

# Sagnik Roy

Dearborn, MI (313) 652-5403

sagroy10@gmail.com, linkedin.com/in/sagnikroy10, sagnik-roy-portfolio.vercel.app

## EDUCATION

### University of Michigan-Dearborn

Master of Science in Data Science (GPA: 3.8)

Dearborn, MI

Expected Graduation: May 2026

Jan 2025 - Apr 2025

Graduate Teaching Assistant

Courses: Applied Regression Analysis, Computational Statistics, Deep Learning, Artificial Intelligence

Chennai, India

### SRM Institute of Science & Technology

Bachelor of Technology in Electronics and Communication (GPA: 9.41/10)

May 2022

Courses: Programming for Problem Solving, Pattern Recognition & Neural Networks, Probability & Stochastic Process

## TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, C/C++
- **Tools:** Excel, Tableau, Microsoft Power BI, Alteryx, Microsoft Azure, AWS, GCP (Dialogflow), Git, Jira
- **Data Science & ML Libraries:** Pandas, NumPy, Scikit-learn, XGBoost, TensorFlow, Keras, PyTorch, Matplotlib, Seaborn, Computer Vision, Hugging Face Transformers, Generative AI, RAG (Retrieval-Augmented Generation)
- **Certification:** Data Science and Machine Learning Program – Indian Institute of Technology Delhi

## WORK EXPERIENCE

### Ernst & Young - Global Delivery Services

Pune, India

#### Associate Software Engineer - Data & Analytics

Aug 2022 - Apr 2024

- Used Azure Databricks and Dataflow to build efficient data pipelines and perform advanced data transformations, streamlining the data processing workflow. Recognized by the client for strong performance
- Utilized Apache Spark on Azure Synapse Analytics to optimize ETL processes, improving data warehouse performance by 35%
- Worked hands-on with MySQL and Azure SQL databases, writing complex queries to improve performance while maintaining data integrity across environments
- Analyzed complex datasets using Alteryx and built 30+ Power BI dashboards that turned raw data into actionable reports. Received a spot award for exceptional contribution

### Ernst & Young - Global Delivery Services

Chennai, India

#### Data & Analytics Trainee

Feb 2022 - Apr 2022

- Developed a COVID-19 analytics model by using machine learning and exploratory data analysis for diagnosing COVID-19 status. Integrated the model with third-party APIs for real-time data and health information, and used Power BI to visualize and interpret the insights
- Led a team of four to deliver a complex analytics project on time, encouraging collaborative work environment and achieving strong results

### HighRadius

Bhubaneswar, India

#### Data Science & Product Intern

July 2021 - Mar 2022

- Built a banking chatbot using Google Dialogflow that combines NLP and machine learning to enable customer interactions. The tool improved customer engagement and streamlined key financial services
- Developed a machine learning-based invoice payment date prediction system, selecting XGBoost for its superior performance (94.76% accuracy) and delivering insights through a Power BI dashboard

## PROJECTS

### Hospital Readmission Prediction for Diabetic Patients

- Developed classification models to predict 30-day hospital readmissions, focusing on recall to better identify high-risk diabetic patients and support improved care outcomes
- Applied feature engineering, class balancing and SHAP analysis to improve model performance and interpretability on imbalanced clinical data from the UCI repository

### Generative AI-Powered Data Firewall for Data Pipelines

- Built a GenAI-powered data firewall that detects and automatically fixes schema drift, missing values, unseen categories, and anomalous inputs in real-time data pipelines
- Used LLM-driven reasoning to generate context-aware data corrections, improving data reliability and helping prevent downstream system failures