

# Project Deliverable 4 - Data Labeling

**Project Title:** Automated JIRA Issue Creation and Classification from Meeting Transcriptions with AI/ML

**Team Name:** TabComp

**Members:** Pallavi Bichpuriya, Hitakshi Shirude, Harsh Avinash Kute, Kunal Shah, Niwant Salunke

**Course:** ITIS 6880 / ITCS 6010 - Software Engineering for AI-Enabled Systems

**Semester:** Fall 2025

## 1 - Understand Your Data and Task (20 pts)

### (a) Target Label

For our JIRA-derived dataset, the target column is **issueType**, which classifies each issue as *Bug*, *Task*, *Story*, or *Epic*.

For the MOM-Summary dataset, the goal is similar to detect actionable items or sentiments within meeting transcripts.

### (b) Type of Data

Textual data (NLP domain). Each record contains a short issue *summary* and longer *description* fields from JIRA tickets or meeting transcripts.

### (c) Goal of Labeling

To manually label the first 50 rows so the model can later learn to classify text into appropriate JIRA issue types or priority levels based on language and context.

### (d) Kind of Labels Needed

Categorical labels.

For simplicity in this deliverable, we used: **BUG**, **SUGGESTION**, and **OTHER**, representing defects, feature requests/improvements, and non-actionable statements respectively.

## 2 - Define Clear Labeling Guidelines (40 pts)

### (a) Labeling Guide

Label	When to Assign	Example Text
<b>BUG</b>	Mentions an error, failure, or fix needed (technical fault)	"Login fails after password reset → NullPointerException in API."
<b>SUGGESTION</b>	Requests a feature or improvement without bug context	"Add dark mode support to dashboard for better accessibility."
<b>OTHER</b>	Not actionable / general comment / status / organizational	"Discussed deployment timeline in next meeting."

### **Edge cases handled:**

- If both a bug and suggestion are mentioned, prioritize **BUG** (since fix is immediate action).
- Meta discussion or management notes fall under **OTHER**.

### **(b) Manual Labeling of 50 Rows**

We removed the existing label column from the first 50 records and manually labeled them following the above guide.

Each teammate labeled the same 50 records, and our labels were aggregated in FleissInput within Data\_Labeling.ipynb.

#### **Sample Labeled Rows**

<b>Text Excerpt</b>	<b>Label</b>	<b>Explanation</b>
"UI crashes when adding a new user."	BUG	Explicit failure requiring a fix.
"Add CSV export option for project reports."	SUGGESTION	Improvement request (no error described).
"Reviewed test plan in today's stand-up."	OTHER	Organizational note - not an issue type.

The labeled subset was saved as **Project-Del-4-TABCOMP-LabeledData.csv**.

### **3 - Resolve Labeling Disagreements (40 pts)**

#### **(a) Data Preparation**

Rows = 50 items; Columns = BUG, SUGGESTION, OTHER.

Each cell contains the number of raters who assigned that category to that item.

Example (first 5 rows):

<b>row_id</b>	<b>BUG</b>	<b>SUGGESTION</b>	<b>OTHER</b>
1	5	0	0
2	5	0	0
3	0	4	1
4	0	3	2
5	0	5	0

## (b) - (d) Fleiss' Kappa Calculation

The formula for Fleiss' Kappa is:

$$\kappa = (P^a - P^e) / (1 - P^e)$$

Where:

- $\kappa$  is Fleiss' Kappa
- $P^a$  is the average proportion of agreement among all pairs of raters
- $P^e$  is the average proportion of agreement expected by chance

For each item  $i$ , we calculate the proportion of agreement among all possible pairs of raters

$$P^a_i = (1 / (n * (n - 1))) * \sum (k_{ij} * (k_{ij} - 1))$$

Where:

- $n$  is the number of raters
- $k_{ij}$  is the number of raters who assigned category  $j$  to item  $i$
- The summation is over all categories  $j$

## Notebook Output Snapshot:

Fleiss' Kappa

Iteration 1:

$$K = (P^a - P^e) / (1 - P^e)$$

We have 5 team members  $n=5$   
Customer reviews  $N=50$

For  $P^a_i$ :

$$P^a_i = \frac{1}{5(4)} \times \sum (k_{ij} * (k_{ij} - 1))$$

$P_1 = 1$     $(5(4) - 12(0)) = 1$

$P_2 = 1$

$P_3 = 0.6$

$P_4 = 0.4$

$P_5 = 1$

$\bar{P} = 0.784$

For  $P^e$

$$P_j = \frac{1}{5 \times 50} \times \sum k_{ij}$$

$P_{\text{positive}} = 0.48$   
 $P_{\text{negative}} = 0.32$   
 $P_{\text{neutral}} = 0.2$

$$\bar{P}^e = 0.373 \quad [\sum P_j^2]$$

$$K = (0.784 - 0.373) / (1 - 0.373)$$

$$K = 0.656$$

Interpretation = Substantial agreement.

Iteration 2:

$$K = (\bar{P} - \bar{P}_e) / (1 - \bar{P}_e)$$

$$n = 5, N = 50$$

$$\text{For } \bar{P}_j = \frac{1}{5(5-1)} \times \sum (k_{ij} / (k_{ij}-1))$$

$$P_1 = 1$$

$$P_2 = 1$$

$$P_3 = 0.6$$

$$P_4 = 0.4$$

$$P_5 = 1$$

$$\bar{P} = \frac{1}{N} \times \sum \bar{P}_j = 0.892$$

For  $P_e$

$$P_j = \frac{1}{5 \times 50} \times \sum k_{ij}$$

$$P_{\text{positive}} = 0.528$$

$$P_{\text{negative}} = 0.276$$

$$P_{\text{neutral}} = 0.196$$

$$\bar{P}_e = \sum (P_j)^2$$

$$= (0.528)^2 + (0.276)^2 + (0.196)^2$$

$$= 0.393$$

$$K = (0.892 - 0.393) / (1 - 0.393)$$

$$\boxed{K = 0.822}$$

Interpretation - Almost perfect.



Fleiss' Kappa: 0.822

$P_i$  (first 5): [1. 1. 0.6 0.4 1. ]

$$\bar{P} = 0.892$$

$$p_j = [0.528 0.276 0.196]$$

$$\bar{P}_e = 0.393$$

$$K = 0.822$$

Interpretation: Almost perfect

### Interpretation

A  $\kappa \approx 0.822$  indicates **almost perfect** among raters, showing that the labeling guidelines were clear and consistently applied across BUG vs SUGGESTION vs OTHER cases.

### 4 - Member Contribution (2.5 pts)

Member	Contribution Description	Overall Contribution%

Pallavi Bichpuriya	Manual labeling of first 50 items, Led label guideline design, built Excel FleissInput table, computed $\kappa$ in Data_Labeling.ipynb	100
Hitakshi Shirude	Manual labeling of first 50 items, review of guidelines	100
Harsh Avinash Kute	Manual labeling of first 50 items, validation of counts table	100
Kunal Shah	Manual labeling of first 50 items, Helped verify data alignment across raters	100
Niwant Salunke	Manual labeling of first 50 items, pairwise agreement check	100