Honeywell

Multivariate Anomaly Detection for IoT Medical Sensors

SUBMITTED BY –

RONIT SINGH

DEPLOYED PROJECT LINK: https://ai-anomaly-detection.streamlit.app/

TABLE OF CONTENTS

1.INTRODUCTION

1.1 Project Overview

The primary goal of this project is to provide a robust, user-friendly platform for the real-time detection of anomalies in multivariate time series data sourced from Internet of Things (IoT) sensors within the medical field. In modern healthcare, continuous patient monitoring and medical device supervision generate vast streams of complex data. This platform addresses the critical need to automatically analyse this data to identify subtle, clinically significant deviations from normal physiological or operational baselines. The solution is an interactive web application, built with Stream-lit, that leverages powerful machine learning algorithms to uncover hidden patterns and flag potential issues. This enables a crucial shift from reactive to proactive clinical intervention and predictive equipment maintenance.

1.2 Key Features

1.2.1 Medical IoT Focus: The platform is specifically tailored for the unique challenges of medical sensor data, such as heart rate, SpO2, blood pressure, and device temperature.

1.2.2 Multiple ML Models: The system integrates several proven anomaly detection algorithms, including Isolation Forest, PCA-based detection, and Autoencoders, allowing users to select the most appropriate model for their specific dataset.

1.2.3 Intuitive Clinical UI: A clean, interactive user interface is designed for clinical researchers and biomedical engineers, enabling seamless data upload, configuration, and analysis without requiring deep programming knowledge.

1.2.4 Explainable AI (XAI): The platform goes beyond simple detection by providing feature attribution analysis, which pinpoints the specific sensor readings (e.g., "a sudden drop in pressure") that contributed most to a detected anomaly.

1.2.5 Advanced Visualizations: It features a comprehensive analytics dashboard with interactive timelines, severity distribution charts, and detailed tables to facilitate in-depth exploration of anomalous events.

1.2.6 Real-time Alerting: The application includes an integrated alert system capable of sending SMS notifications via Twilio when severe anomalies are detected, enabling immediate response.

1.3 Target Audience

This document and the accompanying platform are designed for a professional audience, including:

1.3.1 Biomedical Engineers: For monitoring and maintaining the performance and safety of medical devices.

1.3.2 Data Scientists & Clinical Researchers: For analysing patient sensor data to uncover trends and predict adverse health events.

1.3.3 Healthcare IT Professionals: For deploying and managing data-driven monitoring tools within a clinical environment.

2.SYSTEM ARCHITECTURE

2.1. Architectural Diagram
This is the most important part of the section. You should create a simple flowchart or block diagram that visually represents how data moves through your application and how the different Python modules interact with each other.
Example Flow: **User Interface (app.py) → Data Processing (data_processor.py) → ML Model (anomaly_models.py) → Feature Attribution (feature_attribution.py) → Results & UI (app.py)**

2.2. Module Descriptions Here, you'll briefly explain the role of each Python file, giving context to the diagram above.

2.2.1 **app.py:** This is the main entry point of the application. It handles the user interface, manages the overall workflow, and calls the other modules to perform specific tasks.

2.2.2 **data_processor.py:** This module is responsible for all data preparation, including loading the CSV file, cleaning the data (e.g., handling missing values), and preprocessing it for the machine learning model.

2.2.3 **anomaly_models.py:** This file contains the core machine learning logic. It includes the different algorithms (Isolation Forest, PCA, etc.) used to detect anomalies in the sensor data.

2.2.4 **feature_attribution.py:** After an anomaly is detected, this module analyses it to determine which specific sensors or features were the most significant contributors, providing explainability for the results.

2.2.5 **utils.py:** This is a collection of helper functions used across the entire application for common tasks like formatting scores, validating data, and creating summary reports.

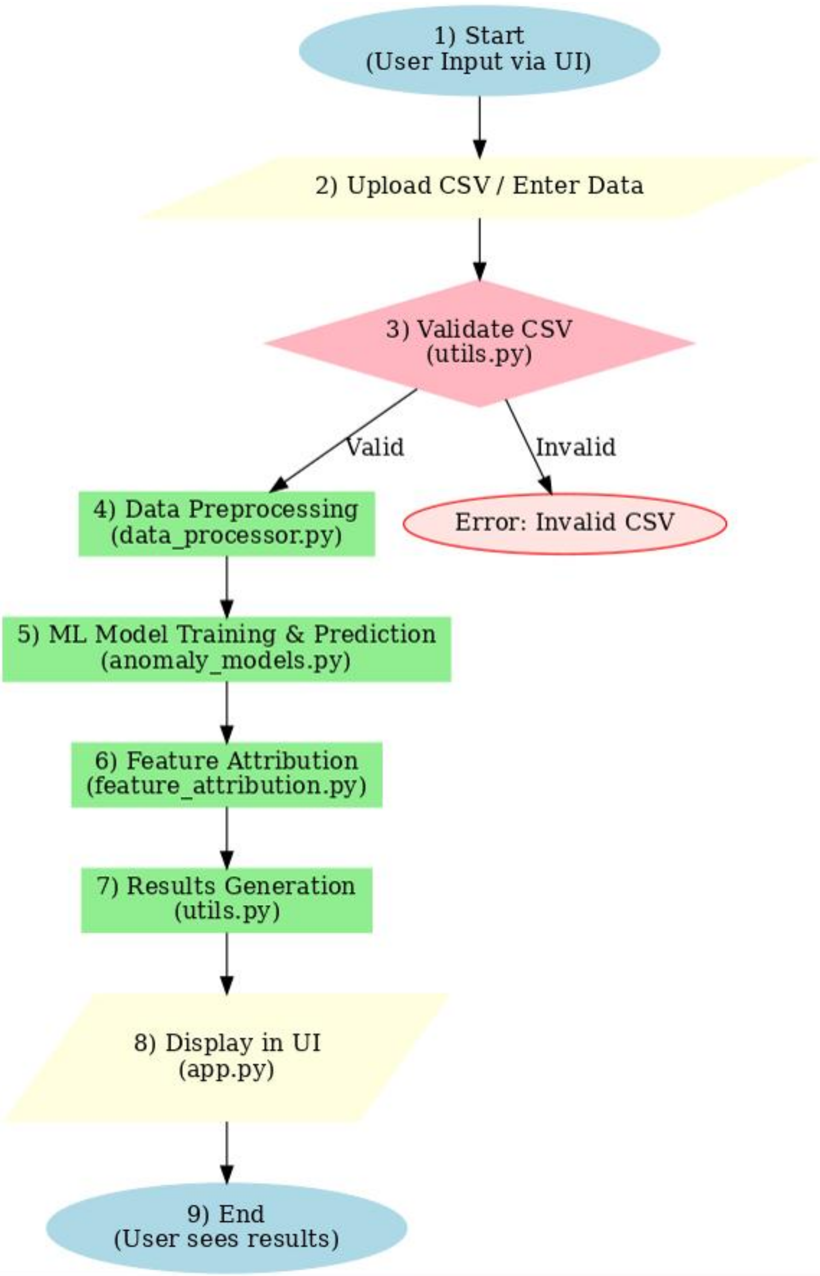


Fig: Application logic and data flow diagram

3.USER GUIDE

3.1. Prerequisites

3.1.1 Python: Version 3.8 or higher.

3.1.2 pip: The Python package installer.

3.2. Technical Stack & Setup

This project is built with the following technologies. All required libraries can be installed via a requirements.txt file.

3.2.1 Programming Language

Python is the primary language used for development.

3.2.2 Libraries & Dependencies To install all necessary libraries, create a file named requirements.txt, paste the content below into it, and then run the command pip install -r requirements.txt.

a) Numpy 1.26.4
b) Pandas 2.2.3
c) scikit-learn 1.5.2
d) tensorflow 2.17.0
e) Plotly 5.24.1
f) Streamlit 1.38.0
g) python-dotenv 1.0.1
h) Twilio 9.7.1

- Numpy : Used for efficient numerical computations and array operations.
- pandas: Powers high-performance data manipulation and preprocessing.
- scikit-learn: Implements traditional ML models like Isolation Forest and PCA.
- Tensorflow: The deep learning framework used for Autoencoder and LSTM models.
- Plotly: Creates the interactive visualizations in the dashboard.
- Streamlit: Builds the frontend web application and user interface.
- twilio: Manages sending SMS alerts when critical anomalies are detected.

iii) Hardware Requirements

a) The application runs on **standard computing resources** (CPU/GPU).

b) No specialized hardware is required, but performance on large datasets will benefit from a more powerful CPU or a compatible GPU for TensorFlow models.

iv) SMS Notifications

a) The platform integrates with the **Twilio API** for sending real-time SMS alerts.

b) Alerts are automatically triggered when a calculated anomaly score exceeds a threshold of **90**.

