMA for the prediction.
  - **Tools & Technologies**: Used Python programming and worked with libraries such as Pandas, Scikit-learn, NumPy, Matplotlib, and Seaborn.
- **Virtusa Polaris** Hybrid
  Machine Learning Engineer *June 2018 - Feb 2021*
  - **Multi-Class classification of Trading Products**: Devised solution for Product Controllers to reduce person-hours spent on tasks of identifying breach in trading activity. Used multi-class classification and ETL operations to perform pre- and post-breach analysis.
  - **NLP Models to Assess Commentary using Spacy**: Developed NLP solution to identify the inconsistency in PnL commentary generated by Product Controllers. Milestones included building RestAPI to render user request and ETL operations to handle user requests.
  - **Impact**: Built solutions has increased the numbers of tagging 15k /hour to 1 million /hour
  - **ML Models/Approaches**: Implemented various ML algorithms like Random Forest, Logistic regression, Support Vector Machine, Gradient Boosting and parameter tuning the model for the optimum solution
  - **Tools & Technologies**: Used Python programming and worked on libraries like Pandas, Scikit-learn, NumPy, Matplotlib, Seaborn, Spacy

Projects

- **Fine-Tuned LLM for Domain Specific tasks using LoRA**: Performed Fine-tuning of LLM models(llama 3.1 8B, mistral 7B) on 80 GB A-100 GPU on custom build dataset for Manufacturing domain.
- **LLM Based Agentic flow for Text to Code Solution**: Agent based solution that convert requirements to code. The solution uses multi-agents (debugger, coder, tester etc) to write the bug-free and optimized code.

Education

- **National Institute of Technology, Manipur** India
  Bachelor of Technology - Information Technology; GPA: 8.5 *July 2014 - June 2018*

  Courses: *Operating Systems, Data Structures, Analysis Of Algorithms, Artificial Intelligence, Machine Learning, Networking, Databases*

Honors and Awards

- Awarded Excellence Client Performance Award in EY
- Honoured to be among in top 10 students in CSE Department
- Achieved (Among Top 1 % in board) in Entire district in HSC exam.