

NIKHIL KUMAR SHARMA

Data Scientist
Reliance Jio, Bengaluru

+91-9968-944-825
nikhil.sharma1294@gmail.com
[LinkedIn](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

TOTAL EXPERIENCE: ~7 years

Results-driven Data Scientist with ~7 years of experience in Machine Learning, NLP, and Deep Learning and Large Language Models (LLMs). Proven expertise in Telecom, Banking, Automobile, and Information Security domains. Skilled in end-to-end project delivery, from data gathering to deployment, driving business growth with cutting-edge AI technologies. Experience with OpenAI, Azure AI, Foundation Models.

SKILLS AND TOOL SETS

Languages & Framework	Python, LangChain, Pytorch, TensorFlow, Unsloth,etc.
Core Competencies	LLMs, SST, TTS, Supervised, Un-supervised, Classification, NLP, Deep Learning, Statistical Model, Scripting and Automation
Data Visualisation	Power BI, Matplotlib, Seaborn, Plotly.
Databases	SQL Server, MySQL, PostgreSQL, MongoDB.
Tools & Software	Azure, OpenAI, Jupyter Lab, Anaconda, Spyder, Git, Docker, Jenkins, VS Code
Big Data	Pyspark, Hadoop, S3 bucket, Blob storage
Libraries	OpenAI, HuggingFace, Transformers, Pytorch, Tensorflow, SpaCy, Gensim, OpenCV, Scikit-Learn, DASH, requests, regex.
Domain	Telecom, Banking and Finance, Internet Security, Automobile

EXPERIENCE

■ RELIANCE JIO

DATA SCIENTIST

Oct 2021 - Till Date

- **LLM-based Chatbot:** Developed a chatbot using Retrieval-Augmented Generation (RAG) method with OpenAI GPT models and also fine-tuning GEMMA and LLAMA models, to be used by its internal Jio International Roaming Plan Consultants.
- **STT (Speech-to-Text) for customer support (To replace Azure STT)**
Designed and implemented a custom STT solution by fine-tuning **Whisper-tiny** and **Whisper-large-v3** models to enhance transcription accuracy and reduce latency for real-time customer support interactions.
- **TTS (Text-to-Speech) for Jio Translate Platform**
 - Developed a high-quality TTS system leveraging **MetaVoice** and **RVC** models to generate natural and expressive speech output.
 - Developed a pipeline for **Voice Cloning Platform** utilizing the **Parler-TTS** model, enabling personalized voice synthesis with high speaker similarity.
 - Integrated advanced audio processing modules, including **RE** (reverberation effects) and **UVR** (Ultimate Vocal Remover) for voice separation and noise reduction, enhancing audio clarity in diverse acoustic environments.
- **Implemented network monitoring dashboards to identify anomalies and optimize eTilt.**
- **Connected Components**
Clustering of telecom affinity data on the geospatial region to get the data segregated cluster to work on eTilt-optimisation.
- **eTilt Optimisation (cvxOpt)**
Implemented a custom in-house developed optimization algorithm to optimize the e-tilts of the cells on the network tower with the sole aim to improve the user’s network experience.
- **Anomaly Detection (Network Security)**
Build an anomaly detection end-to-end pipeline to detect port scans with a negative intent, which can come from within or outside of the network.

Tech Stack Used: Pyspark, spectral clustering, Jupyter-lab, data transformation, HDFS, Git, Kepler-maps, etc.

Mercedes-Benz Research & Development India (MBRDI) – Global Software Engineering Platform (GSEP)

- Created NLP models for spelling correction and chatbot performance dashboards.
- Developed price simulation dashboards for Mercedes-Benz using Azure & Plotly Dash.
- Automated Data Extraction Pipeline with Jenkins and Postgres SQL | Cron Job triggered daily at 00:00

Tech Stack Used: Python, NLP, Plotly, Dash, SQL on Azure, Azure Services, Azure Devops, Spello Language Model (based on SymSpell-Fuzzy Search), Postgres dB, Docker, Network error handling.

ANZ – Bank Guarantee for Trade & Finance – Digital Transformation

- Built document classifiers & NER models for ANZ Bank, automating form processing by 80%.
 - Classification of Bank Documents (ex. BG, Request Forms, ASIC, CLAA, Email, Others).
 - Extraction of multiple entities like Applicant Name, Beneficiary Name, ABN/ACN Numbers, Issue Date, Expiry Date, Purpose etc. from Request forms, Bank Guarantees and Emails.

Processes: Abby + Data Cap OCR, XML Parsing, Text Cleaning, data preparation, Text Analytics, Feature Engineering, Algorithm Selection, data tagging, model building, model training, and model validation.

- Automated Form Processing with AI
 - Insertion of specific entities if they are present and their location when those entities are not present in the Bank Guarantee Document Template of ANZ or of its clients.
 - Tech Stack: LSTM + Rule Engine + Regex + Logic-based solution to automate the form-filling process by the Maker and Checker that reduced the time required by 80% on an average.

Capgemini Centre of Excellence (COE)

- Contributed to Cognitive Document Processing (CDP) platform with OCR & NLP capabilities.
OCR, Text Extraction, Text Classification, Language Detection, Natural Language Generation.
- PII Redaction: Developed a tool for automated PII masking in legal documents, ensuring compliance with GDPR.

- Automated domain fingerprinting and server audits.
- Developed DDoS prevention modules and web scraping pipelines.

EDUCATION

Graduation	HMRITM, Delhi (B. Tech)	2012 – 2016
12 TH	Kendriya Vidyalaya, Delhi (CBSE)	2011 – 2012
10 TH	Kendriya Vidyalaya, Delhi (CBSE)	2009 – 2010