PRATISH DEWANGAN

+91-7987224140 | dpratishraj7991@gmail.com

- Experienced **(3.10+ years)** in developing end to end Python Digital products and chatbot based application leveraging next gen technologies such as **Genai (LLM Model), Azure AI Cloud Integration, RAG, LangChain, LangGraph, Prompt Engineering**.
- Have hands on experience in **Automating the Deployment of ML Models** in Azure Machine Learning Studio by **Azure Devops CI/CD pipeline and also in GithubActions** with model registering and parameter monitoring in **ML Flow**, setting Data Drift with Azure Data Assets.
- Have experience in building and training **NLP and Machine Learning Models like BERT, Random Forest, XgBoost** and Deploying it and also in Fine Tuning LLM Based models like **BERT base and LLAMA 2.0 34B parameters Model**.
- Have hands on experience **in Docker, Machine Learning Algorithms, writing scalable python code, tensor flow code and SQL queries, and have good proficiency in statistics,** and also have exposure to work in Angular and **Django Rest Frame work** as backend.
- Have Good exposure to Cloud services with **4X Microsoft** certified in **Azure AI Engineering Associate, Azure Data Scientist Associate.**

TECHNICAL SKILLS

Skills - MlOps in Azure Machine Learning Studio, Machine Learning, Deep Learning (CNN, NLP), RAG Model, GenAI, LLMs, Prompt Engineering, LLM Fine Tuning, LLMOps in (NVIDIA Nim, NVIDIA Neo)
Language – Python (Pandas, Matplotlib, Seaborn), TensorFlow and Pytorch, Lang Chain, Lang Graph
Techniques - Data Structure and Algorithms, Design Patterns, OPPS in Python
Azure - Azure AI Studio, Azure AI Search, Azure Machine Learning, Azure DevOps (CI/CD), Docker Registry, Azure Prompt Flow
FrameWork/Tools - Django REST Framework, Github Actions, Air Flow, DVC, Dashhub, ML Flow, Docker

EDUCATION

- Master of Technology in **Industrial Engineering & Operations Research from IIT Delhi** (2019-2021)
  Course Work – **Supply Chain Management**, **Operation Research, Stochastic Optimization**, Operation Plaining and Control, and more

WORK EXPERIENCE

Machine Learning Engineer | **Tiger Analytics** *[Nov 2024 – Present]*

Project- DataFlyWheel

- Building a **LLM Fine Tuning Pipeline** start from Data Curation (Data Extraction, Data Filtering, Duplication Removal, PII) pipeline in NVIDIA Nemo framework to LoRa Fine Tuning via NVIDIA Microservice.

Project - PowerBIChatBot

- Build a **PowerBI Chatbot** that's interacts with PowerBI Dashboard, and interacts with the KPI/Measures in the Dashboard and gives answer to complex user questions asked around the measure/KPI.
- Used **GPT-4o mini** as a GenAI Model to generates the DAX queries for the user questions and executes in the semantic layer and extracts the relevant data and at the end used **summarization model** for final data summary. Also Integrated Agentic Flow for auto correction of Dax generated by the model up to 5 attempts.
- Build the complete architecture in python core that can handle group by Based, Comparison based, filtered based questions, with total latency of 7sec per KPI, further optimize and reduce the latency to 4 sec with **multi-threading parallel code execution**.
- Written **Automation Script** for model validation that uses genAi model for auto verification of user questions and gives the accuracy
- Further Improve the accuracy to 100% by feeding DAX example and with **Prompt Engineering** techniques.
- Used **Microsoft Bot Framework** as a bot UI and deployed it into in Teams bots via Azure VM and used **Cosmos DB** to store bot conversation

Impact – Reduce the effort of manually giving the filters and analyzing the data for Low level worker who don't have access to PowerBI

AI/ML Engineer | HCL Tech *[Aug 2023-Oct 2024]*

Project – GenAI

- Automated an entire SDLC Cycle from generation of User stories to generation of backend code as well as frontend code to deployment of generated application in Azure Web App via Docker Registry by creating docker Image using **Generative A**I and **Langchain**.
- Integrated recent released Local LLM's model such as **LLAMA 2.0, Mistral MoE** to undertake data security.
- Used **LangChain** and **Prompt Engineering Techniques** to integrate a sequence of prompts in generation of codes as well as mermaid file.
- Written logic of zipping generated data code file while calling the application API with its prompt input and transferring and storing its file location in **Azure Blob Container**. Written logic of files transferring files structure from blob to blob and deletion in blob in python backend.
- Perform a Instruction Fine tuning in **LLAMA 2.0** 34B parameter model using **PEFT Lora and QLora** technique to our self-made custom data
- Integrated **Lang Graph** as a **AI Agent** framework to account for cyclic nature of backend and frontend code development, and also for model Validation.
- Design entire architecture with backend logic in Django and deployed this application in **Azure Web App** by **CI/CD pipeline** in **Azure DevOps.**.
- Build **RAG system** to account for maintaining latest code version of python-based backend code, used **Azure AI Search Service** for Indexing the code documents and used searching techniques to get relevant context for the model

*Impact – Reduce production effort of **backend code by 80%** and **frontend by 70%,** results in quick initialization of complete SDLC application with just a single prompt from user, & with predefined yaml settings result in **60%** reduction in effort in deploying the application in Azure Devops.*

