

Anomaly Detection Analysis

This repository contains a Jupyter Notebook for performing anomaly detection using machine learning techniques. The project primarily focuses on data preprocessing, exploratory data analysis, and building a Random Forest Classifier to detect anomalies.

Project Overview

The notebook performs the following key steps:

1. **Data Loading:** The data is loaded from an Excel file named AnomaData.xlsx.
2. **Data Preprocessing:**
 - Converts relevant columns to appropriate data types.
 - Handles missing values through dropping or imputation.
 - Scales the data using StandardScaler to standardize the features.
3. **Exploratory Data Analysis (EDA):**
 - Examines data quality and identifies missing values.
 - Performs visualizations to understand data distributions and anomalies.
4. **Model Training:**
 - Splits the data into training and test sets.
 - Trains a Random Forest Classifier to detect anomalies.
5. **Model Evaluation:**
 - Evaluates the model performance using metrics such as accuracy, precision, recall, and F1-score.

Requirements

To run this notebook, you need the following dependencies:

- Python 3.x
- pandas
- numpy
- scikit-learn
- matplotlib
- seaborn

You can install the required packages using the following command:

```
bash
```

Copy code

```
pip install pandas numpy scikit-learn matplotlib seaborn
```

How to Use

1. Clone this repository to your local machine.
2. Ensure that the data file AnomaData.xlsx is present in the root directory.
3. Open the Jupyter Notebook and run the cells sequentially to reproduce the analysis.

File Structure

- Anamolous.ipynb: Main notebook containing the anomaly detection analysis.
- AnomaData.xlsx: Input data file used for training and testing the model.

Results

The results of the analysis include the identification of anomalies in the dataset and the performance metrics of the Random Forest model.

Contributing

Feel free to contribute to this project by forking the repository and submitting pull requests.

License

This project is licensed under the MIT License.