

RAJAT GUPTA

✉ rajat.y6@gmail.com

☎ +91-9986658010

📍 India

🌐 [in/rajat-gupta-72137516/](https://www.linkedin.com/in/rajat-gupta-72137516/)

Seasoned Data Science Professional with 13 years of experience in developing end to end Machine Learning Systems and Applications across Fintech, Insurance and Aerospace Domain. Proven track record of developing and executing on roadmap of data science projects across a spectrum of domains, encompassing Data Science, Machine Learning and Deep Learning.

Skills

Code	Python (Expert), PySpark (Expert)	DataOps	Dask, Airflow, Celery
ML/DL	Sklearn (Expert), XGBoost, Tensorflow, PyTorch	DevOps	Docker, Jenkins, Openshift, Kubernetes
Gen AI	OpenAI, LangChain, HuggingFace, Transformers	API	FastAPI, Flask, BentoML, TF-Serving
NLP	NLTK, Spacy, Gensim	Cloud ML	AzureML, SageMaker, Databricks, Cognitive Services

EDUCATION

[IIT KANPUR \(2006-11\)](#)

B. TECH – CIVIL ENGG. (7.7/10)

M. TECH – CIVIL ENGG. (9.2/10)

[IMT GHAZIABAD \(2021-24\)](#)

EXECUTIVE PGDM GEN. MGT. (8.6/10)

COMPETENCIES

Data Science & ML

ML System Design

Deep Learning

Natural Language Processing

Generative AI

MLOps, DataOps & DevOps

Credit Risk Modelling

Computer Vision

Data Engineering

Big Data Analytics

Cloud Computing

Python Full Stack Development

Data Quality & Governance

People Management

Stakeholder Management

IT Infrastructure Budgeting & Planning

WORK EXPERIENCE

[Paytm](#) (One97 Communication Ltd) – Lending

Jun 2022 - Present | Noida, India

Manager Data Science – Credit Risk & Collections

- Leading Data Science and Risk Analytics function to support 6+ partnerships of Co-Branded and other Credit Cards on Paytm Platform to generate **2.5 Million Leads and 70K+ cards per month**.
- Leading the ML development and strategy for Data Room Exercise (Joint Modeling) for Credit Cards to qualify **80 Million** Paytm Customers.
- Leading the ML development for [Paytm CreditMate](#) – Model base digital collections platform.
- Managing Internal and External Stakeholders from Business, Risk and Policy to deliver insights on New Cards Acquisitions, Portfolio Behavior and Risk Strategy.
- Leading the team of 4 Data Scientists to develop in collaboration with Data Engineering, Software Engineering and DevOps.

Technical:

- Leading the development of **Real time recommendation Engine for Credit Cards** based on Prediction of Risk Approval and Credit Policy to maximize the lead conversions.
- Neural Network** based model for multi-action recommendation for maximum repayment.
- Owner and Co-Creator of various models and data products: **Alternate Risk Scores**, Bureau based custom risk scores, **Approval Prediction scores**, Location stability score.
- Developed **short address** and **incorrect address detection models** using geo-coding and text-based features to reduce address related issues in Credit Cards Digital Journey.
- Co-created various large scale spark pipelines to create feature store of 3000+ features and model inferencing pipelines for 125 Million customers for more than **100 Billion Transactions / year** and **50 Million bureau reports** with feature history of 3+ years.
- Developed credit risk model based on **XGBoost** and **Logistic Regression** for unsecured lending products – Credit Cards, BNPL and Personal Loan using transaction behavior and credit bureau.
- Consuming and transforming various data sources across Paytm ecosystem – involving all online and offline spends, lending management system, travel and credit bureau.

WORK EXPERIENCE

Enterprise Data Analytics, [Optum](#) (UnitedHealth Group)

Nov 2018 – Jun 2022 | Gurugram, India

Manager Data Science

Text Extraction as Self Service (TESS) – Led 8 member team of Data Scientists, DevOps and ML Engineers.

- Computer Vision based cloud application using micro-services to extract user-defined entities from any templated form (pdf/image).
- Led MLOps, Cloud Ops, ML Serverless API and methodology for templated and non-templated forms.

