# Krishna Karthik Polisetty
**Data Scientist/Analyst**

**Tel:** +1 (716)-305-0078
**Email:** polisettykrishnakarthik@gmail.com

## Summary

**Data Scientist / Data Analyst with 3+** years of experience leveraging statistical analysis, machine learning, and data visualization to drive data-informed decision-making. Proficient in Python, SQL, and cloud platforms such as Azure and AWS for building predictive models, automating data pipelines, and delivering actionable insights. Skilled in developing interactive dashboards using Power BI and Tableau, enabling real-time business intelligence for cross-functional teams. Proven track record of improving operational efficiency, customer retention, and reporting accuracy through data-driven strategies in fast-paced, Agile environments.

## Skills

**Programming & Scripting:** Python, SQL, R
**Data Analysis & Processing:** Pandas, NumPy, PySpark, Excel (Advanced), Power Query
**Machine Learning & Statistics:** Scikit-learn, TensorFlow, Keras, A/B Testing, Hypothesis Testing, Predictive Modeling, Statistical Modeling
**Data Visualization & BI Tools:** Power BI, Tableau, Looker, Matplotlib, Seaborn, Plotly, Google Analytics
**Big Data & Cloud Platforms:** Databricks, Hive, Hadoop, Azure (Synapse Analytics, Data Factory, ML), AWS (S3, Athena, Redshift)
**ETL & Data Integration:** Azure Data Factory, Python-based ETL scripts
**Tools & Environments:** Jupyter Notebook, VS Code, Anaconda, Git, GitHub
**Project Management & Collaboration:** Jira, Agile/Scrum

## Professional Experience

### Machinery Monitoring Systems LLC, Knoxville TN - Data Science Consultant — Aug 2024 – Present

- Spearheaded the end-to-end development of a predictive maintenance solution with 96% model accuracy, overseeing data science workflows, project planning, and stakeholder engagement.
- Engineered automated data pipelines for extraction, preprocessing, and feature engineering using Python and PySpark, optimizing data readiness for model training.
- Deployed the application on AWS EC2 using FastAPI, enabling real-time fault detection and reducing system downtime across production environments.
- Collaborated in an Agile/Scrum environment to deliver sprint-based milestones, integrating feedback from cross-functional teams and driving continuous improvements in model performance.

### Tata Consultancy Services - Data Scientist — May 2022 – Jul 2023

- Developed and deployed real-time fraud detection models (Gradient Boosted Trees, Random Forests), reducing false positives by 15% and enhancing claim validation accuracy.
- Applied unsupervised learning techniques (K-Means, DBSCAN) to segment over 500K policyholders, enabling personalized retention strategies that increased renewals by 12% and reduced churn by 18% within 8 months.
- Automated end-to-end ML pipelines using Docker and Kubernetes, ensuring 99.8% model uptime, and implemented real-time performance monitoring dashboards for proactive maintenance.
- Built and optimized scalable ETL pipelines using Azure Data Factory and Databricks, processed 10M+ records, and enhanced reporting efficiency by 30% via Synapse SQL and real-time Power BI analytics.

### Mukand Sumi Special Steel Ltd - Data Analyst – Operations — Jun 2020 – Apr 2022

- Developed predictive models to optimize Precision Sizing Mill (PSM) process parameters, leading to a 12% improvement in operational efficiency and increased production throughput.
- Utilized advanced data analytics and statistical methods to reduce cold saw blade consumption and improve productivity from 50 to 56 tons/hour, enhancing overall equipment effectiveness (OEE).
- Automated daily data ingestion workflows using Power Automate and integrated email data into Power BI dashboards via Python scripts, cutting manual reporting efforts by 50%.
- Designed and deployed interactive Power BI dashboards using Power Query and DAX, delivering real-time production insights to leadership and supporting strategic supply chain and manufacturing decisions.
- Partnered with cross-functional teams to implement Row-Level Security (RLS) and strengthen data governance and compliance, ensuring secure and scalable business intelligence access.

## Projects

### AI-Driven Temperature Forecasting for Educational Environments in Tanzania with OpenDevEd

Addressed the challenge of extreme classroom temperatures in Tanzanian schools by developing a dataset of 16 million records, integrating geospatial and meteorological data to reflect diverse environmental factors.

Achieved a Mean Absolute Error (MAE) of 0.19C in forecasting indoor temperatures using ML models, including LSTM and XGBoost, enabling data-driven strategies to improve learning environments and resource allocation.

## Education

- **Master of Science, Data Sciences and Applications – CGPA: 3.96/4.0**
  University at Buffalo, The State University of New York, Buffalo, NY
- **Bachelor of Technology May 2020**
  Visvesvaraya National Institute of Technology, Nagpur, India

## Certifications

- **Microsoft Certified: Power BI Data Analyst Associate (PL-300)**
- **IBM Data Science Professional Certificate**