

KARTIK JONNAVITHULA

DATA SCIENTIST

DETAILS

PHONE

+91 7207821892

EMAIL

kartik42sharma@gmail.com

LINKS

[LinkedIn](#)

[Github](#)

SKILLS

Machine Learning

Deep Learning

Reinforcement Learning

Snowflake

Power BI

Data Storytelling

Azure

GCP

Generative AI

Prompt Engineering

Langchain

LANGUAGES

Python

SQL

Javascript

PROFILE

As a dedicated Data Scientist, I focus on enhancing business operations & strategies through applying ML and data engineering. My diverse experience includes developing predictive models and recommendation systems, leading to significant cost savings and improved efficiency. A constant learner, I strive to be a valuable addition to any team and contribute to transformative projects

EXPERIENCE

Data Scientist, Eugenie.ai

2023 — 2024

Key Projects:

Optimizing Tata Metaliks' Sinter Plant with a Digital Twin Solution

- Targeted optimization goals: enhance production, reduce coke usage, minimize particulate matter emissions, and lower return fines.
- Developed an LSTM-based model for accurate KPI prediction, addressing data drift through periodic partial retraining.
- Applied DDPG (RL) Algorithm to optimize plant operations in real-time.
- Achieved impactful results: increased throughput by 11%, reduced return fines by 5%, decreased coke usage by 4%, and cut particulate matter emissions by 5%.

Enhancing Chip Manufacturing with Advanced Anomaly Detection for JABIL

- Addressed the challenge of identifying defective chips before they enter the Reflow oven, aiming to reduce losses and save resources.
- Analyzed inspection data, addressing class imbalance (through synthetic data generation and strategic undersampling) and gaining insights into defect patterns.
- Developed specialized anomaly detection models for key component types using ML and DL algorithms, achieving an F1 score of 75%.
- Identified and Implemented a secondary business case to verify machine-identified defects after reflow oven, improving the F1 score to 92%.
- Achieved an estimated savings of \$200k per production line through effective defect detection and reduced human verification effort.

Data Scientist | Assistant Manager, Tata iQ

2019 — 2023

Worked closely with various Tata companies (Tata Steel, Croma, Tata Chemicals, Tata Consumer Products) in providing solutions in the form of predictive modelling, data analysis, visualization dashboards and data flow automation tasks

Key Projects :

Revolutionizing Tata Steel's Sinter Plant with Digital Twin Technology

- Targeted goals: increase production, reduce coke usage, minimize particulate matter emissions, and lower return fines.
- Developed prediction models to forecast KPIs, addressing data drift using SAMKNN algorithms.
- Built a recommendation system with a simulator to suggest optimal operator actions by simulating all possible actions.
- Achieved impactful results: increased throughput by 5%, reduced return fines by 3%, decreased coke usage by 2%, and cut particulate matter emissions by 4%.
- Recognized at the World Economic Forum as a digital innovation for Industry 4.0 in iron and steel manufacturing.

- Scaled across four sinter plants, resulting in annual savings of ₹8 crore.

Leveraging Data Science to Optimize Coke Blends for Tata Steel

- Targeted goals: optimize coke blend to meet constraints on physical properties, CSR, CSN, AMS, and M40 while reducing costs.
- Created a dataset of historical coal blends and their properties, along with a live dataset of available coals in the stockyard.
- Developed an ensemble of stacked models (XGBoost, SVMs, Deep Learning) to predict coke properties with various virtual blends. Implemented mixed integer programming to select the optimal virtual blend balancing cost and conformity to constraints.
- Achieved significant cost savings: 10-20 lakhs per batch run per plant. Scaled the solution across six coke plants, leading to annual savings of ₹40 crore.

Enhancing Croma's Marketing Strategy with Real-Time Customer Segmentation

- Targeted goals: optimize marketing strategy, enable targeted ads, identify loyal customers and churn rate, and perform live RFM Analysis.
- Developed a robust data pipeline to extract data from AWS data lakes and utilize PySpark for creating and updating customer tables with RFM metrics and purchase locations.
- Stored results in a big data table for continuous updates and created reporting scripts to provide updated insights for upper management.

PROJECTS

TextBookGPT: Learning with an AI Chatbot

<https://github.com/kartik-99/textbookGPT>

- Developed a chatbot using LangChain and OpenAI LLMs to assist users in understanding textbook concepts.
- Enabled input of textbook PDFs, allowing the chatbot to extract and explain key information.
- Enhanced interactive learning by providing detailed, user-friendly explanations.

PIP Package: Python Script Timer

<https://pypi.org/project/scriptTimer>

- Developed a Python package for accurately recording the time taken by various sections of a script.
- Features include timing different sections and subsections, infinite sub-divisions, customizable logging, and automatic adjustment of time units, and exporting timers to csv
- Utilizes an object-oriented approach allowing multiple timers within a script.
- Useful for developers to identify time-consuming code sections and monitor script efficiency with minimal additional code.

EDUCATION

M.Sc Computer Science: AI & ML, Woolf University

(Remote)

Jun 2023 — Jun 2025

B.Tech : Computer Science, SRM University

2015 — 2019

COURSES

Deep Learning Specialization, Coursera

Snowflake Fundamentals, Snowflake

Machine Learning : Andrew NG, Coursera