

KAMESH DUBEY

Mumbai, India

+91-7217563163

kamesh.iitbombay@gmail.com

LinkedIn

GitHub

Portfolio

EDUCATION

Indian Institute of Technology, BOMBAY

MSc. Applied Statistics and Informatics

Aug'22 – Dec'24

Mumbai, India

Gurukul Kangri University

BSc. Mathematics and Computing

Aug'18 – Jul'21

Haridwar, India

WORK EXPERIENCE

Associate Consultant–Financial Risk Management

Jun'25–Present

KPMG Assurance and Consulting Services LLP, Mumbai

- Conducted quantitative validation of **10+** loan scorecards and ECL models using PSI, Gini, Binomial and Chi-Square test 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 validation reports with insights and visualizations; supported presentation of findings to the clients

Data Science Intern

Apr–May '25

Bluepond AI, Chennai

SOV Chatbot for Underwriters

- Built RAG chatbot with tool-calling functionality to generate visualization along with insight from excel files
- Engineered an automated data pipeline to extract, clean, and standardize Statement-of-Value (SoV) excels
- Fine-tuned **BERT (99.54% F1)** classifier on a custom dataset to filter non-insurance queries before API calls

Brokers Co-Pilot

- Built an **extraction tool** that extract **30+** key fields from insurance documents, cutting review time by **80%**
- Revamped document reader module using NLP for space correction and integrated logic for image detection

ACADEMIC PROJECTS

End-to-End 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

- Performed thorough EDA and **feature engineering**, handled missing values, and addressed data imbalance
- 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 a **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 variable in the dataset
- Segmented data using **K-Means** clustering and evaluated clusters via **Silhouette Score** and Elbow Method

TECHNICAL SKILLS

Programming Languages and Tools: Python, SAS, SQL, MLflow, Git, Docker, AWS

Machine Learning Libraries: Pandas, scikit-learn, statsmodels, PyTorch, Langchain

COURSES UNDERTAKEN

Machine Learning

Statistical Inference

Deep Learning

Hypothesis Testing

Time Series Analysis

Regression Analysis

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