

# Sreeja Kurapaty

+1 (480) 310-8212 | sreejakurapatydt@gmail.com | [Github](#)

## PROFESSIONAL SUMMARY

---

Machine Learning Engineer with 7 years of progressive experience delivering data science and analytics solutions across banking, healthcare, and insurance domains. Adept at designing and deploying end-to-end machine learning pipelines, from data ingestion and feature engineering to model training, evaluation, and integration into business workflows. Proficient in Python, SQL, PySpark, and Tableau, with hands-on expertise in predictive modeling, natural language processing (NLP), A/B testing, and statistical experimentation. Demonstrated success in driving measurable business impact through intelligent automation, personalization strategies, and scalable MLOps practices. Known for bridging technical depth with strategic thinking, collaborating cross-functionally to translate complex data into actionable insights and production-ready AI solutions.

## SKILLS

---

**Data Science & Analytics:** Data Analysis, Business Analysis, Business Analytics, Exploratory Data Analysis (EDA), Data Cleaning, Feature Engineering, Trend Analysis, KPI Reporting, Process Optimization, Statistical Modeling, Analytical Modelling, Strategic Planning

**Machine Learning & Deep Learning:** Supervised & Unsupervised Learning, Natural Language Processing (NLP), Transformers, BERT, Large Language Models (LLMs), Deep Learning, Neural Networks, Computer Vision, Information Retrieval

**Programming Languages & Frameworks:** Python (NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn), R, Java, C++

**Big Data & Distributed Systems:** Hadoop, PySpark, Apache Kafka, Docker, GPU Programming, GraphQL

**Web & App Development:** Flask, HTML, Web Scraping

**Data Visualization:** Tableau, Microsoft Power BI, Matplotlib, Seaborn

**Statistical & Analytical Techniques:** Hypothesis Testing, A/B Testing, Regression Analysis, Descriptive & Inferential Statistics

**Project & Quality Management:** Project Management, Leadership, Team Leadership, Quality Assurance, Quality Management

**Business & Domain Knowledge:** Finance, Healthcare, and Auto Insurance

**Tools & Platforms:** Microsoft Excel (Advanced), Microsoft Office Suite, Jupyter Notebook, VS Code, Git/GitHub

**Soft Skills:** Analytical Thinking, Critical Thinking, Problem Solving, Communication, Stakeholder Collaboration, Storytelling with Data, Management

## PROFESSIONAL EXPERIENCE

---

### INFOSYS

#### Machine Learning Engineer | Goldman Sachs

**Mar 2022 – Present**

- Built and productionized predictive models for user engagement and financial behavior segmentation, leveraging techniques like Random Forest, K-Means, and NLP pipelines.
- Spearheaded experimentation design with rigorous control/treatment matching and statistical power analysis, leading to 30% more reliable A/B testing across growth teams.
- Created reusable Python modules for data preprocessing, feature stores, and model deployment, reducing model deployment time by 50%.
- Delivered performance dashboards and ML explainability visualizations using Tableau and SHAP values to enhance model transparency for stakeholders.
- Partnered with cross-functional teams to translate ambiguous business goals into structured ML problem statements, leading to 3 successful production ML launches.
- Implemented MLOps practices by integrating version control (DVC/Git), model registries, and automated CI/CD pipelines using Jenkins and Docker.
- Advised product and marketing teams on leveraging model-driven insights, resulting in a 25% improvement in campaign efficiency and personalization.

- Led internal workshops on ML lifecycle best practices, enabling better collaboration between data scientists, analysts, and engineering teams.

## **COGNIZANT**

**Data Scientist| Aetna**

**Jul 2020 – Mar 2022**

- Developed and tuned customer churn models using Scikit-learn and ensemble techniques, supporting retention outreach that improved satisfaction scores across key segments.
- Engineered large-scale PySpark pipelines on AWS to prepare healthcare data for modeling, ensuring HIPAA-compliant transformations and data lineage tracking.
- Conducted causal inference analysis on communication strategies using uplift modeling and stratified A/B testing, guiding marketing redesigns.
- Built model validation dashboards for business stakeholders to monitor performance drift, model recalibration needs, and ROI metrics over time.
- Created ETL audit frameworks in Python to flag anomalies and maintain data integrity, resulting in a 40% drop in reporting defects.
- Worked closely with clinical, finance, and CX stakeholders to operationalize ML solutions that improved patient engagement and service delivery across digital platforms.
- Established scalable frameworks for experimentation pipelines with reusable code templates and modular design patterns.

## **Data Analyst| Travelers**

**Sep 2018 – Jul 2020**

- Conducted in-depth EDA and regression modeling to identify churn drivers and claim predictors, laying the groundwork for future ML models.
- Supported marketing and underwriting teams with campaign optimization by building rule-based scoring systems and testing hypotheses through pre/post analyses.
- Designed data pipelines in Python to ingest, clean, and aggregate data from Redshift, Oracle, and Snowflake into BI-ready datasets.
- Automated insurance policy performance reports and integrated anomaly detection scripts, reducing manual effort by 60%.
- Collaborated with DevOps and data engineers to migrate legacy data processes to the cloud (AWS), contributing to Snowflake warehouse implementation.
- Explored NLP-based classification techniques on customer feedback and call transcripts to detect early churn signals and service dissatisfaction trends.
- Delivered cross-functional presentations to communicate data insights to executives, legal teams, and marketing leaders for regulatory and strategic actions.

---

## **ACADEMIC PROJECTS**

**Customer Retention Modelling | Tools Used:** Python (Pandas, Scikit-learn, Matplotlib), Excel, Jupyter Notebook, Tableau

- Built a machine learning model to predict customer churn using a telecom dataset with **7,000+** records.
- Performed data cleaning, feature engineering, and exploratory analysis to uncover key churn indicators (e.g., contract type, monthly charges).
- Trained and evaluated multiple models achieving **81%** accuracy.
- Visualized churn trends using Matplotlib and shared actionable insights for customer retention.

**Retail Sales Dashboard | Tools Used:** SQL, Tableau, Excel

- Analyzed historical sales data from a fictional retail chain to identify top-performing products and seasonal trends.
- Wrote complex SQL queries to join and aggregate data across multiple tables (sales, products, stores).
- Created an interactive Tableau dashboard with filters by category, region, and date range.
- Presented findings to faculty panel, highlighting sales optimization strategies that could boost revenue by **12%**.

## EDUCATION

---

Master of Science in Data Analytics | Purdue University | USA

Bachelor of Technology in Computer Science and Engineering | Jawaharlal Nehru Technological University | India

## CERTIFICATIONS

---

- Google Data Analytics Professional Certificate | [Credential](#)
- Google Project management Professional Certificate | [Credential](#)
- Tableau Business intelligence Professional Certificate | [Credential](#)