

# Nikhil Premachandra Rao

+1 774 636 8916 · Atlanta, GA · [npraou@gmail.com](mailto:npraou@gmail.com) · [Linked in](#) · [Portfolio](#) · [Github](#) · [Tableau](#)

## SUMMARY

Data Scientist with 5+ years of experience delivering end-to-end data solutions across financial services and healthcare. Adept at building scalable ETL pipelines, real-time analytics platforms, and predictive models using Python, SQL, Airflow, Kafka, Snowflake, and Azure ML. Proven expertise in cloud-native data engineering (AWS, Azure, dbt), fraud detection, regulatory reporting, and self-service BI dashboards (Power BI, Tableau) that accelerate decision-making and improve operational efficiency. Skilled at collaborating with cross-functional teams to drive data governance, compliance (HIPAA, SOC 2, GDPR), and cost-optimized solutions that enhance client retention, risk management, and business growth.

## PROFESSIONAL SKILLS

- **Programming & Data Processing:** Python, SQL, R, Pandas, NumPy, PySpark, Bash, Shell Scripting
- **Machine Learning & AI:** Scikit-learn, XGBoost, LightGBM, PyTorch, TensorFlow, spaCy, BERT, SARIMA, Prophet, MLflow, Forecasting & Predictive Modeling, Anomaly Detection, NLP Pipelines
- **Data Engineering & Automation:** Apache Airflow, dbt, Delta Lake, Apache Kafka, ETL Pipelines (Batch & Real-Time), Data Cleaning & Transformation, Workflow Orchestration
- **Cloud & Big-Data Platforms:** AWS (S3, Redshift, Glue), Azure (Synapse, Data Factory, ML Studio), GCP (BigQuery, Dataflow), Snowflake, Databricks
- **Databases & Query Engines:** PostgreSQL, MySQL, MongoDB, HBase, Snowflake, Google BigQuery
- **Data Visualization & BI Tools:** Power BI, Tableau, Looker, Excel (Pivot Tables, Dashboards, DAX)
- **MLOps & CI/CD:** Docker, Kubernetes, MLflow, Git, GitHub, GitLab, Jira, Confluence, Automated Model Deployment
- **Analytics & Statistical Modeling:** KPI Tracking, A/B Testing, Forecasting, Time-Series Analysis, Hypothesis Testing, Data Governance & Quality Monitoring
- **Compliance & Governance:** HIPAA, SOC 2, GDPR, Data Cataloging, Schema Drift Monitoring
- **Soft Skills:** Communication skills, teamwork, Critical Thinking, strategic planning, problem-solving, and adaptability.

## WORK EXPERIENCE

### Data Scientist, JPMorgan Chase

Jan 2025 – Present | Remote, USA

- Automated regulatory ETL workflows using dbt, SQL, Python, and Apache Airflow, streamlining ingestion from diverse trading systems and cutting reporting cycle times for risk and compliance teams.
- Orchestrated cloud-native data pipelines on AWS (Glue, Redshift, S3) and Snowflake, boosting query performance by 35% and ensuring secure, governed access for portfolio and liquidity dashboards.
- Architected real-time analytics solutions with Kafka, Kinesis Data Streams, and Azure Stream Analytics, enabling instant detection of suspicious trading activities, market trend shifts, and fraud anomalies.
- Built forecasting and churn-prediction models using scikit-learn, Azure Machine Learning, improving investment product demand forecasting and client retention strategies.
- Designed interactive Tableau and dashboards to visualize risk exposure, liquidity, and transaction KPIs, eliminating 30% of manual reporting efforts and empowering business leaders with self-service analytics.
- Collaborated with compliance and risk partners to deploy automated anomaly-detection pipelines leveraging PySpark and DataFrame optimization, strengthening fraud-prevention frameworks and ensuring audit under guidelines.

### Data Scientist, HCL Tech

May 2022 – Jul 2023 | Remote, India

- Implemented Spark MLlib and PyTorch-based recommendation engines, boosting user engagement, basket size, and session duration, driving an 18% increase in sales conversions across digital channels.
- Orchestrated Airflow- and PySpark-driven ETL workflows integrating multi-source datasets into Delta Lake, reducing processing bottlenecks by 42% and enabling faster, more reliable real-time analytics.
- Built NLP-powered sentiment analysis classifiers using spaCy and BERT, automating customer feedback categorization, cutting response delays, and improving NPS by 21%.
- Designed Power BI dashboards to track churn risk, customer lifecycle metrics, and revenue drivers, enhancing strategic retention initiatives and improving client retention by 17%.
- Optimized LightGBM and XGBoost models with MLflow, improving predictive accuracy by 16% while ensuring reproducibility, seamless deployment, and adherence to regulatory standards.
- Developed an automated anomaly detection framework monitoring schema drift, missing values, and inconsistencies, increasing pipeline stability by 23% and strengthening enterprise data quality governance.
- Created SARIMA and Prophet forecasting models, enhancing seasonal demand prediction by 20%, which optimized marketing campaigns, inventory planning, and resource utilization.
- Collaborated with Agile squads using Jira, Git, and Confluence, accelerating ML pipeline delivery by 30% and enabling scalable production deployment with continuous improvement practices.