3.3. Running the Application

Once the dependencies are installed, you can launch the application with the following command:

Bash

Stream-lit run app.py

3.4. Project Resources & Links

a) **GitHub Repository:** https://github.com/ronitsingh2704/ai-anomaly

b) **Deployed Application:** https://ai-anomaly-detection.streamlit.app/

c) **Dataset:** https://drive.google.com/file/d/1v3u5qxUR616vZjJH7gflpKyKCzzB45wc/view?usp=sharing

3.5. Step-by-Step Walkthrough

1. Data Upload: Explain how to upload a CSV file and the required data format.

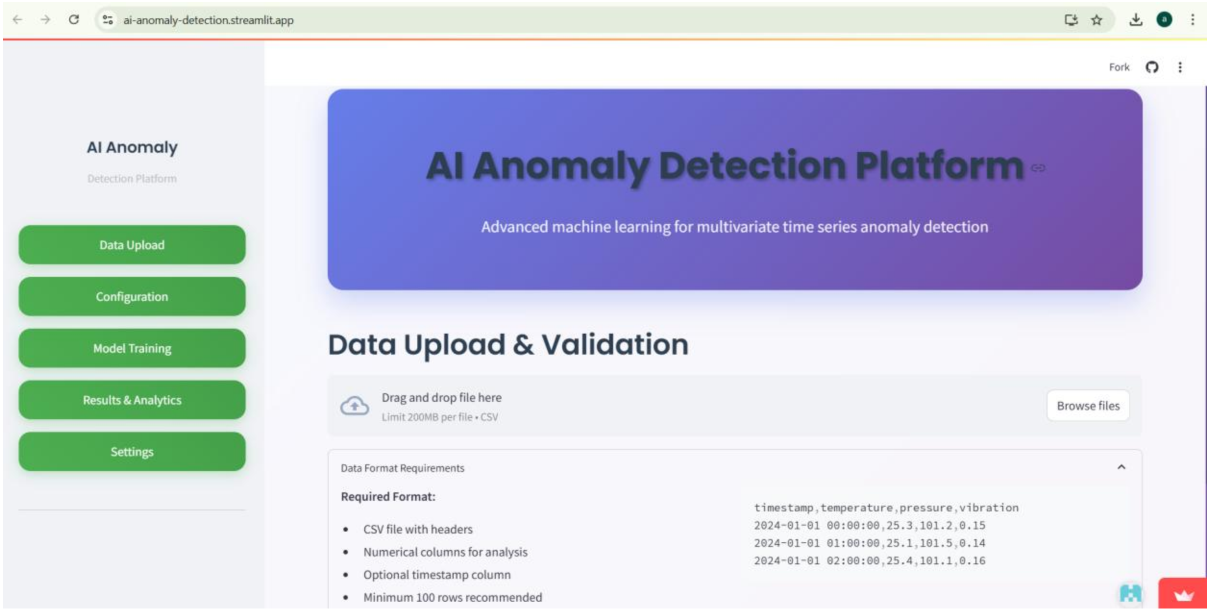
2. Configuration: Guide the user on selecting features, defining the training period, and choosing an algorithm.

3. Model Training: Describe what happens when the user clicks the "Start Training" button.

4. Results & Analytics: Explain how to interpret the results dashboard.

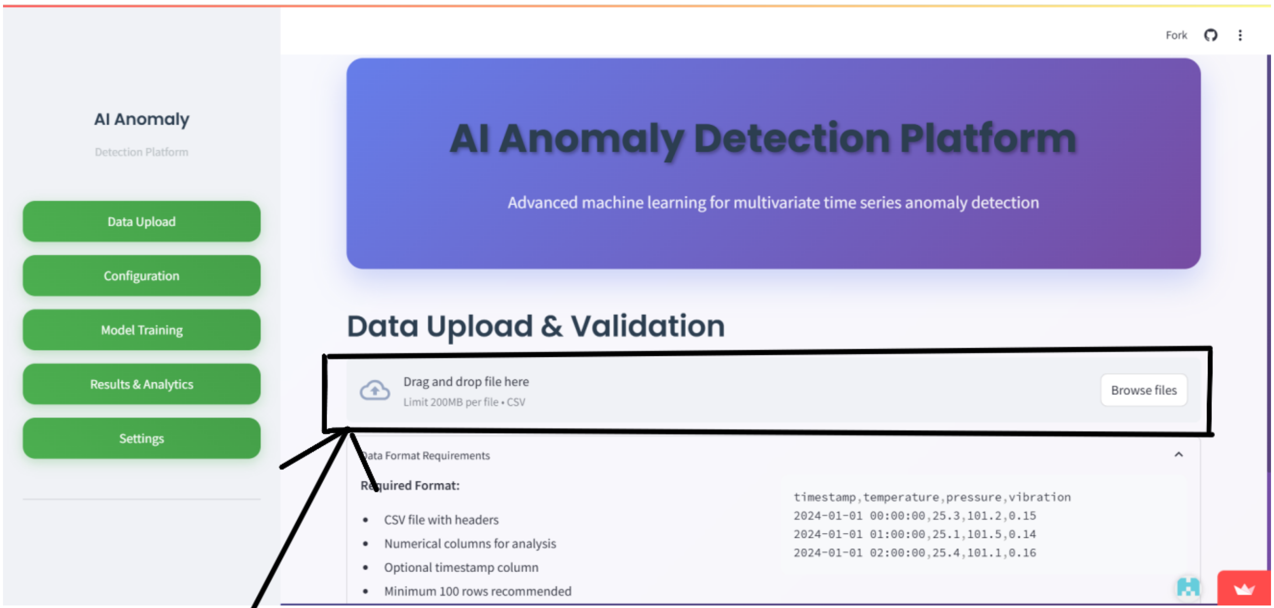
3.6 STEP BY STEP ANALYSIS

3.6.1 The main landing page of the AI Anomaly Detection Platform.



i) Live Link: https://ai-anomaly-detection.streamlit.app/

3.6.2 Uploading Dataset



j) **DATASET:** https://drive.google.com/file/d/1v3u5qxUR616vZjJH7gfIpKyKCzzB45wc/view?usp=sharing

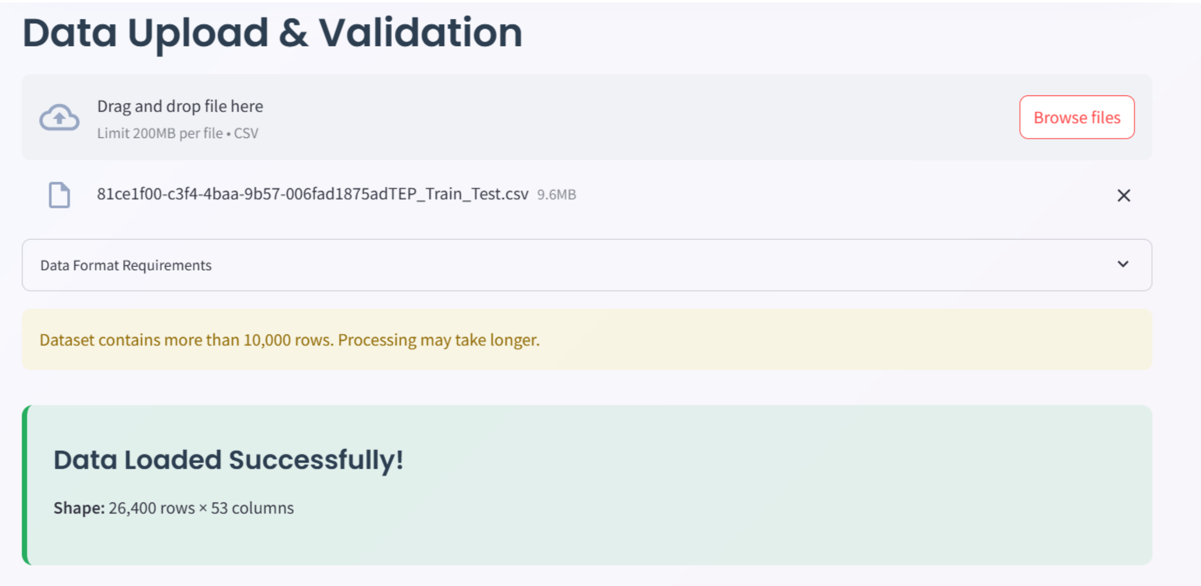
k) The dataset is a multivariate time-series CSV file containing simulated readings from medical IoT sensors, such as temperature and pressure, designed for anomaly detection.

