# AutoML System for Hyperparameter Optimization

### ****Group Description****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. No. | Name | Division | PRN No. | Roll No. |
| 1 | Pravin Kumawat | C | 22210292 | 323042 |
| 2 | Gaurav Pawar | C | 22210838 | 323057 |
| 3 | Ambarish Satbhai | C | 22211511 | 323061 |
| 4 | Rugved Vaidya | C | 22211330 | 323073 |

### ****Introduction****

The AI-Based AutoML System for Hyperparameter Optimization is designed to automate the process of selecting the best machine learning model and its optimal hyperparameters. The system allows users to upload datasets, preprocess data, and optimize model selection through automated techniques. The final trained model can be downloaded for future use. This system simplifies machine learning model selection, making it accessible to users with limited expertise in data science.

The system provides a web-based interface using Flask, where users can upload datasets, select the target column, and automatically preprocess the data. The system supports various machine learning models such as Random Forest, XGBoost, LightGBM, Logistic Regression, and K-Nearest Neighbors. It performs automatic feature scaling, encoding of categorical variables, and missing value imputation. Additionally, the system uses hyperparameter tuning via RandomizedSearchCV to find the optimal model configuration. The trained models are saved for future predictions, allowing users to easily integrate them into their applications.

### ****Motivation****

Machine learning model development often requires extensive manual tuning of hyperparameters, which can be time-consuming and complex. Traditional methods such as trial-and-error or grid search can be inefficient and computationally expensive. Automating this process improves efficiency, ensures better model performance, and reduces human effort in hyperparameter selection.

By leveraging automated hyperparameter optimization techniques, this system democratizes machine learning by making model optimization accessible to non-experts. The integration of SMOTE ensures that imbalanced datasets are properly handled, leading to better generalization and robustness of trained models. Additionally, the use of Flask provides a seamless web-based experience, allowing users to interact with the system without writing code. The motivation behind this project is to bridge the gap between machine learning practitioners and efficient model development by providing an automated, easy-to-use solution.

### ****Problem Statement****

Choosing the best hyperparameters for machine learning models is a challenging task that impacts model accuracy and performance. Traditional methods involve trial-and-error, which are inefficient and do not guarantee optimal performance. This project addresses the problem by implementing an AutoML system that automates hyperparameter tuning using RandomizedSearchCV, ensuring optimal performance with minimal manual intervention.

Furthermore, handling imbalanced datasets and preprocessing features efficiently is another major challenge. The proposed system incorporates automated preprocessing techniques, such as missing value imputation, feature encoding, and standardization, to streamline data preparation. Additionally, it applies SMOTE for classification tasks to handle imbalanced datasets effectively. The system not only optimizes hyperparameters but also selects the best model based on performance metrics such as accuracy, F1-score, and root mean squared error (RMSE) for classification and regression tasks. By addressing these challenges, the project ensures that users can build high-performance machine learning models effortlessly.

### ****Brief Description****

This AutoML system automates the selection of hyperparameters for machine learning models using RandomizedSearchCV. It allows users to upload a dataset, select the target variable, preprocess data, and optimize models such as RandomForestClassifier, XGBoost, LightGBM, Logistic Regression, and KNN. The best model is trained, evaluated, and made available for download in .pkl format for future use.

The system is built using Flask and provides a user-friendly web interface. Users can upload datasets in CSV format, and the system automatically processes the data by handling missing values, encoding categorical features, and scaling numerical features. The system then applies hyperparameter tuning and model selection, identifying the best-performing model. The trained model is stored and can be used for real-time predictions in other applications.

The use of RandomizedSearchCV allows efficient hyperparameter tuning, ensuring that the model achieves the best possible performance. The inclusion of SMOTE addresses class imbalance issues, improving model robustness. The system supports various evaluation metrics and generates a summary of model performance, enabling users to make informed decisions about the trained model.

### ****Tools Used****

**Programming Language**: Python

**Framework**: Flask

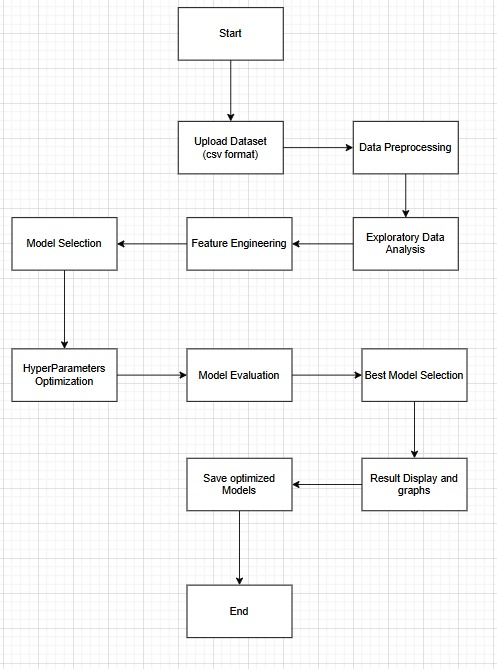
**Libraries**: Pandas, Scikit-learn, Optuna, Joblib, XGBoost, LightGBM, Imbalanced-learn

**Machine Learning Models**: Random Forest, XGBoost, LightGBM, Logistic Regression, KNN

**Development Tools**: VS Code

These tools were chosen for their efficiency and widespread use in machine learning applications. Flask provides a lightweight web framework for deploying the system, while Scikit-learn and other ML libraries ensure robust model development. The integration of Joblib allows efficient model storage and retrieval for real-time predictions.

### ****Flowchart****



### ****Why Automation?****

#### ****Efficiency****

Manual hyperparameter tuning is slow and inefficient. Automation speeds up the process by systematically searching for the best parameters using RandomizedSearchCV. The system evaluates multiple models simultaneously, reducing human intervention.

#### ****Improved Performance****

Automated optimization techniques outperform traditional grid search and random search. The integration of SMOTE ensures better handling of imbalanced datasets, leading to more robust and accurate models.

#### ****Accessibility****

Non-experts can train high-performing machine learning models without extensive ML knowledge. The Flask-based interface ensures an intuitive user experience, making machine learning accessible to a broader audience.

#### ****Scalability****

Automation allows rapid experimentation with different models and datasets, making it easier to scale AI solutions in real-world applications. The ability to download trained models in .pkl format ensures that the models can be deployed in various environments effortlessly.

### ****Conclusion****

The AI-Based AutoML System for Hyperparameter Optimization is an efficient and user-friendly tool for selecting and tuning machine learning models. By automating the tedious process of hyperparameter tuning, it significantly reduces human effort and enhances model performance. The system integrates Flask for seamless interaction, ensuring accessibility for both technical and non-technical users. With its ability to handle preprocessing, model selection, and optimization, this system simplifies machine learning workflows and enables faster deployment of AI solutions.