Internal Tasks -

- Provided support in writing YAML configuration in dockizing and deploying the ML Models in Azure web via CI/CD pipeline.
- Done **NLP Model based Deployment** starts from conversion of code snippets to scalable python script, Optuna Integration of Hyperparameter tunning, writing pytest, logging integration, **Github Actions workflow** yaml configuration, configuration of compute resources for deployment, integration of **ML flow** for Model parameter tracking , and setting the data drift, feature store in Azure ML Studio

AI/ML Developer| HCL Tech *[Nov 2021-Jun 2023]*

Project – Genie

- Build a Domain Driven fully functioning **DataOps platform** for end Data Scientist, **written its scalable python code** undertaking design pattern and OOPs concept.
- Worked as end to end architecture designer and developer. Creates user stories, business logic implementation, and database design strategy
- Written **complex SQL queries** to facilitate the seamlessly fetching of tables joined via foreign key, in doing contextualization and data mapping
- Written **extraction logic** from scratch so it can handle different datatypes input and its parse such as MF4 data, Excel, CSV at the same time.
- Integrated the complete business logic in **Django REST frame work**, build Django Models to creates several CRUD API to integrate in Angular
- Build a **NLP Model** from scratch for doing Fuzzy Matching between Domain Model Table and User Data Table from description column and its name, attend an accuracy of 85% from **Random Forest, XGBoost Based Model**, and 89% accuracy with **LSTM or RNN based mode** and 94% accuracy **with transformer based BERT Model**.
- Integrated and modified **python-based Rule Engine** to give flexibility of writing complex rule in Transformation of User Data once it is mapped to Domain Model.
- Build fully functioning UI of this platform in **Angular**, written typescript code and its function to handle to API from Django Backend.

*Impact – Results in seamlessly handling of Upstream related Data. **Reduced 30%** effort in contextualizing & transforming the data for end D.S*

Internal Tasks -

- Studied Cognite Data Fusion from end to end, dash boarded its open Industry data (Assets, Timeseries events data) in **PowerBI visualization**.
- Build regression based **predictive modeling** with ensemble-based ML Algorithm **like Random Forest, Gradient Boost, LigthGBM** in event data of Cognite Data Fusion.

PROJECTS

Discrete Event Simulation and Optimization Model of Kidney Transplantation | IIT DELHI THESIS *[Jul,2020-June 2021]*

- Collected data of Donor, Recipients, Cities, Hospitals from South ROTTO states and used the data to create a simulation model (of each Individual states as well as combination of whole south ROTTO region) which is used to find the following outputs: **probability of a patient receiving the organ within one to five years of registration**, average number of deaths per year due to lack of donated organs, average time to transplant for wait-listed patients.
- Performed **Hypothesis Testing (Chi-squared Test)** to evaluate probability distribution in data of Organ arrival and Patient arrival.
- Used **Stochastic Simulation Optimization** method to locate additional transplantation center in South ROTTO states to optimize parameters like average wait time, average transportation time and many more, a total of 20+ parameters involved.
- Coded entire simulation in Python from scratch have more than **10k lines of simulation codes**. Used Object Oriented Programming to create objects of Patient and Donor and to optimize reusable codes. Used central **global timeline approach** to facilitate combined simulation of 5 ROTTO states simultaneously.
- Created allocation policies with different combinations of CIT (Cold Ischemia time) radius, Nearest Transplantation Center **(Nearest Neighbor Method Applied)**, Max KAP score and analyzed the outcomes of each allocation policies impacted on the simulation outcomes.

Gurgaon Real State Data Science Application *[Jan,2024]*

- Developed an end to end Real State Price Prediction application on Gurgaon DataSet with Recommendation System. Gathered Data from multiples sources mainly 99acrs using **Selenium** and performed all data preprocessing pipeline such as **Data Cleaning and Merging, Feature Engineering, EDA, Outlier Detection and Missing Value Imputation and Feature Selection**.
- Performed model selection from various Regressor Models such **as LR, SVR, LASSO Regression, Ridge Regression, DT, KNN** etc
- Build a **recommendation system** for selecting real state based on top facilities, location advantage, price details
- Integrated complete pipeline on Django Framework and build UI in Angular. And deployed the application in Azure Web App via CI/CD pipeline in Azure DevOps.

GenAI OpenBox | Visual Studio Code Extension | **Self-Initiated | 160+ Downloads** *[Feb 2024]*

Build a Open Box type GenAi Co-Pilot Visual Code Extension in TypeScript where user can use their own API key to power this extension, it provides code generation and code summarization capabilities. Publish this extension into VS market place with name **GenAI OpenBox**

CERTIFICATIONS

Online Program -

- 2 Month **Software Engineering and Data Science Training | HCL Tech** *[Aug,2021-Oct,2021]*
- 2 Month **MLE** Foundation Training | **Tiger Analytics** *[Nov,2024-Jan,2025]*

ACHIEVEMENTS

- **Passed Azure Fundamental** Certification Exam (Azure AZ 900 certified) *[Feb-2023]*
- **Passed Azure AI Fundamental** Certification Exam (Azure AI-900 certified) *[Feb-2024]*
- **Passed Azure AI Engineer Associate** Certification Exam (Azure AI-102 certified) *[Mar -2024]*
- Passed **Azure Data Scientist Associate** Certification Exam (Azure DP-100 certified) *[Jul 2024]*
- Passed **Problem Solving Test in Python** | Hacker Rank *[2020]*
- Passed **International knowledge management** (IKM-Test in Python | *HCl* *[Mar,2022]*
- Earned **Gold star badges in Java, SQL and Python** (Hacker Rank) *[2021]]*
- **Graduate Aptitude Test in Engineering (GATE)**: Secured **98.46 Percentile** in GATE (Mechanical Engineering) *[2019]*