

Smita Paul

Immediate joiner • Pune Maharashtra • +91 93227 55438 • Smita.Paul.2023@gmail.com

- **Experienced ML Engineer** with 13 years in tech and 6+ years specializing in **ML models**, data visualization, and **data analysis**
- Excel in impactful **communication** and **collaboration** with stakeholders and key users, with outstanding **problem-solving** skills
- Consistently integrates the **latest AI trends** into model development to stay ahead in the field, fostering a continuous learning.
- Extensive background in ML lifecycle , NLP, LLM , and **cloud** platforms to drive impactful projects and technological capabilities.
- Completed **Generative AI** with Large Language Models by **DeepLearning.AI**

Skills

- **Programming Languages, Framework:** Python , PySpark, Sql, NoSql , , TensorFlow, PyTorch, Streamlit, Flask
- **Machine learning:** XgBoost, Random Forest, Logistic regression , KNN, K means clustering , Ensemble, Regularization
- **Generative AI and NLP Technologies:** Open-source and paid LLM models (Llama2, Mistral, OpenAI, Google Gemini Pro), LangChain, LlamaIndex, BERT , Word2Vec , Glove
- **Big Data Analytics:** ETL , HIVE , Hadoop, Spark, Pipeline, Data warehouse , Data lake
- **Vector Databases:** ChromaDB, Pinecone
- **Soft Skills:** Analytical thinking, problem-solving, effective communication, Stakeholder Management , customer engagement
- **Cloud Technology:** Databricks, Azure – ADF, AML, Azure cognitive search, AWS – S3, Glue, Lambda, Athena, Sagemaker

Experience

Sr. Data Scientist | Hexaware Technologies | Oct 21 – June 2024

Credit Risk Modelling – Classification project on Azure

- **Led** and **managed** the end-to-end machine learning pipeline (mlflow) using various ML Algorithm. Worked on **feature** selection , importance, **model explainability** and streamlined deployment processes.
- Application developed by **Pyspark** in Azure Machine Learning Studio - **AML** and Data factory – **ADF**, storage accounts
- Worked on EDA, feature development via **descriptive statistics**, data visualization, training, **fine-tuning**, development, and deployment of explainable, robust and high-performance AI pipelines

Agent chat processing – NLP - LLM Project

- Led the development of production-ready **Conversational AI** - GenAI solutions with LLMs, RAG , foundation models, prompt engineering, fine-tuning, semantic search and using vector databases
- Delivered the **production-ready GenAI** solutions with LLMs, RAG , **Mistral AI**, OpenAI , LLaMa 2
- Focused in NLP & LLMs on top of **RAG** architecture to enable highly accurate multimodal information retrieval to automate Document Question Answering, Document Search, Summarization, Translation, FAQ generation, etc. to significantly boost operational efficiency.
- Building **custom LLMs** for multimodal search and retrieval with **embedding, vector databases, evaluation**

Other task:

- **Collaborated** with stakeholders and business leaders to understand requirements, integrate data sources, productionized ML models, manage databases and data lakes, and **facilitate decision-making** in an Agile environment
- Built, maintain **ETL pipelines** using python, **hive, pyspark**, ADF, handled **scheduling of jobs** till production environment.
- Specializing in analytical skills , handling large datasets , Mlops, Data Mining, Text Analytics and EDA.
- Created and maintained a comprehensive **data governance** program that ensured compliance with industry regulations and improved data quality by 25%

Data Scientist consultant | eClerx Technologies | Nov 2019 - Oct21

Keyword Extraction – NLP – Client: Internet provider

- Automated the **text extraction** and **highlighted** of key information from chat, reducing the manual effort required.
- Used NLP techniques via transformer. Implemented the BertForSequenceClassification for sequence classification.

Clickstream data processing – Client: Top shoe company

- Successfully leveraged **clickstream** data to understand **customer behaviour** and **segment customers** by clustering.
- Implemented in AWS, **Sagemaker**, PySpark, EMR, Glue, S3 to segment customers behaviour.

Risk Rank analysis - Client: banking

- Developed and utilized ML model for Risk Rank analysis on Trade handling and enhanced the accuracy by > 45% by determining trade severity.

Other task:

- Collaborated for deploy ML , NLP models into production environments, enabling automated decision-making and predictive analytics.
- Worked on AWS based clients by building end to end enterprise big data ingestion ETL\ELT and processing solutions using PySpark and SQL complex transformations, EMR, Glue, S3

Cognizant Technologies | Jan 2014 – Apr 2017

Technical Project Lead| Jan 2014 – Apr 2017

- Led a team for **data analytics** in the design and execution of a data warehousing project, resulting in a 50% reduction in data processing time and a 20% improvement in data accuracy.
- **Architect**, design, developed and **recovered 100% transaction** failures in database rollback, redesign **DB transaction**
- Built **PowerBI**, visualization **dashboards** and charts informing customer insights, improving data visibility by 25%
- Optimized database query response time by 50% through the **optimization** of indexes and database schema design, resulting in improved application performance and customer satisfaction.
- Decreased application downtime 60% by implementing multiprocessing report generation (**Thunderhead** and SQL).
- Thoroughly reviewed user stories, participated in analysis and estimation, Sprint planning, daily scrum, as well as review and retrospective meetings

Sr. Software Engineer | Cognizant Technologies |Fintech, Healthcare, Ecomm Domain | Apr 2010 – Dec 2014

- Experienced as Java developer with a strong emphasis on **code quality**, crafting clean, well-structured code that balances **readability**, **maintainability**, and performance. Solutions contributed on documentation of **HLD** and **LLD**.
- Led the analysis, detailed design, and functional specs preparation. Contributed to **technical architect**, coding enhancements, performed unit tests, and managed the test cycle for seamless deployment.

Software Engineer | Cognizant Technologies |Fintech, Healthcare, Ecomm Domain | Nov 2007 – Apr 2010

- Developed web application for banking domain, which was keeping track of trades. Technology used are Java, Sql.
- Worked on windows application using sql

Education

Masters On Computer Management | Pune University| 2001-2003