

# Ms. Sadiya Farheen

AI-ML Engineer

+91 - 9845600555; +44 7776683182

sadiya.farheen005@gmail.com

LinkedIn

Exhibit a proactive and impactful approach as a Data Science expert, offering diverse experience in Advanced Analytics, Machine Learning, and Natural Language Processing, aimed at driving innovation.

## Education

- 2023: M.Sc. in Data Science from University of Glasgow, Glasgow G12 8QQ, United Kingdom
- 2017: B.E. in Information Science and Engineering from PES University, Electronic City Campus, Bangalore

## Areas of Expertise

Predictive Analytics

Data Mining

Statistical Modeling

Cloud Computing Fundamentals

Data Governance

Business Intelligence

User Experience Design

Data Quality Assurance

Agile Project Management

Digital Transformation

## Technical Skills

- Advanced Machine Learning
- Natural Language Processing (NLP) Applications
- Generative AI and Transformers
- Deep Learning Architectures (Neural Networks) and Computer Vision
- Apache Spark for Big Data Processing
- Python Programming (Pandas, NumPy, PyTorch)
- Data Visualization (Matplotlib, Seaborn)
- SQL Database Management (MySQL)
- Java for Software Development
- Version Control Systems (GIT)
- MS Office Suite

## Profile Summary

- Possess 4 years of professional experience, with 3 years dedicated to Data Science, specializing in Artificial Intelligence and Machine Learning, while consistently striving to provide effective and transformative data-driven solutions.
- Working on the position of AI-ML Engineer at Diverse Lynx India, actively engaged in creating cutting-edge solutions for the Aadhaar project, enhancing user experience.
- Gained a wealth of knowledge and skills throughout the career, including advanced proficiency in Machine Learning algorithms, Deep Learning, data preprocessing techniques, and the application of NLP to solve complex business challenges.
- Created and deployed an advanced chatbot tailored for Aadhaar inquiries, leading to a notable enhancement in user engagement and satisfaction through immediate assistance and efficient document retrieval.
- Expert in Generative AI, Transformers, Apache Spark, Python (utilizing Pandas, PyTorch, NumPy, Matplotlib, and Scikit-Learn), Java, and MySQL.
- Displayed proficiency in constructing retrieval-augmented generation pipelines, applying sophisticated methodologies to improve user interactions and optimize information retrieval workflows.
- Contributed to significant improvements in operational efficiency by performing data analysis, preprocessing, and model building, leading to a 96% quality boost in service performance and user satisfaction metrics.
- Showcased exceptional analytical and problem-solving capabilities, alongside proficient communication and collaboration skills, fostering teamwork to achieve project goals and drive successful outcomes.

## Work Experience

Diverse Lynx India Pvt. Ltd. | Bengaluru

Apr'24 - Present

Designation: AI-ML Engineer (Client: UIDAI-Aadhaar)

### Responsibilities:

- Working on the design and development of an intelligent Chatbot, aimed at providing users with immediate assistance for Aadhaar-related inquiries, thereby enhancing user engagement and satisfaction.
- Implementing real-time data retrieval mechanisms that allow users to download Aadhaar documents and check their enrollment status seamlessly.
- Transforming unstructured PDF documents into structured data formats, cleaning and organizing data to ensure accuracy and usability.
- Constructing a sophisticated Retrieval-Augmented Generation (RAG) pipeline that integrates FAQs and policies, utilizing memory buffers to enhance user interactions.
- Designing a modular workflow with specialized agents, integrating Aadhaar APIs to handle user requests. LLaMA 3.1 80B powers decision-making, ensuring seamless automation and efficient task orchestration.
- Developing a Computer Vision model for face matching by utilizing MTCNN and OpenCV for image preprocessing, ensuring document quality for verification.
- Applying RESNET50 to extract and compare embeddings from authorized documents and enrollment photos, using GAN techniques (SRGAN, ESRGAN) to enhance blurry images for fraud detection.
- Collaborating with cross-functional teams to ensure the successful deployment of Machine Learning solutions, aligning project goals with organizational objectives.

