

AutoDS: Simplified Data Analysis and Machine Learning Modelling

Assistant Professor A. C. Sawant, Satyam Shende, Jayesh Mahajan, Alfaj Tamboli

Department of Information Technology, SKN Sinhgad Institute of Technology & Science, Lonavala,
Maharashtra, India

Abstract- The AutoDS project report presents a web application designed to simplify and automate machine learning (ML) model training and deployment for users with limited coding expertise. The platform supports various traditional ML and deep learning models, automates data preprocessing, exploratory data analysis (EDA), model training, evaluation, and report generation, making ML accessible across diverse domains. This review articulates the project's motivation, design, methodology, applications, advantages, limitations, and future scope, highlighting its contribution to democratizing ML workflows.

Keywords- Automated Data Science, AutoDS, Machine Learning, Data Analysis, Predictive Analytics, Automated Machine Learning (AutoML)

I. INTRODUCTION

Machine learning is integral to many fields, yet complex technical requirements obstruct wide use, particularly for non-programmers. The AutoDS project addresses this gap by creating a generalized ML web application that automates key processes including data loading, preprocessing, exploratory analysis, model training, evaluation, and result reporting.

This facilitates users' focus on data insights rather than coding. The platform supports ML algorithms like decision trees, random forests, support vector machines, and neural networks, promoting flexibility for structured and unstructured data.

To address these challenges, the AutoDS (Automated Data Science) system has been developed as a comprehensive web-based platform that simplifies and automates the entire ML workflow. AutoDS aims to democratize machine learning by making it accessible to users with minimal or no programming experience.

1. Sub-Heading

- Complexity of machine learning workflows for non-programmers.
- Time-consuming manual data preprocessing and model training.
- Lack of automated tools for beginners in data science.
- Difficulty in extracting meaningful insights without technical expertise.

2. Sub-Heading

- Limited accessibility of machine learning tools for non-experts.
- Need for automated data validation and preprocessing.
- Challenges in selecting appropriate ML models.

- Lack of integrated platforms combining EDA, training, and reporting.

II. PROBLEM STATEMENT

Machine learning has become an essential tool across industries, yet its adoption is limited due to technical complexity and the need for programming expertise. Users often struggle with multiple steps such as data cleaning, feature engineering, model selection, and evaluation.

Existing tools either require coding knowledge or provide limited automation, making them unsuitable for beginners. Additionally, switching between different platforms for preprocessing, analysis, and model training increases inefficiency and time consumption.

There is a need for a unified system that automates the entire data science pipeline, from data ingestion to model deployment, while maintaining accuracy and usability.

AutoDS addresses these challenges by providing an integrated, automated, and user-friendly platform that simplifies machine learning workflows and enhances accessibility.

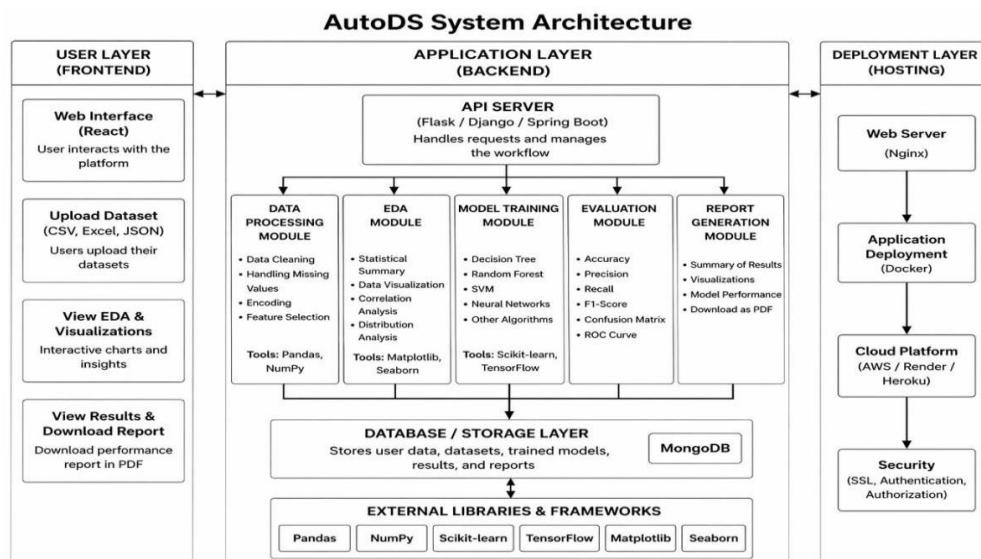


Figure 1: System Architecture

Figure 1: System Architecture

III. CONCLUSION

The AutoDS platform successfully simplifies machine learning workflows by providing an automated and user-friendly environment for data analysis and model development. It reduces the complexity of traditional ML processes by integrating data preprocessing, model training, evaluation, and reporting into a single system.

The platform improves efficiency, reduces manual effort, and enables users from diverse domains to leverage machine learning without requiring advanced technical knowledge.

Future enhancements may include support for unstructured data, real-time analytics, integration with advanced AutoML techniques, and improved scalability for handling large datasets.

REFERENCES

1. Feurer, M., & Hutter, F. (2019). Automated Machine Learning: Methods, Systems, Challenges.
2. Thornton, C., et al. (2013). Auto-WEKA: Combined Selection and Hyperparameter Optimization.
3. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System.
4. Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). Automated Machine Learning