How to Upload

a) Drag and drop your CSV dataset file directly onto the upload area.

b) Alternatively, click the "Browse files" button to select the dataset from your computer.

c) Once uploaded, a preview of the data will appear for validation.



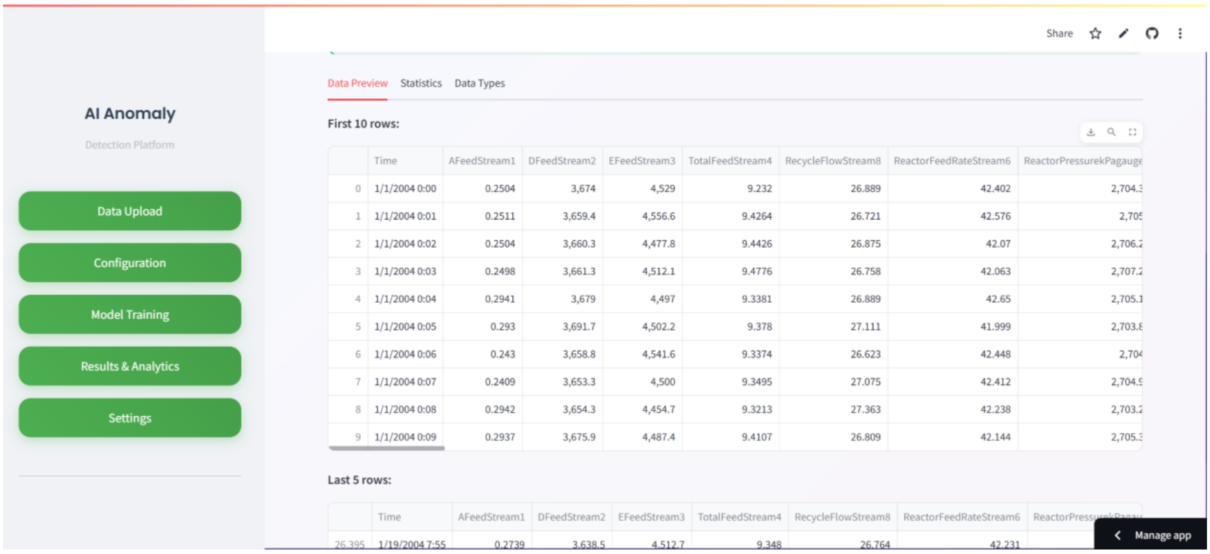
3.6.3 Data Preview:

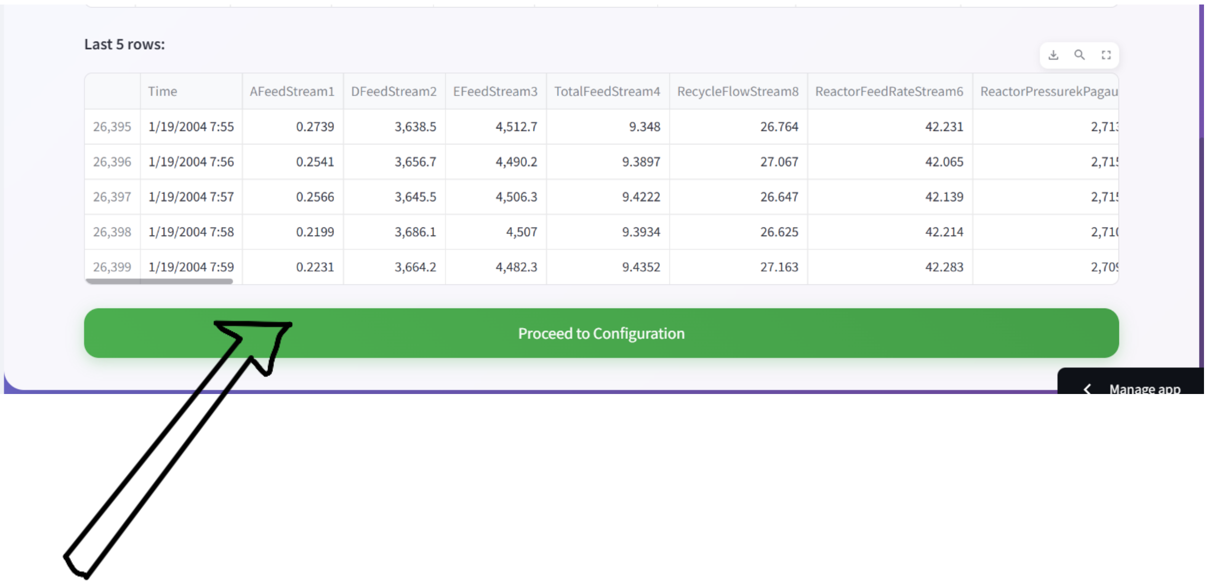
In this preview, you should verify the following:

a) **Correct Columns:** Ensure that all column headers are present and have been read correctly.

b) **Data Alignment:** Check that the data in the rows appears in the correct columns (e.g., numbers are under the temperature column, dates are under timestamp).

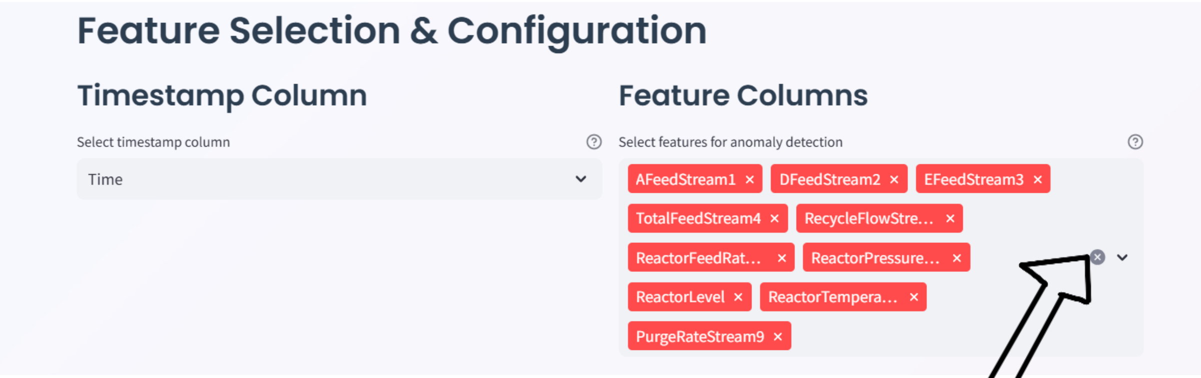
c) **Successful Parsing:** Confirm that the table appears as expected, which indicates a successful upload.



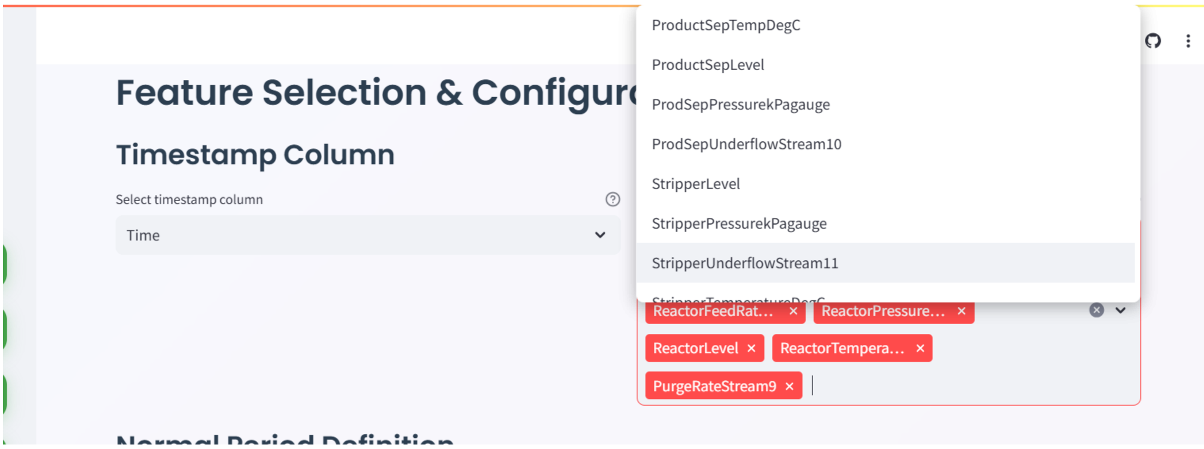


- Once you have verified the data, proceed to the **Configuration** page to set up the analysis parameters.

