Rohit Agrawal

Data Science | Machine Learning | Analytics

agrawalrohit937@gmail.com | +91 8279414117 | India | https://github.com/agrawalrohit937/
https://www.linkedin.com/in/rohit-aggrawal | https://agrawalrohit937.github.io/PortFolio/

Education

- Currently **Bachelor of Technology (Computer Science Engineering)**
  Invertis University, Bareilly | 2023 – 2027 | CGPA: **9.0**
- **Class XII (UP Board)** | 2023 — **92%**
- **Class X (UP Board)** | 2021 — **90.8%**

Skills

- **Programming & Databases:** Python, SQL, PostgreSQL
- **Data Science & Machine Learning:** Pandas, NumPy, Data Cleaning, EDA, Feature Engineering, Classification, Model Evaluation
- **ML Tools & Libraries:** scikit-learn, XGBoost, TensorFlow (Basics), GridSearchCV, RandomizedSearchCV, Optuna, Ensemble Learning
- **Data Visualization & BI:** Power BI, Matplotlib, Seaborn, Git, Jupyter Notebook

Experience / Internship

Data Science Intern – Altair (Virtual) *Oct 2025–Dec 2025*

- Worked on real-world datasets involving data cleaning, EDA, and feature engineering using Python.
- Built and evaluated machine learning classification models using scikit-learn.
- Visualized insights using Matplotlib and Seaborn following industry-standard workflows.

AI/ML Intern – Google (Virtual) *Apr 2025- Jun 2025*

- Implemented end-to-end ML workflows including preprocessing, model training, and evaluation.
- Gained hands-on experience solving real-world AI problems using structured datasets.

ITI – ICTSM (Information & Communication Technology System Maintenance) *Aug 2023–Jul 2025*

- Performed system maintenance, OS installation, and basic networking troubleshooting.
- Supported hardware diagnostics and system optimization tasks.

Projects

Heart Attack Risk Prediction (ML | Deep Learning | Federated Learning)

- Built an **end-to-end medical ML pipeline** on the **UCI Heart Disease dataset**, covering data cleaning, feature engineering, modeling, and evaluation.
- Performed **EDA, correlation analysis, and outlier handling (IQR & statistical bounds)** to identify key clinical risk factors.
- Trained and compared **Logistic Regression, SVM, KNN, Random Forest, XGBoost, and DNN**, evaluated using **ROC-AUC and Stratified CV**.
- Applied **hyperparameter tuning** using **GridSearchCV, RandomizedSearchCV, and Optuna** to improve model performance.
- Built **ensemble models (Voting & Stacking)**, with **stacking achieving 86.5% accuracy and 0.91 ROC-AUC**, outperforming individual models.
- Implemented **Explainable AI (SHAP & LIME)** for global and patient-level interpretability; saved final models and visualized results via **Power BI dashboards**.

Tech: Python, scikit-learn, XGBoost, TensorFlow, SHAP, LIME, Optuna, Power BI

Links: GitHub | Medium Report | Power BI Dashboard | Live Demo

Text Emotion Detection (Sentiment Analysis)

- Built an **NLP-based emotion classification system** (joy, anger, sadness, fear) using **TF-IDF features**.
- Trained, tuned, and evaluated multiple models (Logistic Regression, SVM, Random Forest, XGBoost) with ensemble methods.
- Deployed the best model for **real-time emotion prediction**.

Tech: Python, NLP, scikit-learn, XGBoost, Flask

Links: GitHub | Medium Report | Live Demo

Participation & Achievements

- **IEEE Student Member**
- Participated in **HackBhoomi 2025** (Internal Smart India Hackathon), **Invertis University**

Certifications

- **Google AI-ML Virtual Internship —** EduSkills | Supported by Google
- **Altair Data Science Master Virtual Internship —** EduSkills | Supported by Altair & AICTE
- **Human Research – Data or Specimens Only Research (CITI Program), MIT Affiliates** | Jan 2026
- **Python for Data Science —** IBM / Coursera
- **Tata Data Visualization: Empowering Business with Effective Insights —** Tata Group (Forage)