

# SAKSHI RATHI

Bloomington, IN | +1 (812)-778-4798 | [sakshi.k.rathi@gmail.com](mailto:sakshi.k.rathi@gmail.com) | [linkedin.com/in/sakshi-k-rathi](https://www.linkedin.com/in/sakshi-k-rathi) | [github.com/sakshiiiir](https://github.com/sakshiiiir)

## Education

### Indiana University, Bloomington

Aug 2022 - May 2024

Master of Science in Data Science

GPA 3.73 / 4.00

**Coursework:** Statistics, Exploratory Data Analysis, Time Series Analysis, Engineering Cloud Computing, Data Mining, Deep Learning Systems, Computer Vision, Social Media Mining, Elements of Artificial Intelligence.

### Vishwakarma Institute of Technology, Pune

Aug 2017 - May 2021

Bachelor in Technology in Electronics and Telecommunication

GPA 9.04 / 10.00

**Coursework:** Data Structures, Mathematics, Multivariate Data Analysis, Signal and Image Processing, Natural Language Processing.

## Experience

### Data Analyst

May 2024 - Current

Hoosier Community Network

Remote, USA

- Orchestrated **5+ data pipelines** with Azure Data Factory to process and store CSV and JSON data from 70+ U.S. Census counties in Azure Blob Storage, improving data discoverability.
- Validated API testing in Postman to validate data integrity and correctness, which resulted in **30% reduction in data errors** and enhanced data reliability.
- Developed analytical dashboards in **PowerBi** with **DAX and Power Query** to visualize time-series and geospatial education data, to make data-driven decisions.

### Data Scientist

May 2023 - Aug 2023

Magna International Inc.

Troy, MI

- Devised 5+ ETL pipelines in python to retrieve sensor data via APIs, applied data imputation and cleaning, and stored in **InfluxDB**, managing **80+ GB of time-series** big data from 10+ sensors, optimizing data consistency.
- Integrated the database with **Grafana** using APIs to create interactive dashboards for time series forecasting with ARIMA for **sensor anomaly detection**, resulting in 20% improved functions through real-time monitoring.
- Classified chiller tank flow using **statistical models** - logistic regression, decision trees, random forest, with power consumption data from 100+ injection molding machines, **resulting in 10% reduction** in malfunctions.
- Analyzed chiller efficiency using chiller metrics, identified seasonality trends, and visualized the **efficiency curve** in Grafana, resulting in a 10% boost in performance.

### Data Scientist

Aug 2021 - May 2022

Continental AG

Bengaluru, India

- Collaborated on improving vehicle speed control in Advanced Driver Assistance Systems - ADAS by developing a **PointNet++** drop-in drop-out classifier and integrating **sensor radar data**, amplified responsiveness by 20%.
- Extracted point cloud from **MongoDB** using **10+ APIs** for data integration, which reduced data retrieval errors by 15%.
- Processed point cloud data by obtaining 3D positions and relative speeds, linking with **semantic segmentation** classes, and preparing input features, which led to a 25% increase in data integration efficiency.
- Deployed and maintained the model in ADAS, monitored speed changes in **Tableau** dashboards, elevated model stability by 10%.

### Machine Learning Engineer

Jan 2021 - May 2021

Anubhooti Solutions

Pune, India

- Engineered an ASR system by storing 100+ audio recordings in AWS S3, developing **Python ETL pipelines** for data access and processing, and performing data segmentation, normalization, and **MFCC feature extraction** with librosa.
- Optimized the model using **Kaldi** libraries, maximized accuracy from 85% with pre-trained model to 90% after training with **triphone model** and fine-tuning hyper-parameters.

## Skills

**Programming Languages:** C++, Python (NumPy, Pandas, Scikit-Learn, NLTK, Keras, TensorFlow, PyTorch), R, MATLAB, SQL, Git

**Data Visualization:** Matplotlib, Seaborn, Plotly, Grafana, Microsoft PowerBi, Tableau, Microsoft Excel, Looker

**Database:** MySQL, InfluxDB, MongoDB, MS SQL Server, PostgreSQL

**Big Data and Cloud platforms:** PySpark, Hadoop, Azure (Certified-Data Fundamentals), Databricks, AWS - (S3, Sagemaker, IAM, Lambda), Apache Airflow, Kafka, Spark

**Data Analysis and Statistics:** Hypothesis Testing, A/B Testing, Statistical Modeling, Regression Analysis, Predictive Analytics

## Projects

### Skin Cancer Using AWS | Python, Machine Learning, Deep Learning, AWS

Aug 2024 - Dec 2024

- Constructed an end-to-end skin cancer image classification pipeline using **AWS(S3, SageMaker, Lambda, API Gateway)** and deployed a React.js web application, achieving 85% classification accuracy with **DenseNet** across 7 classes.

### GAN Face Detection | Python, Machine Learning, Deep Learning

Jan 2023 - May 2023

- Derived facial features from images using **Haar Cascade** and trained an ensemble of statistical models, achieving 80% accuracy; compared to **XceptionNet**, which achieved 95% accuracy in distinguishing GAN images.

### Energy Analytics Prediction | Python, Machine Learning, Deep Learning

Jan 2023 - May 2023

- Implemented a **Flask**-based web application for time series analysis of natural gas data, forecasting availability and prices using **LSTM** and **ARIMA models**, created interactive visualizations for strategic planning.