|  |  |
| --- | --- |
| **NIKHIL KUMAR SHARMA**  Data Scientist  Reliance Jio, Bengaluru | [**+91-9968-944-825**](tel:+91-9968944825)  [**nikhil.sharma1294@gmail.com**](mailto:nikhil.sharma1294@gmail.com)  **[LinkedIn](http://linkedin.com/in/itsnikhilksharma" \o "Go to Nikhil's LinkedIn Profile) |** [**Portfolio**](https://nikhilkrsharma.github.io/) |

**PROFESSIONAL SUMMARY TOTAL EXPERIENCE:~7 years**..

Data Scientist with ~7 years of experience in LLM, Machine Learning, NLP, and Deep Learning. Expertise in Telecom, Banking, and Information Security. Skilled in end-to-end project delivery, from data gathering to deployment.

**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, Jupyter Notebooks & Lab, Anaconda, Spyder, Git, Docker, Jenkins, VS Code |
| Big Data | Pyspark, Hadoop, Spark, HDFS, S3 bucket, Blob storage |
| Libraries | Numpy, Pandas, Scipy, Scikit-Learn, Requests, DASH, BeautifulSoup, TextBlob, re, NLTK, SpaCy, Gensim, OpenCV. |
| Domain | Telecom, Banking and Finance, Internet Security, Automobile |

**EXPERIENCE** .

**🞒 RELIANCE JIO DATA SCIENTIST** *Oct 2021 - Till Date.*

* **Developed RAG-based chatbot, fine-tuned GEMMA & LLAMA models, evaluated using BLEU & ROUGE scores.**
* **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)**

Implementated 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.

**🞒 COGNIZANT DATA SCIENTIST - CONSULTANT** *Sep 2020 – Sep 2021.*

**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.

**🞒 CAPGEMINI DATA SCIENTIST - CONSULTANT** *Nov 2018 – June 2020.*

**ANZ – Bank Guarantee for Trade & Finance – Digital Transformation**

* **Multi-Class Document Classification | Text Classification**

Classification of Bank Documents e.g., BG, Request Forms, ASIC, CLAA, Email & Others coming in various file formats.

* **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.

* **Developed PII-redaction tools for secure document handling.**

**🞒 CONFIDENTIAL PYTHON DEVELOPER** *April 2018 – Nov 2018.*

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

**EDUCATION** .

* [Graduation](http://www.hmritm.ac.in/) MAE HMRITM, Delhi 72 % 2012 – 2016
* [12TH](https://no2delhicantt.kvs.ac.in/) Science + CS Kendriya Vidyalaya, Delhi (CBSE) 75 % 2011 – 2012
* [10TH](https://no2delhicantt.kvs.ac.in/) Regular Kendriya Vidyalaya, Delhi (CBSE) 81.7 % 2009 – 2010