DIPANSHU MISHRA

Data Science Intern | ML Engineer | Data Analyst

+91-8454081928 mdipanshu713@gmail.com https://github.com/Dipanshu712 Dahisar East, Mumbai

DM

EXPERIENCE

06/2023 - 12/2023

Mumbai

Web Developer Intern

Save As Web

A web development company focusing on creating innovative online solutions

- Designed and developed 3 responsive web applications using WordPress and modern web technologies, implementing data-driven solutions that increased user engagement and page views by 40%
- Applied statistical analysis to user behavior data and implemented secure authentication systems using Laravel framework, achieving 20% improvement in system security metrics
- Developed business intelligence dashboard with interactive data visualization features, improving inventory management efficiency by 15% through automated reporting and predictive analytics

EDUCATION

01/2024 - 01/2026

Mumbai, Maharashtra

Master of Science in Information Technology

Thakur College of Science and Commerce

01/2020 - 01/2023

Mumbai, Maharashtra

Bachelor of Science in Information Technology

Mumbai University

PROJECTS

Medical Image Classification - Breast Cancer Detection

11/2024 - 12/2024

Research project focused on statistical analysis and predictive modeling for breast cancer detection

- Analyzed complex datasets and benchmarked 12 ML algorithms (Logistic Regression, SVM, XGBoost, Neural Networks) achieving 97.36% diagnostic accuracy
- Implemented advanced statistical analysis with F1 score (0.966) using Gradient Boosting; XGBoost achieved 100% precision through hyperparameter optimization
- Conducted comprehensive model validation maintaining F1 scores above 93% across all algorithms through systematic cross-validation and performance tuning
- Deployed ensemble models (Random Forest, XGBoost, LGBM) with robust feature engineering for clinical decision-making applications

Healthcare Analytics - Tuberculosis Detection via X-Ray Analysis

08/2025 - 03/2024

Personal project focused on tuberculosis detection via X-ray analysis using advanced CNN and AutoML

- Developed end-to-end machine learning pipeline with automated dataset type detection and CNN feature extraction using pre-trained ResNet50, achieving 99.29% classification accuracy
- Performed advanced feature engineering using Principal Component Analysis (PCA) to optimize computational efficiency, reducing CNN features from 2,048 to 100 dimensions
- Built comprehensive AutoML system incorporating exploratory data analysis (EDA), automated model selection, statistical validation, and SHAP explainability for transparent medical diagnostics
- Implemented data mining techniques to balance severely imbalanced dataset (700:140 ratio) using systematic sampling and data augmentation methods

SKILLS

CSS | Data Structures | Deep Learning | Django | EDA | Java | JavaScript | Jupyter | Laravel | MatplotLib | Microsoft Excel | ML Algorithms | NLP | Numpy | OpenCV | Pandas | Python | Random Forest | Scikit-Learn | Seaborn | SQL | SVM | Tableau | TensorFlow