Ojas Singh

ojassingh2805@gmail.com LinkedIn GitHub

Education

Vellore Institute of Technology **Sept 2022 – Apr 2027**
Integrated M.Tech — Computational and Data Science *CGPA: 8.45*

Experience

Database Analyst Intern **Oct 2025 – Dec 2025**
Crimson Systems *Jabalpur*

- Designed and optimized normalized relational databases using SQL and SQLAlchemy to manage large-scale academic data, ensuring high data integrity, consistency, and efficient querying across multiple modules.
- Developed advanced SQL queries and automated reporting workflows (joins, aggregations, subqueries) to deliver actionable performance insights while reducing manual reporting effort and turnaround time.
- Implemented secure role-based access control and data validation frameworks, collaborating with stakeholders to translate business requirements into scalable data models and analytical schemas.

Projects

Diabetic Retinopathy Detection System | Python, TensorFlow, OpenCV, Scikit-learn, XGBoost **July 2025**

- Preprocessed and analyzed 2,750+ retinal images using normalization, augmentation, and noise reduction, conducting EDA to understand class imbalance and feature variability for informed modeling decisions.
- Built and trained an ensemble model combining CNN, Random Forest, and XGBoost, achieving 75.27% classification accuracy through optimized training strategies.
- Improved model robustness and generalization by addressing class imbalance with weighted loss functions and evaluating performance using precision, recall, F1-score, confusion matrix, and ROC-AUC.

Zomato EDA and Regression Analysis | Python, Pandas, Matplotlib, Seaborn, Scikit-learn **May 2025**

- Conducted comprehensive EDA on restaurant datasets to identify trends affecting customer ratings, pricing, and service quality, using statistical analysis and data visualization techniques.
- Cleaned, transformed, and engineered 20+ features, handling missing values, outliers, and encoding issues, and built regression models (Linear Regression, XGBoost) achieving a strong R^2 of 86.4%.
- Reduced multicollinearity and improved model interpretability through correlation analysis and feature selection, effectively communicating insights using visualizations such as heatmaps and box plots.

Algerian Forest Fire Prediction System | Python, Pandas, Scikit-learn, Flask **Apr 2025**

- Analyzed and engineered features from the Algerian Fire Weather Index dataset (FFMC, DMC, DC, ISI, BUI, FWI) to identify fire-risk patterns and enhance predictive signal.
- Developed and evaluated ensemble machine learning models using stratified K-fold cross-validation, achieving 93% accuracy with strong recall for fire-prone conditions.
- Deployed a Flask-based analytical API providing real-time predictions, feature-importance insights, and reliable early-warning metrics validated using accuracy, MAE, recall, and confusion matrices.

Technical Skills

Languages: Python, SQL, Java
Data Analysis: Pandas, NumPy, Exploratory Data Analysis (EDA), Statistical Analysis
Visualization: Matplotlib, Seaborn, Tableau
Databases: SQLite, Relational Databases (RDBMS), SQLAlchemy (ORM)
Machine Learning: Scikit-learn, XGBoost, TensorFlow
Tools: Google Colab, VS Code, Git, GitHub
Soft Skills: Analytical Thinking, Stakeholder Communication, Problem Solving, Team Collaboration

Certifications

- Applied Machine Learning in Python — University of Michigan (Coursera)
- Introduction to Machine Learning — NPTEL