

KAMESH DUBEY

Mumbai, India

+91-7217563163

kamesh.iitbombay@gmail.com

LinkedIn

Portfolio

PROFESSIONAL SUMMARY

Data Scientist skilled in credit-risk and AI product development, applying statistical and machine learning methods to strengthen credit-decision. Proficient in Python, SQL, and the ML model development lifecycle.

EDUCATION

Indian Institute of Technology, BOMBAY

M.Sc. Applied Statistics and Informatics

Aug'22 - Dec'24

Mumbai, IN

Gurukula Kangri University

B.Sc. Mathematics and Computing

Aug'18 - Jul'21

Haridwar, IN

WORK EXPERIENCE

Associate Consultant–Financial Risk Management

KPMG Assurance and Consulting Services LLP

Jun'25–Present

Mumbai, IN

- Conducted quantitative validation of **10+** loan scorecards and ECL models using PSI, Gini, Chi-square, and Binomial tests to assess conceptual soundness, ensuring compliance with regulatory model-risk frameworks
- Automated the validation workflow delivering complete quantitative validation results in client-ready format
- Prepared and presented comprehensive validation reports summarizing key findings and actionable insights

Data Science Intern

Bluepond AI

Apr–May '25

Chennai, IN

SOV Chatbot for Underwriters

- Built a RAG chatbot with tool-calling capabilities to generate visualizations and insights from the Excel files
- Designed an **AI-driven data pipeline** to extract, clean, and standardize Statement-of-Value (SoV) Excel files
- Fine-tuned a **BERT (99.5% F1)** classifier on a custom dataset to filter non-insurance queries before API calls

Brokers Co-Pilot

- Built an **extraction tool** that extracted **30+** key fields from insurance documents, cutting review time by **80%**
- Revamped the document reader pipeline to enhance text parsing precision and improve extraction accuracy

TECHNICAL SKILLS

Technical Skills: Python, R, Excel, SQL, Scikit-learn, PyTorch, Pandas, Plotly, LangChain, Git, Docker, AWS

Key Coursework: Machine Learning, Deep Learning, Regression Analysis, Statistical Inference, Optimization

ACADEMIC PROJECTS

Reviews Sentiment Analysis – Course Project | GitHub

Aug–Nov '24

Guide – P. Balamurugan, Department of Industrial Engineering and Operations Research, IIT Bombay

- Implemented **NLP** techniques like data cleaning, tokenization, and **lemmatization** to boost model accuracy
- Utilized **TF-IDF** for vectorization and **LSTM** for classification and achieved **87.17%** accuracy on test data
- Developed a **Streamlit** webapp and set up a **CI/CD pipeline** for deployment of Docker container on **AWS**

Oil Well Failure Prediction | MSc. Project

Jan–Apr '24

Guide – Sanjeev V Sabnis, Department of Mathematics, Bombay

Co-Guide – Prof. Sujit K Ghosh, Department of Statistics, North Carolina State University

- Conducted a comparative analysis of **Logistic Regression** and **Tree-Based Models** for the failure prediction
- Optimized Decision Tree by tuning hyperparameters on feature set, achieving **83%** cross-validated F1-score

Bank Customers Segmentation | Course Project | GitHub

Aug–Nov '23

Guide – Prof. Siuli Mukhopadhyay, Department of Mathematics, IIT Bombay

- Enhanced interpretability by leveraging **PCA** and **factor analysis** to uncover latent variables in the dataset
- Segmented data using **K-Means** clustering and evaluated clusters via **Silhouette Score** and **Elbow Method**

SCHOLASTIC ACHIEVEMENTS

- Secured **AIR-50** in Mathematical Statistics, IITJAM 2022 among 2,912 candidates organized by IITR '22
- Achieved a remarkable **Department Rank 3** out of 120 BSc. Students at Gurukul Kangri University '21