Senior Data Scientist

Lead ML Systems Development working closely with Optum Business, Technology and Operations.

ML based Duplicate claim framework

- Realized Annual savings of \$500K by Real time prediction on 2.4 Million claims / year for 2 clients.
- Scalable, Extendible and reusable service consumed by multiple clients.
- Automated training and CI/CD pipeline for cloud deployment. Model monitoring for drift detection.

Medical Document Indexing

- Scanned Document review application (Search, Table of Content) using Computer Vision and NLP.
- Led Heading Detection using Ensemble of Bounding Box classification, Image Processing, Heuristics.

Self-Learning system for Provider Contract Rate Prediction

- Projected Annual Savings of \$200K by real time prediction on 150K claims / year.
- Continuously track and learn from manual decision. Explainable Rate Predictions: Linked to tracked data.

Customer Appeals Prediction

- Led development of ML system to **flag rare instances (0.01%)** of expected claim appeals due to processing **errors**: Estimated reduction of 6000 Appeals / year.
- Scalable system to daily scan 1 million claims, filter to 100k claims and flag up to 150 claims only.
- **Text mining** of call transcripts: Call Drivers detection (OptumRx)
- **Text Mining** of social media: Customer sentiments for UHG.
- **Q-Card**: Scalable ML assisted Data Quality toolkit using PySpark, Pandas and Sklearn

R2 Data Labs, [Rolls Royce](#)

Nov 2015 – Nov 2018 | Bengaluru, India

Data Scientist

Complex engineering problems using Data Science and Capability development with SME, Field Engineers

TRENT 700 HPT Sulphidation Investigations

- ML Model Development to address risk mitigation and fleet management of more than 1600 engines.
- PySpark Pipeline: 25 years data from engineering, operations, manufacturing and environment.

AI Hub Development

- Domain specific question answering system based on [GATE](#), [Rasa](#) and [AllenNLP](#)
- Text Mining and Document Structure Detection of RR Research Papers to ingest data in ElasticSearch.

Enterprise Solutions, [Altair Engineering](#)

Jan 2014 – Nov 2015 | Bengaluru, India

Technical Leader / Senior Software Engineer

Package Solution Offering (PSO) Developed: [Model Verification](#), [Automated Reporting](#) and [Squeak & Rattle Director](#) using Tcl-Tk, Python and Hyperworks.

Edison Engineering Development Program , [GE Aviation](#)

Jul 2011 – Jan 2014 | Bengaluru, India

Airfoils Design Tools: Software and Methodology Development of Finite Element Method based Tools for Bird/Ice Impact Analysis and Airfoil Mechanical Design using LS-Dyna, Ansys, Tcl-Tk and distributed and cluster computing using LSF.

Achievements

- **Currently ranked #1** in PGDM Executive Part Time Batch of 2021-24 at IMT Ghaziabad.
- **Winner**, Optum Global Hackathon 2020 – AI / ML (155 Teams).
- **2 Patents** Filed in the area of unsupervised text outlier detection and rare event prediction. Recognized as [UHG Inventor](#).
- **Best Green Belt** (six sigma) Project across GE Aviation Bangalore for 2012.
- **GE's first** Windows Based Cluster: LSF cluster (20 machines) to realize 20X speed for Design of Experiments.

Certifications

- Azure AI-900 Certification | 2021
- [Introduction to Machine Learning in Production](#) | Coursera – deeplearning.ai | 2021
- [Natural Language Processing with Probabilistic Models](#) | Coursera – deeplearning.ai | 2021
- [Introduction to Microsoft Azure Cloud Services](#) | Coursera | 2021
- [Neural Networks and Deep Learning](#) | Coursera – deeplearning.ai | 2020
- [Introduction to TensorFlow for AI, ML, and Deep Learning](#) | Coursera – deeplearning.ai | 2020
- [Natural Language Processing with Classification and Vector Spaces](#) | Coursera – deeplearning.ai | 2020
- [Data Scientist Masters Program | Simplilearn | 2017](#)
- Certified Scrum Master | Scrum Alliance | 2017
- **A-Course Graduate** [EEDP Program](#) | GE Aviation | 2013
- Green Belt | GE Aviation | 2012