

KAMESH DUBEY

Mumbai, India

+91-7217563163

kameshcodes@outlook.com

kameshcodes

kameshcodes

EDUCATION

Indian Institute of Technology, BOMBAY

MSc. Applied Statistics and Informatics - CPI - 6.7

Aug'22 – Dec'24

Mumbai, India

Gurukul Kangri University

Bsc. Mathematics and Computing - CGPA - 8.6

Aug'18 – Jul'21

Haridwar, India

KEY PROJECTS

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 **79%** cross-validated F1-score

Semi-Supervised Random Forest — Course Project

Jan–Apr '24

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

- Implemented **SSRF** algorithm from existing research and **validated** paper results on *Caltech* – 101 dataset
- Designed the **pipeline** to preprocess and train model, streamlining learning with labeled and unlabeled data
- Improved model accuracy by **9%** by integrating a CNN-based feature extractor and inverse cooling function

Bank Customers Segmentation — Course Project

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**

Residual Analysis and Diagnostic Checking — Course Project

Jan–Apr '23

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

- Identified the influential observations and outliers in dataset by measuring **Cook's Distance** and **DFFITS**
- Diagnosed *multicollinearity* and *heteroskedasticity*, and remedied them through **VIF** and transformations
- Achieved **80.1% adjusted-R²** by selecting predictors guided by **AIC** and **Mallow's C_p** for model selection

TECHNICAL SKILLS

Programming Languages and Tools: Python, R, SQL, MLflow, DVC, Git, Docker, FastAPI, AWS

Machine Learning Libraries: Pandas, scikit-learn, statsmodels, PyTorch, nltk, OpenCV, PySpark

SCHOLASTIC ACHIEVEMENTS

- Awarded the **MCM Scholarship** during both years of MSc. at Indian Institute of Technology Bombay '24
- 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
- Received LIC HFL Vidhyadhan Scholarship for Graduation, awarded to **200** UG students across India '20

POSITION OF RESPONSIBILITIES

PG Convener, Events and PR — E-Cell, IIT Bombay

Aug'22–Feb'23

Asia's leading entrepreneurship-promoting student body with patronages from UNESCO.

- Executed Seed-Star Event, connecting 30+ startups and 20+ VCs/Angels for seed funding in E-Summit'23
- Onboarded 10+ outreach partners for promoting E-Summit'23 events through their different social media

COURSES UNDERTAKEN

| Key Courses | Online Courses |
|--|--|
| Deep Learning – Theory and Practice | Generative AI with LLMs – deeplearning.ai |
| Machine Learning – Principles and Techniques | Deep Learning Specialization – deeplearning.ai |
| Time Series Analysis | Python 3 Programming Specialization – Coursera |
| Statistical Inference – I | MLOps Fundamentals – DataCamp |