

# SOUMYA GOPALAPURAM

Milwaukee, WI | +1 414-595-9792 | [soumyagopalapuram11@gmail.com](mailto:soumyagopalapuram11@gmail.com) | [LinkedIn](#) | [GitHub](#)

## SUMMARY

Data Science graduate with 3+ years of industry experience in data analysis, backend development, and machine learning operations. Skilled in transforming raw data into actionable insights, building scalable RESTful APIs, and delivering data-driven solutions using Python, SQL, Java, and Data Visualization. Certified in Business Analytics with strong foundation in real-world project delivery and agile collaboration.

## EDUCATION

**University of Wisconsin-Milwaukee**, Milwaukee, WI

Jan 2024 - May 2025

- *Master of Science in Data Science, GPA 3.7/4.0*

- *Certificate in Business Analytics, Spring 2025*

**Stanley College of Engineering & Technology, Osmania University**, Hyderabad, India

Aug 2015 – May 2019

- *Bachelor of Engineering in Computer Science*

## TECHNICAL SKILLS

- **Languages & Libraries:** Python, Java, SQL, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, BeautifulSoup
- **Data Visualization & Analytics:** Tableau, Power BI, JMP, Excel, Pivot tables, VB macros, Power Shell, SHAP, SMOTE
- **Databases & Processing:** MySQL, Oracle, SQL Server, Apache Spark, Snowflake, Databricks, Power Query, Power Automate
- **Development & Frameworks:** Spring Boot, Spring MVC, RESTful API Development, SDLC, Postman
- **Tools & Platforms:** Visual Studio, PyCharm, Eclipse, STS, Git, JIRA, Microsoft 365
- **Core Competencies:** Data Analysis, Predictive Analytics, Business Analytics, Machine Learning, Data Integration, Data Modeling, ETL/ELT workflows, Data Wrangling, Data Warehousing, Information Security, Privacy & Continuity.

## WORK EXPERIENCE

**INFOSYS**, *Systems Engineer*, Hyderabad, India

Apr 2022 – Aug 2023

- Developed 20+ RESTful APIs using Java and Spring Boot to modernize a legacy monolithic system to scalable microservices architecture, reducing response time by around 70% supporting client integration needs.
- Enhanced database performance by optimizing vector database queries and SQL procedures in PostgreSQL, resulting in a 50% improvement in data retrieval speed and dashboard load times.
- Conducted functional and regression testing using Postman, leading to a ~30% reduction in production-level issues across multiple sprint cycles and ensuring reliable deliverables.
- Balanced development and production support across 10+ Agile sprints, consistently delivering features on schedule, addressing ~15+ weekly ad-hoc change requests, and conducting peer code reviews to uphold code quality.
- Maintained Git-based version control and implemented CI/CD best practices to streamline deployment cycles, decreasing rollbacks and build failures while reinforcing systems implementation excellence.

**AMAZON**, *Machine Learning Data Associate*, Hyderabad, India

Feb 2020 – Apr 2022

- Analyzed operational data from 50+ autonomous robots to assist in real-time issue resolution and root-cause reporting, contributing to a 15% reduction in system downtime across multiple fulfillment and bolstered business analytics for client operations.
- Validated machine learning software updates through structured A/B testing and regression scenarios, identifying and documenting edge cases to improve model robustness in live environments.
- Partnered with data scientists and engineers to annotate datasets, review model outputs, and perform quality checks, accelerating the refinement of computer vision and navigation models to enhance service delivery.
- Served as interim Subject Matter Expert (SME), enforcing SOPs that improved task consistency across a 25-member team and reduced process errors by 40%.
- Built performance-tracking dashboards in Excel to support weekly reporting, enabling data-driven operations management and continuous team performance monitoring.

## PROJECTS & ACTIVITIES

**Steam Game Recommendation System**, *Class Project, Big Data in Business*

Sep 2024 – Dec 2024

- Engineered a scalable recommendation engine in Azure Databricks using PySpark, leveraging TF-IDF and ALS collaborative filtering to personalize game suggestions for 10,000+ users.
- Designed robust ETL pipelines for data cleaning, feature engineering, and model training, optimizing Spark workflows to process millions of records with Delta-compatible schema design and achieving RMSE of 0.85 and Precision@5 of 72%.

**Customer Churn Prediction**, *Individual Project*

Jan 2025 – May 2025

- Built and tuned classification 5 models (Logistic Regression, KNN, SVM, Random Forest, XGBoost) using Python and scikit-learn to predict telecom customer churn, incorporating SMOTE to address class imbalance.
- Applied SHAP for model interpretability and conducted correlation analysis to identify key behavioral predictors such as complaints, status, and usage frequency.

**Northwestern Mutual Data Science Institute**, *Member*, Milwaukee, WI

Aug 2024 - Present

- Facilitated 3+ data science workshops for 50+ attendees, fostering collaboration between industry experts and academia while expanding access to community-driven data resources and career development for data scientists.