### Achievements:

- Enhanced the accuracy of document verification processes, achieving a fraud detection rate of 92% through the implementation of advanced Computer Vision techniques.
- Streamlined data processing workflows, reducing document retrieval times, thereby improving overall user satisfaction.
- Recognized for outstanding contributions to the Aadhaar project; received commendations from senior management for problem-solving and project execution.

## Certification

- ③ Designing and Implementing a Microsoft Azure AI Solution | Microsoft | Pursuing
- ③ AWS Certified Solutions Architect – Associate Certification | Creedly | Jan'24
- ③ The Complete Prompt Engineering for AI BootCamp (2024) | Udemy | Nov'24
- ③ Master the Coding Interview: Data Structures + Algorithms | Udemy | Jun'21
- ③ Complete Machine Learning and Data Science Bootcamp | Udemy | Jun'21
- ③ Complete Python Developer in 2023: Zero to Mastery | Udemy | Jun'21

## Soft Skills

Communication



Adaptable



Time Management



Analytical



## Personal Details

**Date of Birth:** 1<sup>st</sup> January 1995

**Languages Known:** English, Hindi, Kannada, Urdu

**Address:** Bangalore – 560095

## Accenture | Bengaluru

Dec'17 – Mar'21

**Designation:** Machine Learning Engineer and Network Engineer

### Responsibilities:

- ③ Conducted comprehensive data analysis and verification across diverse datasets, utilizing Python and SQL to ensure the integrity and accuracy of infrastructure assets.
- ③ Designed and implemented robust data pipelines for FTTx network systems, enhancing operational efficiency and data accessibility for stakeholders.
- ③ Developed and optimizing Machine Learning models for sentiment analysis, employing advanced classification and clustering algorithms to derive actionable insights.
- ③ Supervised all phases of data collection and preprocessing, ensuring high-quality data input for model training and evaluation processes.
- ③ Executed troubleshooting and stress testing on Machine Learning applications, resulting in substantial performance enhancements and reliability improvements.
- ③ Extracted and presented actionable insights from data analysis, supporting data-driven decision-making for key stakeholders.
- ③ Created interactive data visualization reports using Matplotlib, effectively communicating complex findings to non-technical audiences.
- ③ Collaborated with cross-functional teams to address infrastructure-related challenges, integrating Machine Learning solutions into existing systems to drive innovation.

### Achievements:

- ③ Received **Trailblazer Award** in the **Team Category in Sep'19** for dedication to delivering high-quality solutions and impacting client value creation at Accenture PLC.
- ③ Achieved an impressive **90% improvement** in operational efficiency and service quality through the successful implementation of Machine Learning solutions.
- ③ Developed a sentiment analysis model that outperformed existing benchmarks, **achieving an accuracy rate of 86%** and providing valuable insights for business strategy.

## Projects

### Fashion Forward: Multi-Model Recommendation System for Amazon Clothing

**Technologies:** Python, Transformers (BERT), Similarity Metrics, Deep Learning, PyTorch, Numpy, Pandas, Transformers, Scikit-Learn

- ③ Developed a hybrid recommendation system for Amazon clothing using collaborative filtering (SVD, KNN, NMF) and content-based filtering (Cosine Similarity, Euclidean Distance, SBERT, RESNET50). Optimized the model with a hybrid approach, achieving an nDCG score of 0.2378.

### Sentiment Analysis: Text as Data for Amazon Food Reviews

**Technologies:** Python, Pandas, Numpy, Scikit-Learn, BERT (Roberta), NLP, ML

- ③ Managed sentiment analysis on Amazon Food Reviews using K-Means clustering and classifiers (Logistic Regression, SVC, Random Forest) with BERT (RoBERTa) context-vectors. Achieved 88.7% accuracy and 69.4% F1 score, outperforming other methods.

### Big Data: Text Search and Filtering Pipeline with Apache Spark

**Technologies:** Apache Spark, Java, Big Data

- ③ Built an Apache Spark pipeline for efficient text search and filtering across large datasets. Implemented DPH ranking and advanced preprocessing, processing 10 GB of data in 50 seconds.

### Deep Learning for Cell Nuclei Classification

**Technologies:** Python, PyTorch, Matplotlib, Numpy, Deep Learning, Pandas

- ③ Developed CNN models (custom and ResNet18) in PyTorch for cancer cell nuclei classification, achieving 97.66% accuracy in Kaggle competition images.