3.6.4 Feature Selection and configuration

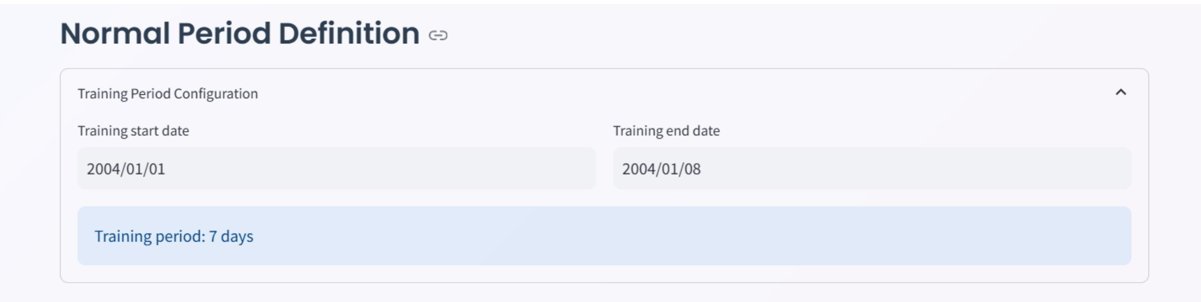


a) Feature Columns: In this multi-select box, you choose the specific numeric columns (sensor readings) that you want the model to analyse for anomalies. You can select one or multiple features.



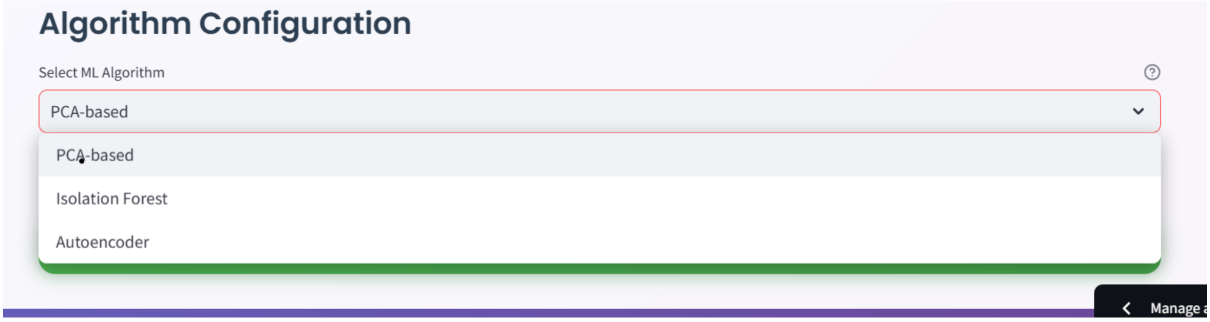
3.6.5 Normal Period Definition:

i) Specify the date range or row indices that represent a known "normal" or healthy baseline period. The machine learning model will be trained exclusively on this portion of the data.



3.6.6 Algorithm Selection:

i) Choose the machine learning algorithm you want to use for detection (e.g., **Isolation Forest**, **PCA**, **Autoencoder**).

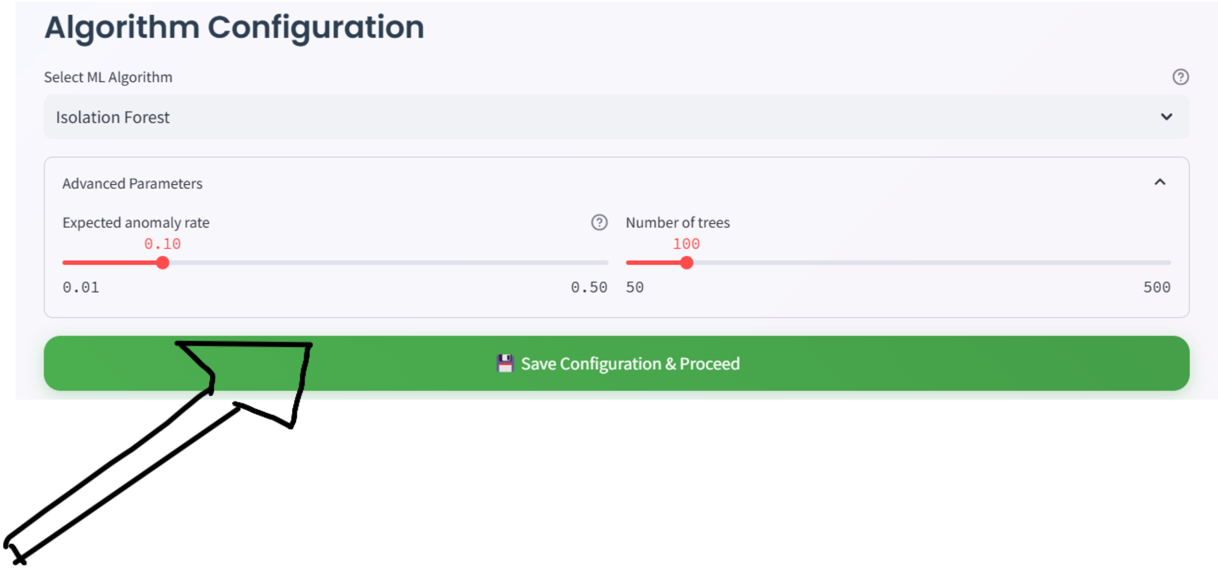


i)PCA-Principal Component Analysis (PCA) is a dimensionality reduction technique that detects anomalies by identifying data points with a high **reconstruction error** when trying to rebuild them from the learned patterns of normal data.

ii)Isolation Forest - Isolation Forest is an algorithm that isolates data points using random trees, identifying anomalies because they are **"few and different"** and therefore require fewer splits to be separated from normal data.

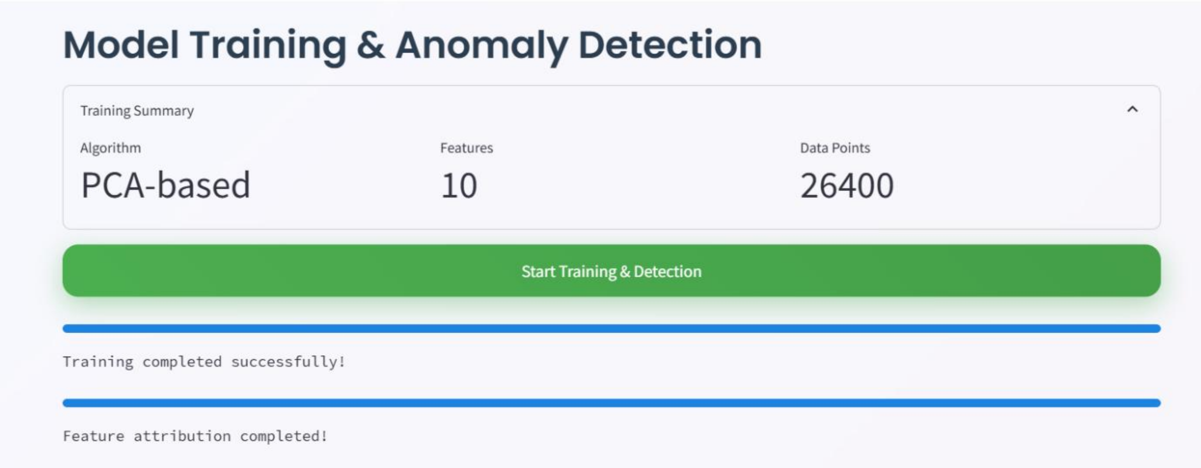
iii)Autoencoders - An **Autoencoder** is a neural network trained to reconstruct its input, detecting anomalies as data points that it fails to rebuild accurately (resulting in a high **reconstruction error**) because they don't match the patterns of the normal data it was trained on.

ii) **Advanced Parameters:** This section allows you to fine-tune the behavior of the selected algorithm. For example, you can adjust the **"Expected anomaly rate"** for an Isolation Forest or the **"Number of PCA Components"** for a PCA model. These settings allow you to customize the model's sensitivity and complexity.



Once you have set all the parameters, click the **'Save Configuration & Proceed'** button to begin the model training process.

