

# PRAVEEN JOGI

Chicago, IL • +1 5513590769 • [My Portfolio](#) • [Github](#)  
[praveenj3006@gmail.com](mailto:praveenj3006@gmail.com) • [LinkedIn](#)

## EDUCATION

University of Illinois Springfield

(Jan 2023 - May 2024)

Masters in data Analytics Engineering

Springfield, IL

Global Academy of Technology

(Aug 2017 - Jun 2020)

Bachelors in computer science & Engineering

Bangalore, KA

## TECHNICAL SKILLS

**Languages:** Python, R, C, C++, Scala, MATLAB

**Data Engineering:** SQL, MySQL, Airflow, ETL, Azure DevOps, AWSS3, Redshift, Power BI, Tableau, Spark.

**Machine Learning:** Keras, Pandas, NumPy, TensorFlow, Decision Tree, Logistic Regression, CNN.

**Big Data Tools:** Apache Spark, Snowflake, Databricks, AWS EMR, and DBT.

**Methodologies:** SDLC, Agile, Waterfall.

**Others:** Deep Learning, Data Analytics, Data Visualization, Data modeling, KPI metrics management, Data warehousing, Excel (VBA macros, Pivot tables, VLookup, Analytic Solver), Healthcare Analytics.

## CERTIFICATIONS

- 'AWS Data Analytics Specialty' from AWS.
- 'Snow pro Core Certified' from Snowflake.
- 'Career Essentials in Data Analysis' from Microsoft.

## WORK EXPERIENCE

**Data Analyst (Real-time Project)**

University of Illinois Springfield, Springfield, IL.

(Feb 2023 - Apr 2024)

- Conceived and executed a sentiment analysis solution that processed and analyzed 500,000 customer reviews, providing actionable insights that led to a 25% increase in user satisfaction metrics within a quarter.
- Accelerated model training time by 30% through advanced techniques and parallel processing using GPU resources.
- Incorporated sentiment analysis APIs into three client-facing applications, enhancing user engagement and satisfaction.
- Implemented the sentiment analysis model in a cloud environment, ensuring scalability and real-time processing for high-volume data streams.
- Built a data pipeline to automate data collection, preprocessing, and model updates, reducing manual intervention by 50%.

**Data Engineer**

Aarmec Technologies Pvt Ltd, Bangalore, India.

(Aug 2020 - Dec 2022)

- Collaborated with product owners to develop healthcare analytics data assets, improving data pipeline efficiency and warehouse performance by 20%.
- Formulated fact-dimension modeling (Star Schema) and created materialized views and T-SQL transformations, optimizing Datawarehouse performance.
- Crafted over 12 dashboards on Google Sheets for data visualization, enhancing business KPI tracking and stakeholder collaboration.
- Automated production processes using test plans, Agile methodologies, and KPI metrics evaluation, resulting in a 30% increase in data mining efficiency.
- Designed solutions to ingest over 2 million healthcare records (including claims and genetic tests) through ETL pipelines in SSIS and SQL transformations, improving data processing speed by 25%.
- Established a standardized data environment using CI/CD practices, optimizing code performance by 40% for data analytics and visualization in Snowflake, DBT, and Azure cloud environments.

## PROJECTS

**Sentiment Analysis on Twitter using Deep Learning**

(Feb 2023 - Jan 2024)

- Engineered a deep learning model that achieved 95% accuracy in classifying Twitter tweets into positive and negative sentiments.
- Enhanced natural language processing (NLP) techniques, increasing the precision of sentiment analysis in social media data by 20%.
- Processed over 1 million tweets, refining the model to interpret informal language and slang accurately.
- Integrated the model into a real-time analytics pipeline, enabling the analysis of 100,000+ live tweets per day for immediate sentiment insights.
- Generated actionable sentiment data that improved market research accuracy, brand management strategies, and public opinion monitoring by 30%.
- Streamlined model performance and reduced latency by 40%, enabling the sentiment analysis tool to scale with a 50% increase in data volumes.

**Customer Segmentation and Analytics for E-Commerce**

(May 2024 - Sep 2024)

- Designed and implemented a scalable data pipeline for an e-commerce platform, using AWS S3 for storage and Apache Kafka for real-time data ingestion.
- Developed ETL processes with Apache Spark to clean and transform data, including handling duplicates, missing values, and deriving features like total spend and purchase frequency.
- Built customer segmentation models based on purchasing behavior, categorizing customers into groups such as high spenders and frequent buyers for targeted marketing.
- Optimized data storage in AWS Redshift, ensuring efficient querying with a structured schema and aggregation techniques, and executed complex SQL queries to extract insights.
- Created visualizations and reports using Power BI or Tableau to present actionable insights and trends to stakeholders, supporting data-driven decision-making.

Explore all my Projects on [GitHub](#)