

# NIKHIL KUMAR SHARMA

Manager - Data Scientist  
Reliance Jio

+91-9968-944-825  
[nikhil.sharma1294@gmail.com](mailto: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
Machine Learning Algorithms	Clustering, DNN, CNN and RNN, LSTM, Linear, Polynomial, Ridge and Lasso Regression, Logistic 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
Libraries	Numpy, Pandas, Keras, Scipy, Matplotlib, Seaborn, 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	25/10/2021 - Till Date
-	<b>Connected Components</b> Clustering of telecom affinity data on the geospatial region to get the data segregated cluster to work on eTilt-optimisation. Tech Stack: Pyspark, spectral clustering, data transformation, HDFS, Git, Kepler-maps, etc.	
-	<b>Network Monitoring</b> Monitoring of user cellular metrics like RSRP, CQI, TA, etc. throughout the day at a city or circle level for all the users in that selected region and to find the anomaly clusters. Built a dashboard for the same to present the data to the business. Tech Stack: Pyspark, jupyter-lab, Kepler-maps, git, etc.	
-	<b>eTilt Optimisation</b> Implementation of 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. Tech Stack: Pyspark, pandas, Kepler-maps, cvxopt, etc.	
-	<b>Anomaly Detection (Network Security)</b> 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: Pyspark, pandas, Jupyter-lab, git, etc.	
-	<b>Credit Risk Modeling</b> Created a credit score model using Telecom features and developed a scorecard for all the partners. Developed scorecard formulation using WoE and IV. Developed a data engineering workflow for various EDA and feature creation. Worked on various statistical methods for feature selection and model building.	

■ COGNIZANT	DATA SCIENTIST - CONSULTANT	09/09/2020 – 21/10/2021
<b>Mercedes-Benz Research &amp; Development India (MBRDI) – Global Software Engineering Platform (GSEP)</b>		
-	<b>Spelling Correction Model [NLP]</b> 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.

Tech Stack: Python, NLP, Spello Language Model (based on SymSpell-Fuzzy Search)

- **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.

Tech Stack: Plotly Dash framework, SQL on Azure, Azure Services, Azure DevOps.

- **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.

Tech Stack: Plotly Dash framework, Postgres dB, Docker, network error handling.

- **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: Python scripting, bash, Postgres dB, Jenkins (Artifactory), Docker.

## ■ CAPGEMINI

## DATA SCIENTIST - CONSULTANT

29/11/2018 - 30/06/2020

### 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

09/04/2018 - 30/11/2018

- 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
-	12 <sup>TH</sup>	Science + CS	Kendriya Vidyalaya, Delhi (CBSE)	75 %	2011 – 2012
-	10 <sup>TH</sup>	Regular	Kendriya Vidyalaya, Delhi (CBSE)	81.7 %	2009 – 2010