**1. Analytics Dashboard – Data Visibility**

1. **Scan History Table**
   * **Fields per Scan:**
     + **Scan ID**: Unique identifier for each scan
     + **User ID**: Which user performed the scan
     + **Timestamp**: When the scan took place
     + **Features**: (Optional in a summarized form) e.g., short listing or count of significant features
     + **Autoencoder Error**: Reconstruction error from the autoencoder
     + **Binary Prediction**: “Benign” or “Malware”
     + **Malware Type**: (If identified) e.g., “Trojan,” “Worm,” etc.
     + **Risk Rating**: A numeric rating (0–10) capturing overall severity
   * **Purpose**: The admin can see a chronological list of all scans with key data points to identify trends and anomalies.
2. **Log of Actions (System & Admin)**
   * **Fields per Log Entry:**
     + **Action**: e.g., “retrain triggered,” “new dataset uploaded,” “parameter updated”
     + **Admin ID**: Which admin performed the action
     + **Timestamp**: When the action occurred
     + **Description**: Additional details, e.g., old vs. new parameter values
   * **Purpose**: Provides an audit trail of all administrative and system events.
3. **Aggregated Charts & Graphs**
   * **Threat Distribution**: Pie/Bar chart of “Benign vs. Malware” ratio or “Malware Types” distribution
   * **Risk Score Over Time**: A line chart showing average risk rating across scans or per user
   * **Scan Volume**: Graph of daily/weekly scan counts
   * **Autoencoder Error Distribution**: A histogram or timeline of reconstruction errors to spot unusual spikes
   * **Model Version History**: A small chart or table showing when re-training occurred and any major parameter changes
4. **Dataset Summary**
   * **Current Dataset Version**: e.g., “v2.1 after 2025-03-12 update”
   * **Number of Samples**: e.g., total benign/malware samples
   * **Malware Type Breakdown**: Count of each malware subtype in the training set
   * **Class Imbalance**: e.g., ratio of benign to malware to help the admin see if additional data is needed

**2. Parameters the Admin Can Update**

**A. Autoencoder Settings**

1. **Epochs**
   * Default: 100
   * Controls how many training passes the autoencoder does on benign data
2. **Batch Size**
   * Default: 32
   * Number of samples processed before the model updates its weights
3. **Learning Rate**
   * Default: 0.001
   * Adam optimizer parameter controlling the step size of gradient updates
4. **Threshold (mean\_error + k × std)**
   * Default: ~0.05 (k usually 3)
   * Boundary above which the autoencoder reconstruction error is considered “anomalous”

**B. Random Forest Hyperparameters**

1. **n\_estimators** (Number of Trees)
   * Default: 100
2. **max\_depth**
   * Default: None (trees grow fully)
3. **criterion**
   * Default: "gini" (admin could switch to "entropy" if desired)
4. **Random Seed**
   * Default: 42 (ensures reproducibility)

**C. Risk Weight (α\alphaα)**

* Default: 0.5
* Balances current anomaly error vs. historical average error
* Adjusting α\alphaα can make the system more/less sensitive to recent scans vs. long-term behavior

**D. Training Pipeline Config**

1. **Data Split Ratio**
   * Default: 70% train / 30% test
2. **Outlier Capping**
   * Default: 99th percentile
3. **Imputation Strategy**
   * e.g., “Drop rows” vs. “Mean fill” if there are missing values

**3. Practical Use-Cases**

1. **Re-Tuning for High False Positives:**
   * Admin sees a spike in “malware” predictions that are mostly false positives. They might lower the autoencoder threshold or adjust RF parameters to reduce over-detection.
2. **New Malware Samples:**
   * The dataset changes significantly. Admin merges new CSV data, re-runs data preprocessing, and triggers a re-training job to reflect the updated distribution.
3. **High Risk Scores Over Time:**
   * If the average risk rating is trending high, the admin might see from the logs if it’s because the autoencoder threshold is too low or the distribution of data changed. They can tweak α\alphaα or retrain the model with more examples.
4. **Version Control:**
   * Each time the admin hits “Re-train,” the system logs the new model version (with updated hyperparameters). The “Model Config” table helps track exactly what changed.

**Summary**

* **Analytics Dashboard (Admin View)**:
  + **Data**:
    - List of scans (user ID, anomaly error, predictions, risk rating)
    - Summaries & charts (threat types, daily scan volume, risk over time)
    - Logs of all admin/system actions
  + **Parameters**:
    - Autoencoder: epochs, batch\_size, learning\_rate, threshold
    - Random Forest: n\_estimators, max\_depth, criterion
    - Risk Weight α\alphaα
    - Additional data pipeline configs (split ratio, outlier handling, etc.)