NIKHIL KUMAR SHARMA

Manager - Data Scientist
Reliance Jio, Bengaluru

+91-9968-944-825
nikhil.sharma1294@gmail.com
LinkedIn | **Portfolio**
Notice Period: 45 Days

PROFESSIONAL SUMMARY   TOTAL EXPERIENCE: 6 years

Experienced Data Scientist specialized in Machine Learning, NLP and Deep Learning. Proven expertise in _Telecom, Banking, Automobile and Information Security domains_. Strong in end-to-end project handling, from requirement/data gathering to development to deployment. Continuously learning new technologies to efficiently complete projects and drive business to success.

SKILLS AND TOOL SETS

Languages & Framework	Python, Pyspark, Keras, HTML, CSS, Flask, FastAPI, Streamlit, etc.
Machine Learning Algorithms	Clustering, DNN, CNN and RNN, LSTM, Linear, Polynomial, Logistic, Ridge and Lasso Regression, Decision Tree, Random Forest, k-NN, SVM, K-means
Core Competencies	Supervised, Un-supervised, Classification, NLP, Deep Learning, Statistical Model, Data Science, Scripting and Automation
Data Visualisation	Matplotlib, Seaborn, Plotly, Power BI, Storytelling.
Databases	SQL Server, MySQL, PostgreSQL, MongoDB, Oracle, SQLite
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, TensorFlow, TextBlob, re, Pytorch, NLTK, SpaCy, Gensim, OpenCV
Domain	Telecom, Banking and Finance, Internet Security, Automobile

EXPERIENCE

▣ RELIANCE JIO   DATA SCIENTIST | ML & BIG DATA ENGINEER   _Oct 2021 - Till Date_

- **Connected Components**
  Clustering of telecom affinity data on the geospatial region to get the data segregated cluster to work on eTilt-optimisation.

- **Network Monitoring**
  Implemented monitoring of key user cellular metrics (RSRP, CQI, TA) at city level, identifying anomaly clusters and optimizing eTilt; created a dashboard for data visualization, enabling strategic decision-making in network optimization.

- **eTilt Optimisation**
  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.

- **Credit Risk Modeling**
  Engineered a proprietary credit score model using Telecom data, creating customized scorecards for all partners; developed an advanced data workflow for EDA, feature generation, and scorecard formulation, leveraging WoE and IV techniques.

- **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)

- **Spelling Correction Model [NLP]**
  Developed and deployed a spelling correction module for the chatbot (RASA).
  Challenge: Domain Specific words, shorthand, URLs and email IDs.
  Process Flow: Data collection, transformation, model training, serialization & other supporting scripts.

- **Dashboard – Price Parts Simulation Dashboard – Price Demand Elasticity – [Trucks]**
  Show KPIs and Price Simulations based on various selections like price change, volume change, and discount group for a selected market and part type. Also displaying the break-even data in a tabular format based on the user's selection.
- **Chatbot Dashboard**
  Development and deployment of a comprehensive Admin dashboard showcasing KPIs essential for chatbot performance evaluation. These KPIs will encompass metrics such as chatbot usage, session time, response ratings, top-performing intents, CISM ticket volume and departmental penetration by the chatbot.
- **Data Extraction Pipeline with Jenkins | Cron Job triggered daily at 00:00**
  Data extraction from Postgres, sourced from RASA's 'events' table, is followed by transformation into structured data and storage in a different table.
- **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 (ex. BG, Request Forms, ASIC, CLAA, Email, Others)

  Process Flow: Data Mining, Abby + Data Cap OCR, XML Parsing, Text Cleaning, Text Analytics, Feature Engineering, Algorithm Selection, Model Validation and Optimization.

- **Named Entity Extraction using LSTM**
  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.
  Process Flow: Abby + Data Cap OCR, XML Parsing, Text Cleaning, data preparation, 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)
- **Cognitive Document Processing (CDP) Platform**
  Capabilities: OCR, Text Extraction, Text Classification, Language Detection, Natural Language Generation.
  Document Types: Driving License, W9 Form, Balance Sheet, Contractual Form, Aadhar Card, Pan Card etc.

- **PII-Redaction** (Personal Identifiable Information censoring)
  A solution is used to censor the PII present in any document, either in the image or pdf format and then generate the same document in the same format with all the PII redacted.

◼ CONFIDENTIAL	PYTHON DEVELOPER	*April 2018 – Nov 2018*

- Automation of Domain Fingerprinting
- DDoS Attack Module & Prevention Module
- Application Testing
- Server Audit Automation for Auditing (LAMP Stack)
- Web Scrapping and Data Wrangling

EDUCATION

-	Graduation	MAE	HMRITM, Delhi	72 %	2012 – 2016
-	12TH	Science + CS	Kendriya Vidyalaya, Delhi (CBSE)	75 %	2011 – 2012
-	10TH	Regular	Kendriya Vidyalaya, Delhi (CBSE)	81.7 %	2009 – 2010