Summary Report for Task-1

Tagging Approach:

The data tagging process was implemented through a comprehensive, systematic methodology focusing on extracting meaningful information from free-text fields. The approach leveraged keyword-based matching and logical reasoning to categorize complex engineering complaint data.

For Root Cause Tagging:

- Developed a keyword-based dictionary capturing common root cause categories

- Analyzed both complaint and cause text for identifying underlying issues

- Categorized root causes like "Not Tightened", "Not Installed", "Loose", "Leaking", and "Faulty"

- Used flexible matching to handle variations in textual descriptions

Symptom Component Identification:

- Created a comprehensive keyword mapping for identifying specific machine components

- Matched keywords across complaint and cause texts

- Captured components such as Cab P Clip, Fuel Door, Compressor Line, Harness, Sensors, and Connectors

- Implemented multi-level tagging to accommodate complex scenarios

Fix Condition Extraction:

- Analyzed correction text to understand repair methodologies

- Identified fix conditions like "Retightened", "Installed", "Replaced", and "Repaired"

- Matched keywords to categorize specific repair actions

- Ensured flexibility in interpreting repair descriptions

Potential Insights from Large-Scale Dataset (10,000 Rows):

With a larger dataset of 10,000 rows, we could generate more robust insights:

1. Predictive Maintenance Patterns

- Identify recurring failure modes

- Develop early warning systems for specific component vulnerabilities

- Create proactive maintenance schedules based on historical data

2. Quality Control Analysis

- Quantify manufacturing defects by frequency and type

- Pinpoint systemic issues in production processes

- Analyze correlation between root causes and specific product categories

3. Component Reliability Assessment

- Calculate mean-time-between-failures for different components

- Develop component reliability indices

- Support engineering design improvements

4. Geographical and Temporal Trend Analysis

- Examine failure patterns across different regions

- Identify seasonal variations in machine performance

- Correlate environmental factors with maintenance requirements

5. Cost-Benefit Analysis

- Estimate potential savings from preemptive maintenance
- Calculate return on investment for quality improvement initiatives
- Develop data-driven resource allocation strategies

Limitations and Recommendations:

- Enhance tagging accuracy through machine learning techniques
- Develop more sophisticated natural language processing models
- Continuously refine keyword dictionaries

The proposed methodology demonstrates a structured approach to transforming unstructured text data into actionable insights, providing a foundation for advanced engineering analytics.