### Data Scientist, Oracle Cerner

Jan 2020 – Apr 2022 | Hybrid, India

(AdventHealth, Baptist Health, Christiana Hospital, Intermountain Health)

- Performed in-depth analysis of clinical and operational datasets using Power BI, uncovering key patterns and trends that informed strategic decisions for healthcare stakeholders.
- Engineered Python-based ETL pipelines to streamline data ingestion and reporting workflows, cutting manual efforts by 25% and improving accuracy and timeliness of insights for cross-functional teams.
- Developed predictive models leveraging historical healthcare data to enhance patient outcome forecasting and fault detection in hospital systems, improving prediction accuracy by 40%.
- Automated HBase table cleaning and SQL query optimization, reducing storage costs by 10% and significantly improving data retrieval speeds for client-facing reports.
- Led the development of an automated JIRA reporting framework, enabling real-time issue tracking and providing leadership with KPI dashboards for operational decision-making.
- Designed and delivered data catalog reports for over 2,500 healthcare clients, incorporating advanced filtering, sorting, and search functionality to improve client data accessibility and governance by 30%.
- Partnered with product and compliance teams to ensure HIPAA-compliant data handling and robust governance practices for all analytics deliverables.

### CMTI, Software Engineer Intern

May 2019 – Oct 2019 | Onsite, India

- Developed a Python-based inspection dashboard to monitor component deviations and ensure product quality.
- Automated generation of statistical reports and graphical analyses, reducing manual inspection time by 30%.
- Collaborated with quality engineers to integrate real-time data visualization and enhance process accuracy.

### Arcapsis, Freelance Data Scientist

May 2018 – Nov 2018 | Hybrid, India

- Designed and implemented 30% of an AutoML API for automated model selection and hyperparameter tuning using Python, Flask, and Scikit-learn.
- Containerized and deployed the system with Docker and Kubernetes, cutting manual tuning time by 25%.
- Integrated experiment tracking and model interpretability tools to prepare the platform for production deployment.

## CERTIFICATIONS

- **Machine Learning** June 2024  
*DeepLearning.AI, Stanford University* - Credential ID: [PCEFF94NBT9G](#)
- **Getting Started with Deep Learning** April 2025  
*NVIDIA, Deep Learning Institute* - Credential ID: [MOS7oiZMQcO9R966P906-Q](#)

## PROJECTS

### ModelMind – Intelligent Model Recommender for Regression, Apr 2025 – May 2025

- Designed an intelligent regression recommender that analyzes dataset feature signatures to identify the most suitable algorithms with confidence scoring.
- Implemented automated metadata logging and side-by-side ranking comparison between predicted and actual model performance (e.g., 92% confidence vs. 89%  $R^2$ ), improving model selection transparency and evaluation efficiency and optimized support for categorical and mixed-type variables, enabling scalable integration across varied regression workflows.

### DATA CHATBOT WITH LOCAL LLM | Python, Flask, Ollama

Mar 2025 – Apr 2025

- Built a real-time CSV analytics chatbot powered by a local LLM (Ollama) to handle natural language data queries securely within a Python-Flask web app.
- Integrated llama-3.2 inference engine to deliver 35% faster response times than comparable cloud APIs, maintaining <2-second response latency for both textual and visual outputs.
- Developed safe code execution modules for Pandas and Matplotlib, enabling seamless on-demand chart generation and data summaries.

### AUTO ML API Development | Python, Flask, Docker, Kubernetes

Aug 2024 - May 2025

- Developed a production-ready API automating ML model training, hyperparameter optimization, and deployment with Dockerized microservices orchestrated via Kubernetes.
- Reduced manual tuning effort by 25% through automated pipeline execution for algorithm selection and parameter sweeps, with built-in plans for interpretability dashboards and performance tracking.

## EDUCATION

### University of Massachusetts - Dartmouth, MA

Aug 2023 - Aug 2025

Master of Science (M.S), Data Science GPA 3.9

## LEADERSHIP & MENTORSHIP

### Secretary – Intelligent Modelling Club | University of Massachusetts

- Led discussions on student projects, reviewing modeling approaches and providing feedback on completeness and accuracy.
- Conducted ML and AI concept sessions 4 days a week, teaching core algorithms, modeling workflows, and hands-on experimentation to improve members' practical skills.
- Fostered a collaborative learning environment, helping peers gain confidence in building and deploying machine learning