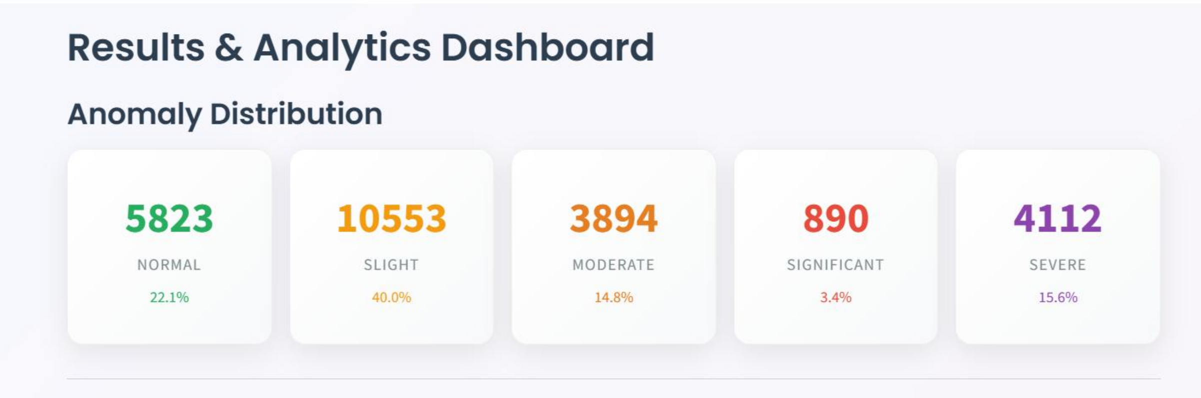
3.6.6 Model training



3.6.7 Results and analytics

i) Anomaly Distribution Summary

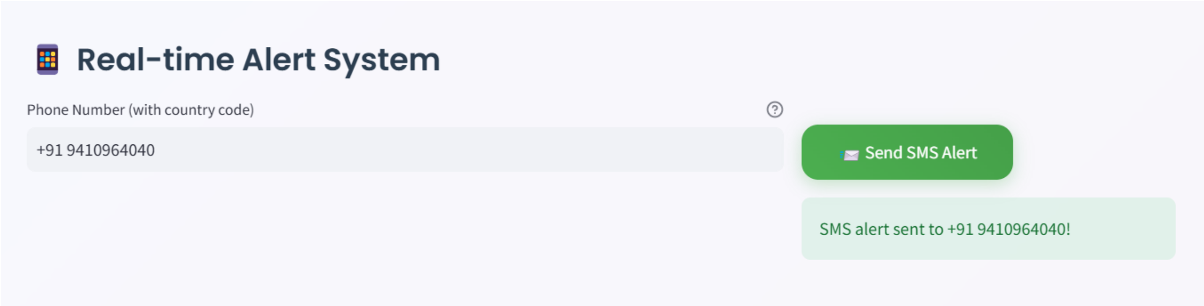
At the top of the dashboard, you will find metric cards that provide a high-level summary of the results. These cards display the total count and percentage of data points categorized by their calculated severity, giving you an immediate overview of the dataset's overall health. The categories include Normal, Slight, Moderate, Significant, and Severe.



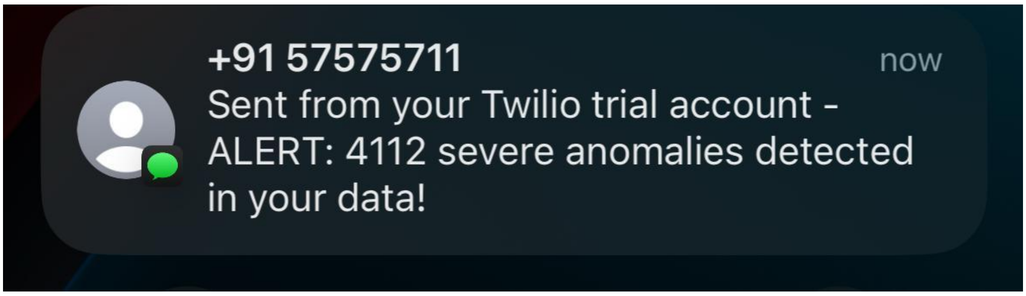
ii) **Real-time SMS Alert System**

The platform includes a real-time alert system to notify you of critical anomalies. You can enter a phone number in the input field and click the **"Send SMS Alert"** button. An alert is sent if the analysis has detected any "Severe" anomalies (score > 90), summarizing the number of critical events found.

Important Note: The integrated Twilio service can only send SMS messages to phone numbers that have been verified in your Twilio account. For testing and demonstration, please use a pre-verified number, such as +91 9410964040.



Received Message:



iii)**Interactive Visualizations**

The dashboard includes several interactive graphs to help you explore the results in detail:

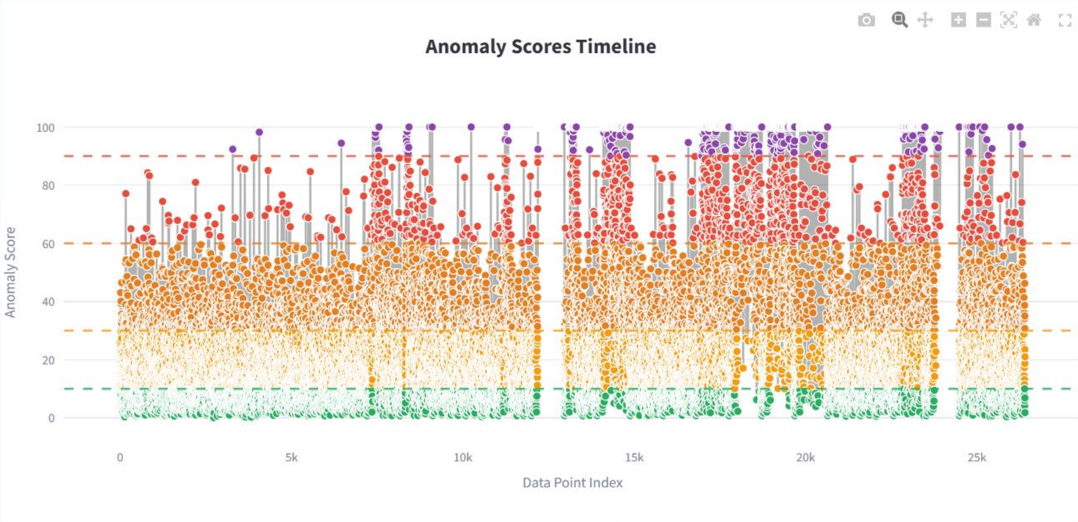
a) **Anomaly Scores Timeline:** This is the main line chart that plots the Abnormality_score for every data point over time. It allows you to pinpoint the exact moments when anomalies occurred and observe their severity.

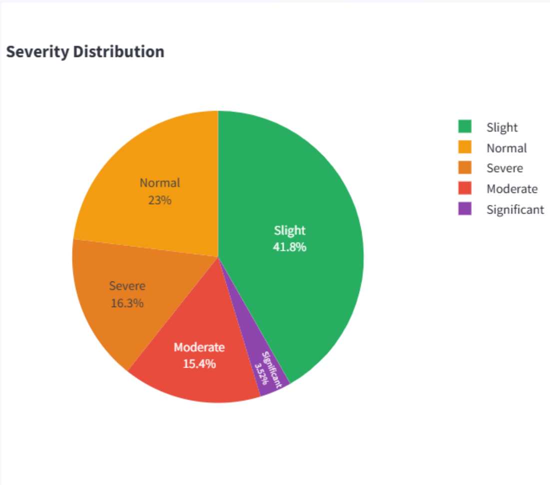
b) **Top Anomalies Bar Chart:** This horizontal bar chart displays the data points with the highest anomaly scores. It helps you quickly identify and focus on the most critical events that require immediate attention.

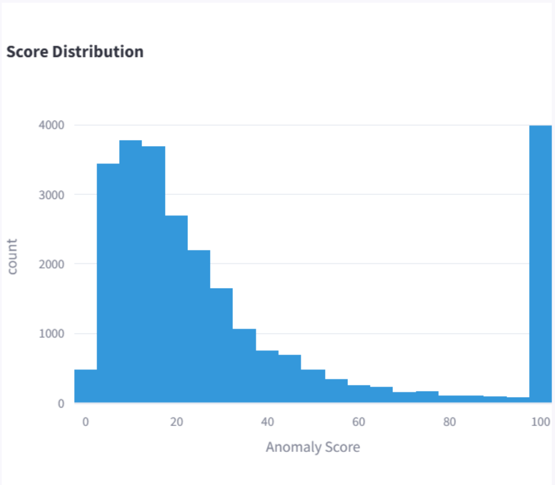
c) **Feature Contribution Analysis:** This chart shows which features (sensors) were the most frequent root causes of the detected anomalies. Use this to understand *why* an anomaly occurred, which is crucial for diagnosis.

