

# Kunal Sheshrao Takalkhede *Data Engineer (AI/ML)*

✉ kunal98takalkhede@gmail.com

☎ 7776838286

🌐 Kunal Takalkhede

📍 Hinjewadi ,Pune 411057

## 👤 PROFILE

- **Experienced ETL Developer** with **3.11** years of expertise in designing and building efficient data pipelines.
- Proficient in **SQL**, **Azure DataFactory**, **Databricks**, **Datastage**, and **BODS**, with a solid understanding of **Azure Services** and data warehousing principles.
- Hands-on experience in **machine learning** using algorithms like **XGBoost**, **Random Forest**, etc. to enhance data-driven decision-making.
- Currently developing industry-specific **data pipelines** for a UK-based banking client, adhering to **SDLC** and **Agile methodologies**.
- Skilled in **data governance** with expertise in SQL-based analysis to identify and resolve data defects.
- Adept at crafting **automated solutions** using **Python** and **Unix scripts** to streamline processes.
- Passionate about leveraging data to drive insights and transformative business solutions.

## 🧠 SKILLS

ETL | SQL | Azure Data Factory | Data Analysis | Azure Databricks | Azure Services | Data Warehouse | Power BI | Machine Learning | Python | Unix | Hadoop | Explanatory Analysis | Numpy | Pandas | NLP | JIRA | Agile | Regression and classification

## 📁 PROJECTS

### Fraud Detection System for Financial Transactions

- Built a fraud detection model using XGBoost on a 6-million-record banking payments dataset to classify potentially fraudulent transactions.
- Performed exploratory data analysis (EDA) to uncover hidden patterns and engineered features to enhance model accuracy.
- Addressed class imbalance using SMOTE and optimized decision thresholds to improve recall for fraud detection.
- Conducted extensive data preprocessing, including handling missing values, outliers, and feature normalization.
- Tuned hyperparameters and evaluated the model using accuracy, ROC-AUC, and recall metrics for optimal performance.
- Delivered a scalable batch processing pipeline and a real-time prediction module for continuous monitoring of fraudulent transactions.

### Health Insurance Premium Prediction

- Developed a predictive model using XGBoost to estimate health insurance premiums from a 700,000-record dataset containing demographics, medical history, lifestyle, and region data.
- Conducted extensive exploratory data analysis (EDA) to uncover hidden patterns and experimented with feature engineering to improve model performance.
- Addressed class imbalance using SMOTE and fine-tuned hyperparameters with GridSearchCV for optimized predictions.
- Focused on comprehensive data preprocessing, including cleaning, normalization, and feature selection.
- Evaluated the model using metrics like RMSE and  $R^2$ , and adjusted decision thresholds to refine results.
- Delivered the solution as a Python batch processing script for bulk premium estimation and integrated real-time prediction capabilities.

### Generative AI Chatbot for Financial and Operational Risk Guidelines

- Developed a chatbot using LangChain and OpenAI models to provide real-time answers to questions on RBI's financial and operational risk guidelines, enhancing accessibility to complex regulatory information.
- Implemented document processing and retrieval system by loading, splitting, and vectorizing RBI guideline documents using FAISS and OpenAI embeddings for efficient and accurate query resolution.
- Performed advanced text preprocessing with a character text splitter to optimize document chunking for token size limitations, ensuring accurate retrieval and generation of responses.
- Delivered a scalable and deployable chatbot solution capable of handling dynamic user queries, integrated with conversational retrieval chains for iterative and context-aware answers.

## 👛 PROFESSIONAL EXPERIENCE

### Data Engineering, Management & Governance Analyst

12/2020 – Present | Pune, India



Accenture

- Developed and optimized ETL pipelines for regulatory reporting using **Azure DataFactory**, **Databricks**, **Synapse Analytics**, **SQL**, **IBM DataStage**, and **SAP BODS**, significantly enhancing data processing efficiency.
- Improved data pipeline performance by **70%** using DataStage and SAP BODS, optimizing data workflows and increasing overall processing speed.
- Automated tasks with **Python** and **Shell scripting**, reducing processing time by **25%** and streamlining workflows in cloud-based environments.

- Enabled Data Analysis teams to address regulatory and risk management challenges by developing ETL pipelines for Credit Card and Loan data, delivering accurate and structured datasets for deeper insights and compliance reporting.
- Proficient in creating interactive and visually compelling reports using **Power BI** to facilitate data-driven decision-making.
- Experienced in data analysis and machine learning model development, with hands-on expertise in **XGBoost, Neural Networks, Random Forest**, and other algorithms.
- Identified and addressed bottlenecks in data pipelines, showcasing strong problem-solving and data analysis abilities.
- Contributed to **code reviews** to maintain industry best practices.
- Led a team of 5+ members, managing projects, ensuring high-quality deliverables, and consistently meeting deadlines. Provided mentorship and fostered a collaborative team environment.
- Collaborated closely with designers and analysts to develop and maintain high-quality data pipelines, ensuring data accuracy and resolving defects promptly.
- Leveraged advanced SQL and **data warehousing** concepts to support data engineering processes effectively, ensuring data integrity and scalability.
- Engaged with business leadership and stakeholders to gain approvals, align projects with strategic objectives, and effectively communicate data product value and business metrics.
- Conducted knowledge transfer sessions for 50+ professionals on processes and tools, enhancing team capabilities and promoting knowledge-sharing.

## CERTIFICATES

---

- Microsoft Certified: Azure Data Engineer Associate 
- Microsoft Certified: Azure Fundamentals 

## EDUCATION

---

<b>ML and Data Science Certification Course</b> <i>Intrvu.ai</i>	03/2024 – 11/2024   Pune, India
<b>Bachelor Of Engineering (Electronics Engineering)</b> <i>Yeshwantrao Chavan College of Engineering</i>	08/2016 – 05/2020   Nagpur, India

## AWARDS

---

**Performer of the Month**  
*Accenture*  
Awarded Performer of the month 5 Times: June(2021), March(2022) ,October (2022) ,October(2023) & May(2024)