

Tagging Approach and Insights

Approach to Tagging Each Field

To ensure accurate classification of the complaint and resolution data, we used a structured NLP-driven approach combined with fuzzy matching techniques. Below is a breakdown of the tagging methodology for each field:

1. Root Cause

- Extracted from the "Cause" column using **text normalization** (lowercasing, removing special characters) and **fuzzy matching** against a predefined taxonomy.
- If no exact match was found, the closest category with a confidence threshold was assigned.
- Example: "Not tighten at factory" was mapped to "Not Tightened."

2. Symptom Condition

- Derived from the "Complaint" column using **keyword extraction**.
- Frequent complaint patterns were identified and mapped to known symptom conditions.
- Example: "Oil running from bottom of machine" → "Oil Leak."

3. Symptom Component

- Identified by **entity recognition** to locate the specific machine part mentioned in the complaint.
- Cross-referenced with a taxonomy of components for accuracy.
- Example: "Fuel door will not stay open" → "Fuel Door."

4. Fix Condition

- Extracted from the "Correction" field using **action-based text parsing**.
- Mapped common fix actions like "Installed," "Replaced," and "Tightened."
- Example: "Installed missing unlocks" → "Installed."

5. Fix Component

- Identified using **fuzzy string matching** to detect which component was modified or replaced.
- Example: "Installed missing unlocks" → "Unlocks."

Potential Insights Generated

- **Frequent Root Causes:** By analysing trends in the "Root Cause" field, we can identify systemic issues in manufacturing or assembly (e.g., "Not Tightened" might indicate quality control gaps).
- **Recurring Symptom Patterns:** Identifying common failures (e.g., "Oil Leak" or "Component Not Installed") helps prioritize design improvements.
- **Effectiveness of Fixes:** Analysing successful resolutions over time can refine troubleshooting procedures and reduce downtime.
- **Predictive Maintenance:** By correlating symptom conditions with root causes, we can predict potential failures before they happen, allowing for proactive interventions.
- **Component Failure Trends:** Understanding which components fail most frequently helps optimize spare parts inventory and service scheduling.

Evaluation Criteria for Task 1

To ensure effective data tagging and structured categorization, the following evaluation criteria are considered:

- **Tagging Accuracy:** A minimum of **50% tagging accuracy** is expected. This ensures that the assigned labels effectively represent the nature of the complaints and resolutions, allowing for meaningful analysis and actionable insights.
- **Critical Thinking:** The ability to handle ambiguous cases is essential. In cases where multiple classifications are possible, logical reasoning and documentation of observations must be applied to ensure consistency and traceability in tagging.

Conclusion:

This data-driven tagging approach allows for structured analysis, leading to better decision-making in maintenance, quality assurance, and product design improvements.