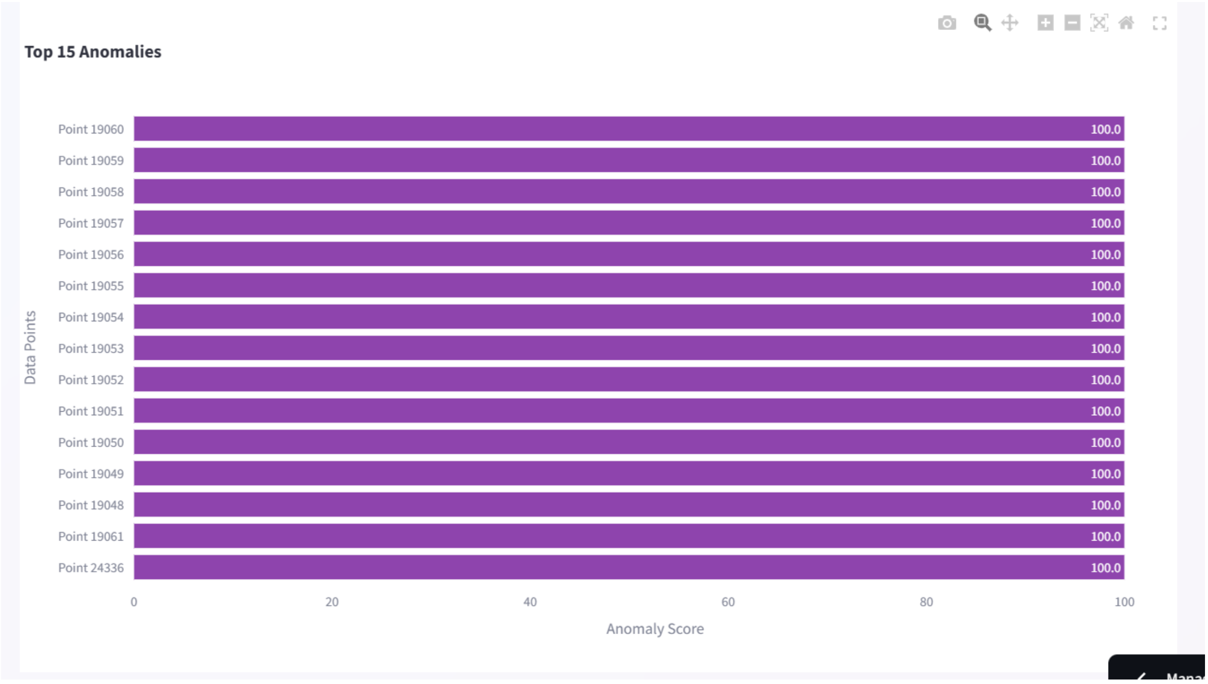
d) **Severity Distribution Pie Chart:** This pie chart provides a visual breakdown of the percentage of anomalies in each severity category, offering a clear view of the overall proportion of alerts.

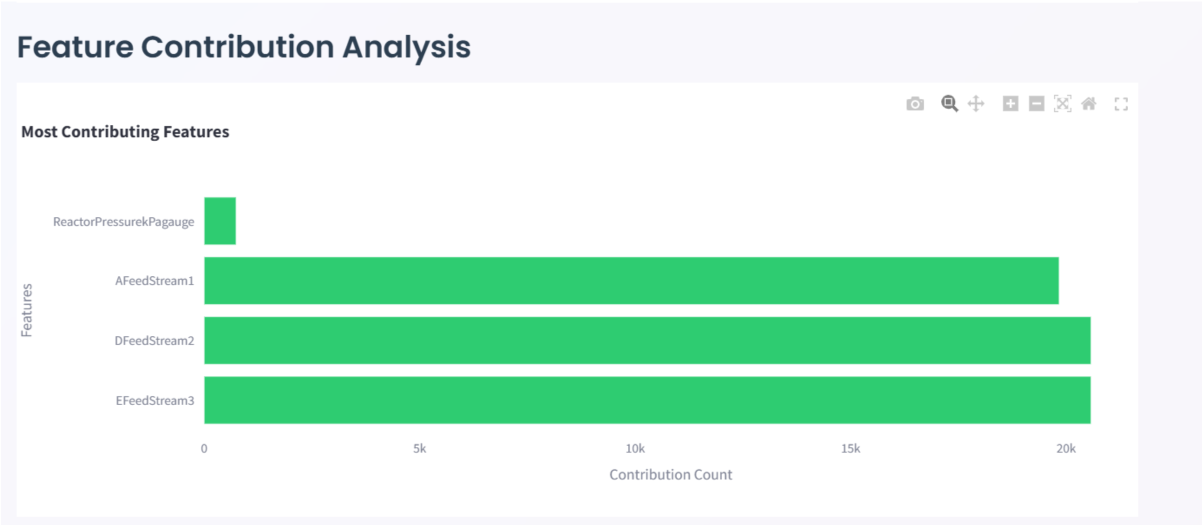
Interactive Visualizations









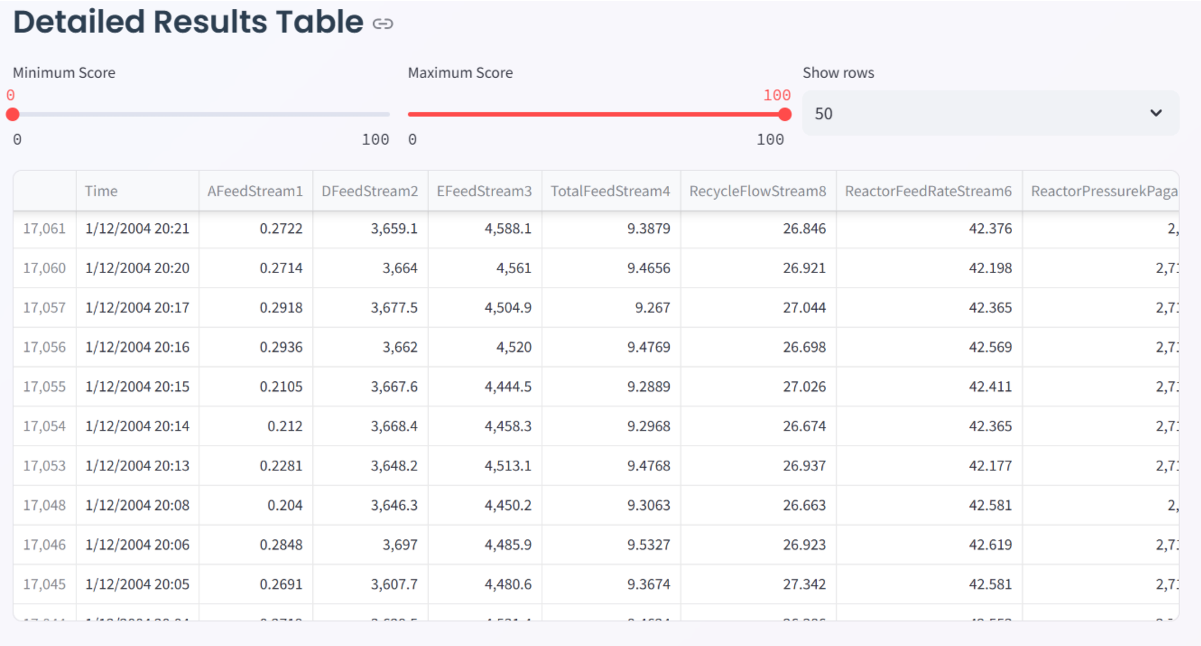


e) Detailed Results Table

At the bottom of the dashboard is an interactive table that provides the most granular view of the results. It contains your original data enriched with the new output columns: Abnormality_score and the top contributing features (top_feature_1, etc.).
Above the table, you will find several controls that allow you to customize the view:

Adjust by Anomaly Score: You can filter the results using the "Minimum Score" and "Maximum Score" sliders to view only the data points that fall within a specific score range.

Adjust Rows Displayed: You can control how many rows you need to see by selecting a value from the "Show rows" dropdown menu. This is useful for focusing on a specific number of top anomalies.



