

ANBU SURIYA KUMAR

Chicago, IL | (312)-843-7255 | anbusuriyak@gmail.com | <https://www.linkedin.com/in/anbu-suriya-kumar/>

PROFESSIONAL SUMMARY:

Data professional with a master's in data science and 3+ years of experience in analytics, ETL, machine learning, and generative AI. Skilled in SQL, Python, Snowflake, Azure, Power BI, and Tableau, with expertise in natural language processing (NLP), large language models (LLMs). Passionate about transforming data into actionable insights and building AI-driven solutions.

WORK EXPERIENCE:

Genpact

Data Analyst

July 2022 - Dec 2022

- Successfully integrated **Adobe Analytics** to monitor user activity, clicks, and search patterns on the intranet portal of a leading pharmaceutical company, enabling better insights into user engagement.
- Utilized **Business Intelligence** tools and **Azure Data Factory** to analyze complex datasets, deliver actionable insights, and provide strategic recommendations, leading to a 15% improvement in operational efficiency.
- Designed and implemented incremental **data ingestion** workflows to transfer data from **on-premises SQL** databases to **Snowflake**, ensuring accurate, up-to-date reporting while optimizing storage costs.
- Automated routine tasks, such as **incremental updates**, using **Power Automate** to improve data accuracy and simplify workflows.
- Participated in client meetings and executive presentations, translating technical solutions into business insights, managing escalations, and aligning solutions with business goals.

Avasoft

Data Analyst and M365 Specialist

Feb 2020 - July 2022

- Developed **ETL** pipelines using custom **C# SSIS** job to transfer data from on-prem **SQL Server** to SharePoint Lists, ensuring compliance and reducing data retrieval time by 30%. Collaborated with teams to define migration strategies and present solutions to stakeholders.
 - Created **Power BI dashboards** embedded in **SharePoint** to provide real-time insights, collaborating with data and cloud teams to ensure scalability and security.
 - Migrated **unstructured** Lotus Notes databases to structured **Azure SQL** and **Dataverse** tables, standardizing data for use in PowerApps business applications.
 - Collaborated with data and cloud teams to build a **data warehouse** on **Azure**, integrating multiple datasets for enhanced reporting capabilities in **Azure Synapse**.
 - Automated multi-level approvals, data parsing, and **JSON-to-Excel** transformations by developing **RPA workflows** in Power Automate, reducing **manual effort and processing time by 40%**.
 - Migrated legacy applications from **.NET** and **SQL Server** to Microsoft 365 (PowerApps, Power Automate, SharePoint), ensuring smooth **data migration and validation** while addressing stakeholder concerns.
 - Prepared a **project charter** for a digital transformation initiative, detailing objectives, **ETL strategy**, and reporting framework, while coordinating cross-team collaboration to meet deadlines.
 - Took ownership and developed a demo PowerApps business use-case applications featuring **Power BI** and **Dataverse**, which was later presented by **Microsoft** at a **PowerPlatform** Conference with 10,000+ attendees.
-

PROJECTS:

AI-Powered PDF Query System

- Developed a system powered by large language models (**LLMs**) to enable natural language search within PDF documents.
- Fine-tuned **Llama2** with **LoRA** and integrated **LangChain** and **ChromaDB** for efficient and accurate information retrieval.

Supermarket Sales Analysis & Prediction:

- Conducted exploratory data analysis (**EDA**), **feature engineering**, and **data cleaning** on historical sales data to optimize marketing strategies.
- Used **Random Forest**, **Decision Trees**, and **K-Means Clustering** for customer segmentation and demand forecasting.

Car Accident Severity Prediction

- Created an **interactive dashboard** in **Tableau** to analyze high-risk areas and times for car accidents.
 - Implemented **XGBoost**, **Random Forest**, and **Decision Trees** for severity classification, achieving 85% accuracy.
 - Applied **A/B testing**, **feature engineering**, and **data balancing techniques** to improve model accuracy and reliability.
-

SKILLS:

- **Programming and Tools:** Python (NumPy, pandas, scikit-learn, keras, PyTorch, matplotlib, seaborn, OpenCV, scikit-image, PIL), React, SPFx, REST API, Graph API
 - **Machine Learning and AI:** GenAI, Deep Learning, Predictive Analytics, XGBoost, Random Forest, Gradient Boosting, Decision Trees, K-Means Clustering, NLP (LangChain, Llama2, ChromaDB, GPT, Transformer)
 - **Database and Visualizations:** MySQL, MS SQL Server, T-SQL, Oracle, SSMS, SSIS, PowerBI, DAX, Tableau, Excel
 - **Cloud and Platforms:** Azure Data Factory, Azure Blob Storage, AD, Snowflake, SharePoint, Dataverse, PowerAutomate, Visio
 - **Development and Productivity Tools:** Jira, Jupyter Notebook, VS Code, Anaconda, PowerShell
 - **Project Management and Soft Skills:** Data Storytelling, Data Modelling, Cross-Team Collaboration, Agile Methodologies, Stakeholder Engagement
-

EDUCATION:

DePaul University

Master of Science in Data Science

Pondicherry University

Bachelor of Technology in Computer Science

Chicago, IL

June 2024

Puducherry, India

April 2020