Anomaly Detection Workflow with Python
A Step-by-Step Guide Using Machine Learning and Automation Tools
Date: March 24, 2025

1. Importing Libraries

- pandas | Data manipulation (e.g., CSV, DataFrames)
- numpy | Numerical operations (arrays, matrices)
- sklearn.ensemble.IsolationForest | Anomaly detection using Isolation Forest
- sklearn.preprocessing | Encoding categorical variables, scaling data
- smtplib | Sending emails
- email.mime.text/multipart | Structuring and formatting emails
- logging | Logging actions/errors for debugging
- requests | HTTP requests for external APIs
- openai | Leveraging OpenAI's language models
- jira | Interacting with Jira API
- os, sys | System path and error handling

2. Logging Setup

- logging.basicConfig(level=logging.INFO) | Sets up logging to track INFO-level messages

3. API and Configuration Setup

- OpenAI API Key | openai.api_key = "your-openai-api-key"
- Jira Setup | jira_url, jira_user, jira_token, JIRA instance
- Agentic AI API Key | AGENTIC_API_KEY = "your-agentic-api-key"

4. Data Loading and Preprocessing

- File Check | Verifies file existence with os.path.exists
- Load Data | pd.read_csv with categorical dtype
- Column Validation | Ensures required columns exist
- Label Encoding | Encodes categorical columns (Company, Account, etc.)
- Date Conversion | Converts Asofdate to datetime

5. Feature Engineering

- Data Cleaning | Converts numeric columns, fills NaN with 0
- Sorting/Grouping | Sorts by Asofdate, groups by Company, Account, AU
- Feature Creation | Days_Since_Last, GL_Change, IHub_Std, etc.
- Features for Model | Returns dataframe and feature list

6. Contamination Estimation

- Z-Score Calculation | Estimates outliers using mean and std
- Contamination Level | Caps between 0.01 and 0.5

7. Anomaly Detection (Isolation Forest)

- Data Scaling | Uses RobustScaler to handle outliers
- Model Training | IsolationForest with estimated contamination
- Anomaly Labeling | Assigns anomaly scores and labels (-1 as 1)

8. Generating Anomaly Explanations Using OpenAI
- Prompt Creation | Formats anomaly data into a detailed prompt
- OpenAI Call | Uses text-davinci-003 for explanation
- Response Handling | Extracts and returns explanation text

9. Creating Tasks in Agentic AI and JIRA
- Agentic AI | Task Creation | POST request with task details
- JIRA | Ticket Creation | Creates issue with anomaly summary

10. Sending Email Notification
- Email Setup | Configures sender and checks for anomalies
- Formatted Email | Converts anomalies to HTML format
- SMTP Sending | Uses Gmail SMTP to send email

11. Feedback Loop and Model Retraining
- Feedback Processing | Adjusts model based on user feedback
- Retraining | Updates contamination and retrains model

12. Main Execution
- Data Loading | Loads data and feedback files
- Feature Engineering | Prepares features for model
- Model Training | Trains Isolation Forest
- Feedback Loop | Retrains if feedback provided
- Email Notification | Sends anomaly report

13. Execution Example
- File Paths | Specifies paths to data and feedback files
- Main Call | Executes main function

14. Summary of Key Components
- Data Loading | Loads and preprocesses data
- Feature Engineering | Creates features for anomaly detection
- Anomaly Detection | Uses Isolation Forest

- LLM Explanations | Generates explanations with OpenAI
- Task Management | Creates tasks in Agentic AI and JIRA
- Email Notifications | Sends anomaly reports
- Feedback Loop | Improves model with feedback