

PROFILE SYNOPSIS

- A Data Scientist having 9+ years of Software engineering experience, including 6+ years of Data Science and Machine Learning experience.
- Having the drive to push boundaries and make significant real-world business impacts using **Artificial Intelligence**.

TECHNICAL SKILLS SUMMARY

Azure Services – Azure AI Search, Azure Machine Learning, Azure AI Services, Azure AI Studio

Languages – Python, C#, Java

LLMs – AI Agents, Llama3, GPT, PaLM, BERT, XLNET, T5 etc

Deep Learning Frameworks –Tensorflow, PyTorch, keras

IDE – Jupyter Notebook, PyCharm, VS Code, Visual Studio,

Tools – **langchain**, **liamaindex**, pandas, scipy, seaborn, keras, sklearn, keras, tensorflow, pycaret, spacy, tesseract (OCR), PySpark, PyTorch, OpenCV etc

Web Frameworks – Flask, Django

Deployment – docker, docker-compose, **MLOPs- Azure and AWS**, Kubernetes, Jenkins

Source Control – bit bucket, git, Azure Machine Learning Repo

Operating Systems – Linux, windows

Cloud Platform – AWS and Azure

OCR Tools – Tesseract and AWS Textract

Distributed Data Systems - Spark

SOFT SKILLS:

- Interactive person with quick learning capabilities, analytical skills, good communication and language proficiency
- Core team member of **Mindtree Foundation – Corporate Social Responsibility** wing and had worked with several NGOs
- **Other Information: Prepared for civil services** for three years and have deep knowledge of general issues, geographical locations, and ethical conduct.

EDUCATIONAL QUALIFICATIONS

- Master of Computer Applications
- Bachelor of Computer Applications

PROFESSIONAL CERTIFICATIONS

- Executive PG Certification in MLOps by IIT Roorkee

- AI Master Engineer - IBM and Simplilearn certified
- Deep Learning with TensorFlow IBM and Simplilearn certified
- Big Data Hadoop and Spark Developer Simplilearn Certified
- Advanced Analytics with Amazon SageMaker - AWS Certification
- **Tableau Desktop Certified** Professional Certification – Simplilearn

WORK EXPERIENCE

Role: AI TECHNICAL LEAD/COE LEAD

Responsibilities:

- As an AI Technical Lead, I design, develop, and deploy advanced artificial intelligence solutions, using the latest algorithms and technologies and achieve organizational goals with a focus on scalability, efficiency, and seamless integration.
- I lead a talented team of engineers to revolutionize processes, optimize performance, and unlock new opportunities across diverse domains.
- Developed LLM based product called **AssistAI** for DAMCO where **GenAI** is used extensively for functionality implementation.
- Developed **Intelligent Document Processor** a proprietary product of Damco Solutions to process unstructured documents.
- Worked on LIAMA2 fine tuning and hosting for a medical requirement.
- Worked on deploying MLOps pipelines for a sports analysis company.

Role: SENIOR DATA SCIENCE ENGINEER

Responsibilities:

- Deployed ML model on **Azure MLOps** platform.
- Implemented AI OCR based solution for US based **Business Management Firm** which help OEMs in global dealer data assimilation and analysis, and deciding sales targets. Used **AWS Textract and LSTM models** for prediction.
- Worked on weather-based staffing prediction project (**Insurance Domain**).
- Deployed Decision Tree based ML models for loan prediction (Banking Domain).
- Got good exposure to the **AWS services (EMR, SageMaker, Athena)** while implementing Machine Learning project for Insurance vertical. **Random Forest, XGBoost and CatBoost Classifiers** were used for classifying claims assignment records. **PySpark** understanding improved during project implementation
- Also implemented **Spacy based NLP/NER model** for extracting text from resumes
- Rewarded with **Gold Star** for optimizing git branching and environment deployment process
- Deployed machine learning solution in production using **Docker**

Role: TEAM LEAD ARTIFICIAL INTELLIGENCE

Responsibilities:

- Got the opportunity to manage team of AI Engineers
- Placed in a customer facing role. Developed whole AIOCR based solution for client using **Tesseract** and **CNN** model. Also got the chance to use **AWS** services like **DynamoDb, S3, IAM role and SQS**
- Learned Docker and containerized several ML applications for the deployment.