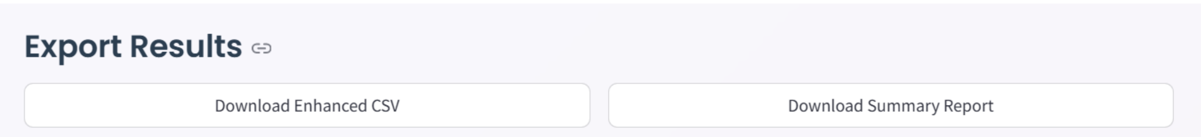
	Time	AFeedStream1	DFeedStream2	EFeedStream3	TotalFeedStream4	RecycleFlowStream8	ReactorFeedRateStream6	ReactorPressurekPaga
17,061	1/12/2004 20:21	0.2722	3,659.1	4,588.1	9.3879	26.846	42.376	2,
17,060	1/12/2004 20:20	0.2714	3,664	4,561	9.4656	26.921	42.198	2,7
17,057	1/12/2004 20:17	0.2918	3,677.5	4,504.9	9.267	27.044	42.365	2,7
17,056	1/12/2004 20:16	0.2936	3,662	4,520	9.4769	26.698	42.569	2,7
17,055	1/12/2004 20:15	0.2105	3,667.6	4,444.5	9.2889	27.026	42.411	2,7
17,054	1/12/2004 20:14	0.212	3,668.4	4,458.3	9.2968	26.674	42.365	2,7
17,053	1/12/2004 20:13	0.2281	3,648.2	4,513.1	9.4768	26.937	42.177	2,7
17,048	1/12/2004 20:08	0.204	3,646.3	4,450.2	9.3063	26.663	42.581	2,
17,046	1/12/2004 20:06	0.2848	3,697	4,485.9	9.5327	26.923	42.619	2,7
17,045	1/12/2004 20:05	0.2691	3,607.7	4,480.6	9.3674	27.342	42.581	2,7

f) **Export Results**

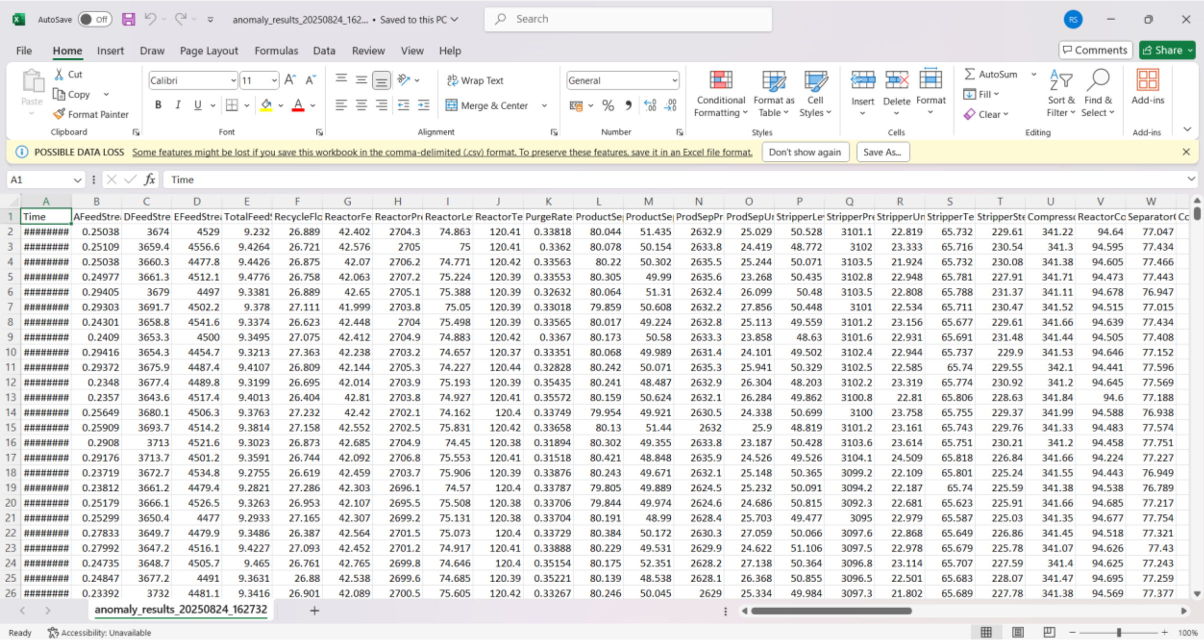
The platform allows you to export your analysis for offline use or reporting. You have two functions for downloading the results, available as buttons at the bottom of the page:

i) **Download Enhanced CSV:** This button exports the complete dataset as a CSV file. This file includes all of your original data columns along with the newly generated Abnormality_score and the seven top_feature columns, making it ideal for further analysis in other tools like Excel or Tableau.

ii)**Download Summary Report:** This button generates a human-readable text file (.txt) that summarizes the key findings of the analysis. This report includes overall statistics, the distribution of anomalies by severity, and a list of the top 10 most critical anomalies, making it perfect for quick reports and stakeholder updates.



