Automated Machine Learning Model Builder & Visualisation Web Application

MKSSSS Cummins College of Engineering for Women, Pune.

Team Members:

Shraddha Mane (UCE2024011),
Siddhi Khomane (UCE2024012),
Sneha Bhalerao (UCE2024015),
Samruddhi Kshatriya (UCE2024016)

The Challenge: Bridging the ML Skills Gap

Machine learning holds immense potential, but its complexity often creates a barrier for many. We've identified several key obstacles:

Coding & Technical Expertise

Traditional ML development demands extensive coding knowledge and a deep understanding of various algorithms.

Complex Data Preprocessing

Handling missing values, encoding categorical data, and scaling features are time-consuming and intricate tasks.

Model Selection & Evaluation

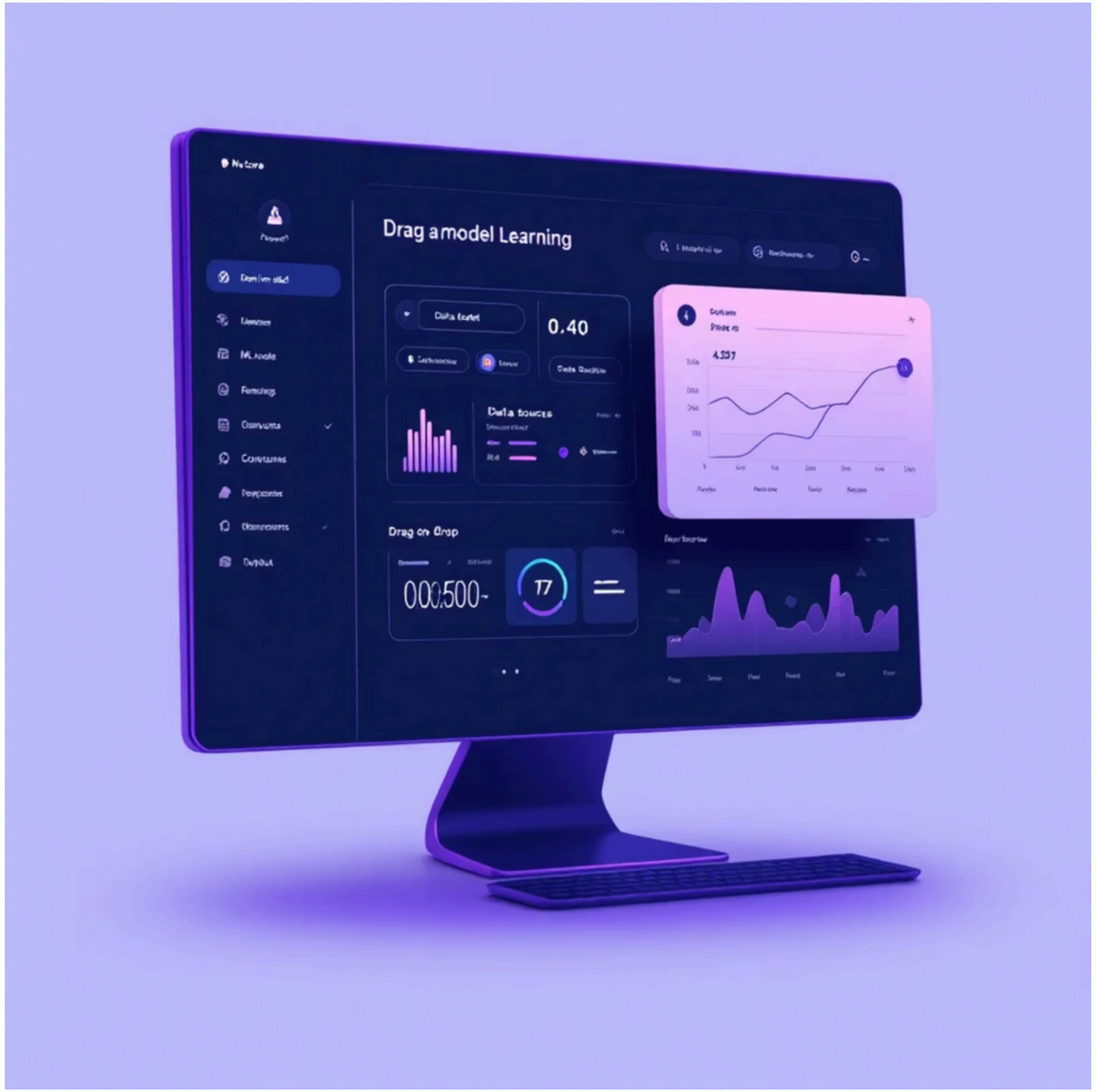
Choosing the right model and effectively evaluating its performance can be a daunting, iterative process.

