

# Sayam Kumar - Data Scientist | Data Analyst

[sayamk565@gmail.com](mailto:sayamk565@gmail.com) | (437) 876-4544 | Toronto, ON | [LinkedIn](#) | [GitHub](#) | [Website](#) | [Kaggle](#) | [Tableau](#)

## OBJECTIVE

**Data Scientist** with 2+ years' experience in **machine learning, predictive analytics, and data visualization** across tech, manufacturing, and service domains. Skilled in **Python, SQL, Power BI, Scikit-Learn, TensorFlow, PyTorch, and cloud platforms** (Azure, AWS, GCP). Seeking data science, analytics & machine learning roles.

## CORE SKILLS

**Programming & Libraries:** Python, R, SQL, Scala, NumPy, Pandas, Scikit-Learn, TensorFlow, PyTorch

**Machine Learning & AI:** Classification, Regression, NLP, Clustering, A/B Testing, Feature Engineering, Computer Vision, Time Series Forecasting, LangChain, LangGraph, RAG, Vector Databases, Prompt Engineering, AI Agents, MCP

**Data Visualization & BI Tools:** Power BI, Tableau, Matplotlib, Seaborn, ggplot2, Plotly

**Data Engineering & Databases:** MySQL, PostgreSQL, MongoDB, SQL Server, PySpark, Hive, Hadoop, Databricks

**Cloud & MLOps:** Azure (ADF, Synapse), AWS (S3, EC2), GCP (BigQuery), MLFlow, Docker, CI/CD, FastAPI

**Analytics & Reporting:** Data Storytelling, Dashboard Design, Statistical Modelling, Causal Inference, KPI Tracking

## WORK EXPERIENCE

### AI Developer Intern

Toronto, ON

*Iter Innovandi*

Dec 2024 - Apr 2025

- Improved **data retrieval latency** by **30%** by building & optimizing FastAPI backend APIs integrated with **MongoDB**.
- Accelerated development cycles by **25%** by deploying **GenAI LangChain pipelines** to streamline AI application workflows.
- Enhanced team productivity** by integrating **AI-powered solutions** into internal processes, driving measurable efficiency gains across multiple projects.
- Strengthened system scalability and performance** through backend optimizations and **AI-driven process automation**.

### Data Scientist

Delhi, India

*Mechademy Incorporated*

Aug 2023 - Nov 2023

- Boosted fraud detection accuracy by **5%** by engineering NLP-driven ML pipelines using Python, Scikit-Learn, NLTK, XGBoost, and Hugging Face Transformers.
- Eliminated **32+ manual hours** weekly by automating ML lifecycle tasks via MLFlow and AutoML, streamlining end-to-end workflows.
- Reduced system downtime by **69%** via stable CI/CD integration with Docker, Kubernetes, & **Azure DevOps**.
- Optimized **ML workflow efficiency** by automating pipeline orchestration and monitoring with Prometheus & Grafana.

### Data Scientist

Bengaluru, India

*ExactSpace Technologies*

June 2022 - July 2023

- Designed **Azure-based ETL pipelines** using **PySpark** to process **100K+** industrial records daily.
- Deployed **anomaly detection models** to predict failures, **reducing unplanned downtime by 25%**.
- Built **Power BI dashboards** to monitor operational KPIs, improving **fault diagnosis speed**.

## EDUCATION

- Postgraduate Diploma – Business Insights & Analytics**

Toronto, ON

*Humber College, 2025*

- Bachelor of Technology – Computer Science Engineering**

New Delhi, India

*Guru Gobind Singh Indraprastha University, 2023*

- Certifications** – IBM Data Analyst Professional Certificate, IBM Data Science Professional Certificate, IBM Applied Data Science Specialization, Statistics with Python Specialization (Michigan University)

- Research Publications** – [Fake Reviews Filtering System using Supervised Machine Learning](#) (2022)

*IEEE International Conference on Data Science and Information System (ICDSIS)*