Role: ARTIFICIAL INTELLIGENCE DEVELOPER

Responsibilities:

- Since it was a startup, I got very good exposure to different technologies, frameworks and tools like **Tableau 10, TensorFlow, Keras, Pandas, Numpy, Matplotlib, SciPy, Spark** etc.
- Developed Lobby Manager for Indian railways. For image processing function I used **Convolution Neural Network** in this project

Role: PROGRAMMER ANALYST

Responsibilities:

- SharePoint portal development and management for Client
- Single handed managed server hosting and support work for this project
- Used C#.net, Visual Studio 2010 with SharePoint 2007 and SharePoint 2010
- Got a very good understanding of Agile Methodology
- Migration from SharePoint 2007 to SharePoint 2010

Role: TRAINEE

Responsibilities:

- Developed File management system development using .net framework (C#, SQL Server)

KEY PROJECTS

AHI

- It was one of the recent project where problem statement was to create a bot which can answer end user query from the data stored on Azure SQL database.

THERAGRAPH

- In this project, health record history of patients had to be displayed on a dashboard. To build this product, a hybrid approach was used where classical AI, OCR, LLMs have been incorporated. Hosted LIAMA2 model on-prem for health record processing.

IDP

- It's a proprietary product of Damco Solutions. To build this product, a hybrid approach is used where classical AI, OCR, LLMs have been incorporated.
- This IDP tool is able to process any kind of document irrespective of it's type, format and layout. This product allows user to process unstructured documents in bulk.
- I played the key role in designing and creating this whole solution.

ASSIST AI

- It's a proprietary product of Damco Solutions. To build this product, I used latest Generative AI techniques such as LLMs and Prompt engineering.
- This product allows user to query organizational knowledgebase through conversational APIs.
- Generative AI models - GPT 4 and GPT 3.5 Turbo, BERT and PaLM 2,
- Framework - PyTorch, BeautifulSoup for scraping
- Language - Python
- Cloud - AWS
- I use Pipeline (CD).
- This was done for a short term loan lending organization - NCR Finance.

PORBIS

- Implemented solution for processing different kinds of reports through a tool which process and store desired information in tabular format.
- For solution AWS Textract service was leveraged and for identifying the target location of report data, LSTM model was trained and used.
- Optimized code for better performance and to implement best practices.

MLOPS DEPLOYMENT

- Created MLOPs CI/CD pipelines for faster and accurate machine learning experiments on Azure Machine Learning platform.
- Three pipelines were built - Infrastructure as Code Pipeline, Training Pipeline (CI) and Release Pipeline (CD) using Azure DevOps services.

PORBIS

- Implemented solution for processing different kinds of reports through a tool which process and store desired information in tabular format.
- For solution AWS Textract service was leveraged and for identifying the target location of report data, LSTM model was trained and used.
- Optimized code for better performance and to implement best practices.

CLAIMATIC

- Predict Staffing requirement for next one week based on historical claim and weather data.
- Analyze Nine years legacy Big Data and find patterns which can improve efficiency of Intelligent Claims Automation process

- Amazon Elastic Map Reduce was used to process data from AWS data lake and for ML implementation PySpark was used in EMR notebook

NCTEK

- Loan Prediction Model – Predict the outcome of loan based on applicant's bank statements.
- Bank Transaction Classifier Model – It classifies bank transactions in loan, payroll, govt security and other category.
- Loan amount prediction Model – The model predicts the amount to be given to the applicant as loan.

RESUME PARSING

- Implemented **Spacy based NLP/NER model** for extracting candidate's information from resumes

LENSIGHT

- An **AI and OCR** based project to extract data from scanned documents
- For OCR and AI, **Google Tesseract and CNN model** was used respectively. I leveraged **Tensorflow and Keras** libraries along with other libraries for model building. **AWS services** were also required for complete integration of the application

LOBBY MANAGER

- In this project my task was to identify employee based on the image provided
- I have used CNN model for this task. Through this project my image processing skills have improved. I did it using **Keras and Tensorflow**

CLARITY

- It was a web portal designed for Novartis. I had used **SharePoint server 2007** for initial development and later migrated it to SharePoint Server 2010

AWARDS & RECOGNITION

- Performance Excellence Award, Customer Service Excellence Award, Gold Star, 3 Silver Star from Damco
- Award of Excellence for OCR and AI project from pOrbis
- Mindtree's Got Talent Award