Accessibility for Learners & Researchers

Many aspiring learners and researchers are unable to leverage ML due to these technical hurdles.

There's a pressing need for a simplified platform that allows users to train ML models without needing to write a single line of code.

Our Innovative Solution: An Interactive ML Platform



Web-Based & Interactive

Access powerful ML tools through a user-friendly interface accessible from anywhere.

Automated Data Preprocessing

Intelligent algorithms handle data cleaning, encoding, and scaling automatically.

Versatile Data Import

Upload CSV files or seamlessly import datasets directly from Kaggle for instant analysis.

Comprehensive ML Support

Supports Classification, Regression, Clustering, and Association Rules to cover a wide range of analytical needs.

Scope & Objectives: Empowering Non-Programmers

Our Scope:

> A comprehensive, full-stack ML model training and evaluation system, meticulously designed for non-programmers.

Key Objectives:

Intuitive UI with React.js

Develop a highly responsive and intuitive user interface using React.js for a seamless experience.

Robust Backend with FastAPI + Python

Ensure efficient and scalable backend processing through FastAPI and Python's powerful libraries.

Support for 20+ ML Algorithms

Integrate a wide array of machine learning algorithms to cater to diverse problem types.

Real-time Visualisations

Provide dynamic, real-time visualisations for immediate insight into model performance and data characteristics.

System Architecture: The Engine Behind Simplicity

Client (Frontend)

Powered by React.js for a dynamic and responsive user experience.

API Layer

Utilises FastAPI to ensure fast, efficient, and reliable communication.

ML Processing

Leverages Python with scikit-learn, XGBoost, LightGBM, and mlxtend for diverse algorithm support.

Dataset Storage

Supports local file storage and direct integration with Kaggle for data.

Core Features & Extensive Algorithm Support

Our platform offers a rich set of features to streamline the ML workflow, backed by an impressive array of algorithms for diverse applications.

Key Features:

Data Handling

Seamless dataset upload and direct import from Kaggle for quick access to data.

Automated Preprocessing

Intelligent algorithms for cleaning, encoding, and scaling data, reducing manual effort.

Model Training

Supports Classification, Regression, Clustering, and Association Rule Mining.

Comparison & Metrics

Evaluate and compare models using various performance metrics for informed decision-making.

Algorithms Supported (26+ Total):

Classification

Logistic Regression, KNN, SVM, Naive Bayes, Decision Tree, Random Forest, XGBoost, CatBoost, Gradient Boost.

Regression

Linear, Ridge, Lasso, ElasticNet, Random Forest, Gradient Boost, XGBoost, CatBoost.

Clustering

K-Means, DBSCAN, Agglomerative, Birch.

Technology Stack: Powering Performance

Our platform leverages a modern and efficient technology stack to deliver a seamless and powerful machine learning experience.

Frontend	React.js, Chart.js
Backend	FastAPI
Language	Python, JavaScript
ML Libraries	scikit-learn, XGBoost, LightGBM, mlxtend
Integration	Kaggle API

This combination ensures robustness, speed, and extensibility, making our platform a reliable tool for all your ML needs.

Experimental Results: Demonstrating Superior Performance

Our platform has been rigorously tested across various datasets and machine learning tasks, consistently demonstrating strong performance. Here are some summarised results:

Classification (Breast Cancer Dataset)

Using the Breast Cancer dataset, our platform quickly identifies the most accurate models for classification.

Logistic Regression	95.6%
Random Forest	96.5%
XGBoost (Best)	**97.4%**

Regression (California Housing Dataset)

For predicting housing prices, our models achieve competitive R^2 scores.

Random Forest	0.467
XGBoost (Best)	**0.490**

Clustering (Iris Dataset)

The platform effectively groups similar data points, highlighted by optimal silhouette scores.

Best silhouette score: **0.552 (K-Means)**

Future Scope of Project:

Integration of AutoML for Complete Automation

Automatic hyperparameter tuning, model selection, and optimization to achieve higher accuracy without user intervention.

Cloud Deployment & Scalability

Run heavy ML models on cloud (AWS, Azure, GCP) with multi-user access and real-time collaboration.

Real-Time Data Pipeline

Enable live model training using streaming sources like Kafka, IoT sensors, or APIs.

Mobile & Desktop Application

Expand platform into a mobile app and desktop software for broader accessibility and offline usage.

Conclusion

- The platform simplifies machine learning for beginners and non-programmers.
- It automates data preprocessing, model training, and evaluation.
- Supports a wide range of ML algorithms across classification, regression, clustering, and association rules.
- Provides clear visual insights and model comparisons.
- Enables fast experimentation using local datasets and Kaggle integration.
- Overall, the system makes ML more accessible, efficient, and user-friendly.

Thank youu!