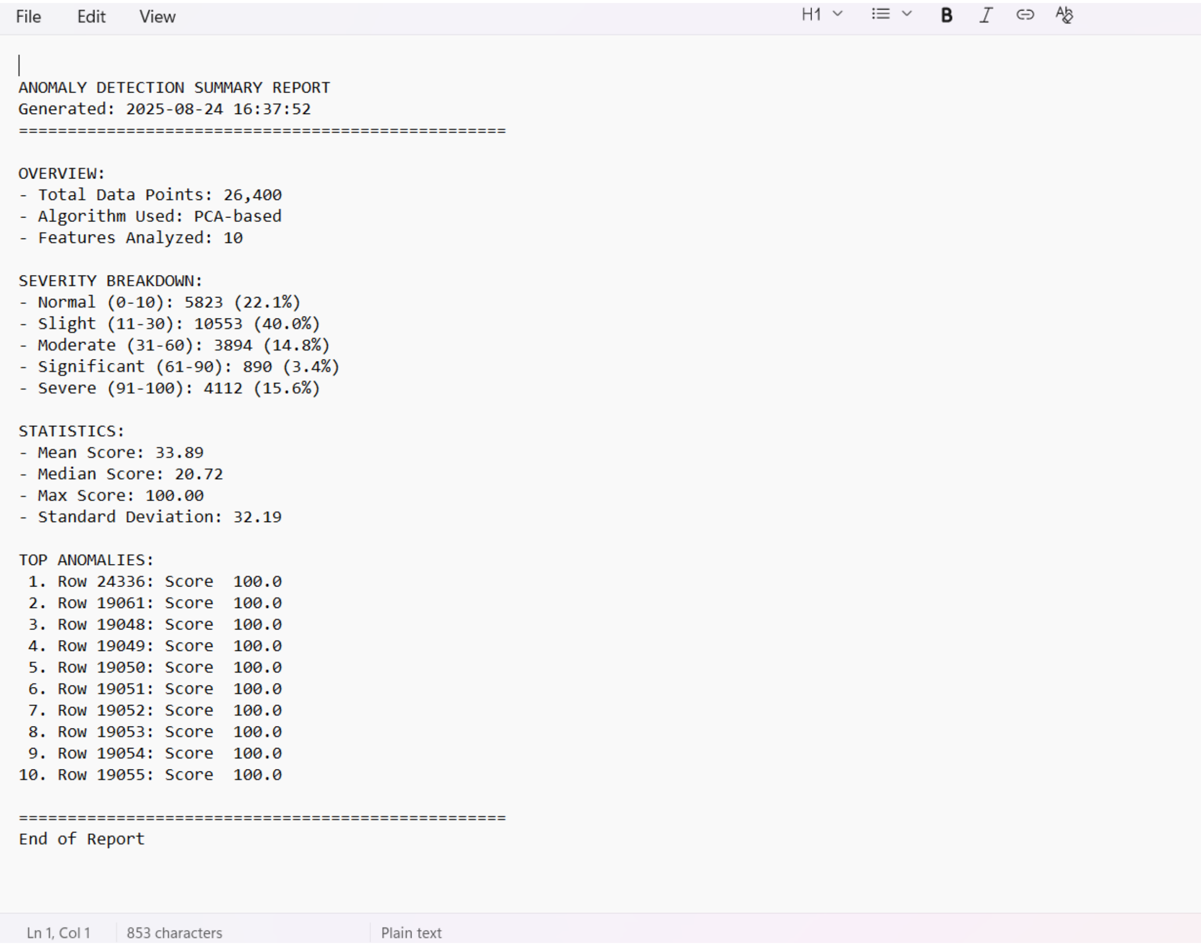
Downloaded Enhanced CSV FILE →



	A	B	C	D	E	F	G	H	I	J	K	L
1	Time	AFeedStre	DFeedStre	EFeedStre	TotalFeed	RecycleFlo	ReactorFe	ReactorPr	ReactorLe	ReactorTe	PurgeRate	ProductSe
2	########	0.25038	3674	4529	9.232	26.889	42.402	2704.3	74.863	120.41	0.33818	80.044
3	########	0.25109	3659.4	4556.6	9.4264	26.721	42.576	2705	75	120.41	0.3362	80.078
4	########	0.25038	3660.3	4477.8	9.4426	26.875	42.07	2706.2	74.771	120.42	0.33563	80.22
5	########	0.24977	3661.3	4512.1	9.4776	26.758	42.063	2707.2	75.224	120.39	0.33553	80.305
6	########	0.29405	3679	4497	9.3381	26.889	42.65	2705.1	75.388	120.39	0.32632	80.064
7	########	0.29303	3691.7	4502.2	9.378	27.111	41.999	2703.8	75.05	120.39	0.33018	79.859
8	########	0.24301	3658.8	4541.6	9.3374	26.623	42.448	2704	75.498	120.39	0.33565	80.017
9	########	0.2409	3653.3	4500	9.3495	27.075	42.412	2704.9	74.883	120.42	0.3367	80.173
10	########	0.29416	3654.3	4454.7	9.3213	27.363	42.238	2703.2	74.657	120.37	0.33351	80.068
11	########	0.29372	3675.9	4487.4	9.4107	26.809	42.144	2705.3	74.227	120.44	0.32828	80.242
12	########	0.2348	3677.4	4489.8	9.3199	26.695	42.014	2703.9	75.193	120.39	0.35435	80.241
13	########	0.2357	3643.6	4517.4	9.4013	26.404	42.81	2703.8	74.927	120.41	0.35572	80.159
14	########	0.25649	3680.1	4506.3	9.3763	27.232	42.42	2702.1	74.162	120.4	0.33749	79.954
15	########	0.25909	3693.7	4514.2	9.3814	27.158	42.552	2702.5	75.831	120.42	0.33658	80.13
16	########	0.2908	3713	4521.6	9.3023	26.873	42.685	2704.9	74.45	120.38	0.31894	80.302
17	########	0.29176	3713.7	4501.2	9.3591	26.744	42.092	2706.8	75.553	120.41	0.31518	80.421
18	########	0.23719	3672.7	4534.8	9.2755	26.619	42.459	2703.7	75.906	120.39	0.33876	80.243
19	########	0.23812	3661.2	4479.4	9.2821	27.286	42.303	2696.1	74.57	120.4	0.33787	79.805
20	########	0.25179	3666.1	4526.5	9.3263	26.953	42.107	2695.5	75.508	120.38	0.33706	79.844
21	########	0.25299	3650.4	4477	9.2933	27.165	42.307	2699.2	75.131	120.38	0.33704	80.191
22	########	0.27833	3649.7	4479.9	9.3486	26.387	42.564	2701.5	75.073	120.4	0.33729	80.384
23	########	0.27992	3647.2	4516.1	9.4227	27.093	42.452	2701.2	74.917	120.41	0.33888	80.229
24	########	0.24735	3648.7	4505.7	9.465	26.761	42.765	2699.8	74.646	120.4	0.35154	80.175
25	########	0.24847	3677.2	4491	9.3631	26.88	42.538	2699.6	74.685	120.39	0.35221	80.139
26	########	0.23392	3732	4481.1	9.3416	26.901	42.089	2700.5	75.605	120.42	0.33267	80.246

	M	N	O	P	Q	R	S	T	U	V	W
1	ProductSe	ProdSepPr	ProdSepUr	StripperLe	StripperPr	StripperUn	StripperTe	StripperSt	Compress	ReactorCo	Separator
2	51.435	2632.9	25.029	50.528	3101.1	22.819	65.732	229.61	341.22	94.64	77.047
3	50.154	2633.8	24.419	48.772	3102	23.333	65.716	230.54	341.3	94.595	77.434
4	50.302	2635.5	25.244	50.071	3103.5	21.924	65.732	230.08	341.38	94.605	77.466
5	49.99	2635.6	23.268	50.435	3102.8	22.948	65.781	227.91	341.71	94.473	77.443
6	51.31	2632.4	26.099	50.48	3103.5	22.808	65.788	231.37	341.11	94.678	76.947
7	50.608	2632.2	27.856	50.448	3101	22.534	65.711	230.47	341.52	94.515	77.015
8	49.224	2632.8	25.113	49.559	3101.2	23.156	65.677	229.61	341.66	94.639	77.434
9	50.58	2633.3	23.858	48.63	3101.6	22.931	65.691	231.48	341.44	94.505	77.408
10	49.989	2631.4	24.101	49.502	3102.4	22.944	65.737	229.9	341.53	94.646	77.152
11	50.071	2635.3	25.941	50.329	3102.5	22.585	65.74	229.55	342.1	94.441	77.596
12	48.487	2632.9	26.304	48.203	3102.2	23.319	65.774	230.92	341.2	94.645	77.569
13	50.624	2632.1	26.284	49.862	3100.8	22.81	65.806	228.63	341.84	94.6	77.188
14	49.921	2630.5	24.338	50.699	3100	23.758	65.755	229.37	341.99	94.588	76.938
15	51.44	2632	25.9	48.819	3101.2	23.161	65.743	229.76	341.33	94.483	77.574
16	49.355	2633.8	23.187	50.428	3103.6	23.614	65.751	230.21	341.2	94.458	77.751
17	48.848	2635.9	24.526	49.526	3104.1	24.509	65.818	226.84	341.66	94.224	77.227
18	49.671	2632.1	25.148	50.365	3099.2	22.109	65.801	225.24	341.55	94.443	76.949
19	49.889	2624.5	25.232	50.091	3094.2	22.187	65.74	225.59	341.38	94.538	76.789
20	49.974	2624.6	24.686	50.815	3092.3	22.681	65.623	225.91	341.66	94.685	77.217
21	48.99	2628.4	25.703	49.477	3095	22.979	65.587	225.03	341.35	94.677	77.754
22	50.172	2630.3	27.059	50.066	3097.6	22.868	65.649	226.86	341.45	94.518	77.321
23	49.531	2629.9	24.622	51.106	3097.5	22.978	65.679	225.78	341.07	94.626	77.43
24	52.351	2628.2	27.138	50.364	3096.8	23.114	65.707	227.59	341.4	94.625	77.243
25	48.538	2628.1	26.368	50.855	3096.5	22.501	65.683	228.07	341.47	94.695	77.259
26	50.045	2629	25.334	49.984	3097.3	21.802	65.689	227.78	341.38	94.569	77.377

Downloaded Report Summary→



ANOMALY DETECTION SUMMARY REPORT
Generated: 2025-08-24 16:37:52
==

OVERVIEW:
- Total Data Points: 26,400
- Algorithm Used: PCA-based
- Features Analyzed: 10

SEVERITY BREAKDOWN:
- Normal (0-10): 5823 (22.1%)
- Slight (11-30): 10553 (40.0%)
- Moderate (31-60): 3894 (14.8%)
- Significant (61-90): 890 (3.4%)
- Severe (91-100): 4112 (15.6%)

STATISTICS:
- Mean Score: 33.89
- Median Score: 20.72
- Max Score: 100.00
- Standard Deviation: 32.19

TOP ANOMALIES:
1. Row 24336: Score 100.0
2. Row 19061: Score 100.0
3. Row 19048: Score 100.0
4. Row 19049: Score 100.0
5. Row 19050: Score 100.0
6. Row 19051: Score 100.0
7. Row 19052: Score 100.0
8. Row 19053: Score 100.0
9. Row 19054: Score 100.0
10. Row 19055: Score 100.0

==
End of Report

4. Appendix

The appendix contains supplementary material that is important for reference but doesn't fit in the main flow of the report.

4.1 Full requirements.txt

For reproducibility, this file lists the exact versions of all Python libraries used in the project. This allows anyone to create an identical environment by running pip install -r requirements.txt

```
numpy==1.26.4
pandas==2.2.3
scikit-learn==1.5.2
tensorflow==2.17.0
plotly==5.24.1
streamlit==1.38.0
python-dotenv==1.0.1
twilio==9.7.1
```

4.2 **Configuration Parameters**

This table lists the user-configurable parameters available in the application's UI, along with their default values and a brief description.

Parameter	Default Value	Algorithm	Description
Expected Anomaly Rate	0.1	Isolation Forest	The estimated proportion of anomalies in the dataset.
Number of Trees	100	Isolation Forest	The number of random trees to build in the forest.
PCA Components	5	PCA-based	The number of principal components to use for the model.
Encoding Dimension	10	Autoencoder	The size of the compressed data representation in the neural network.
Training Epochs	50	Autoencoder	The number of times the neural network will be trained on the data.

THANK YOU