AI-Powered Knowledge Engine for
Smart Support & Ticket Resolution

Assignment -
1. Data collection (Tickets)
2. Tagging and Categorisation of Tickets
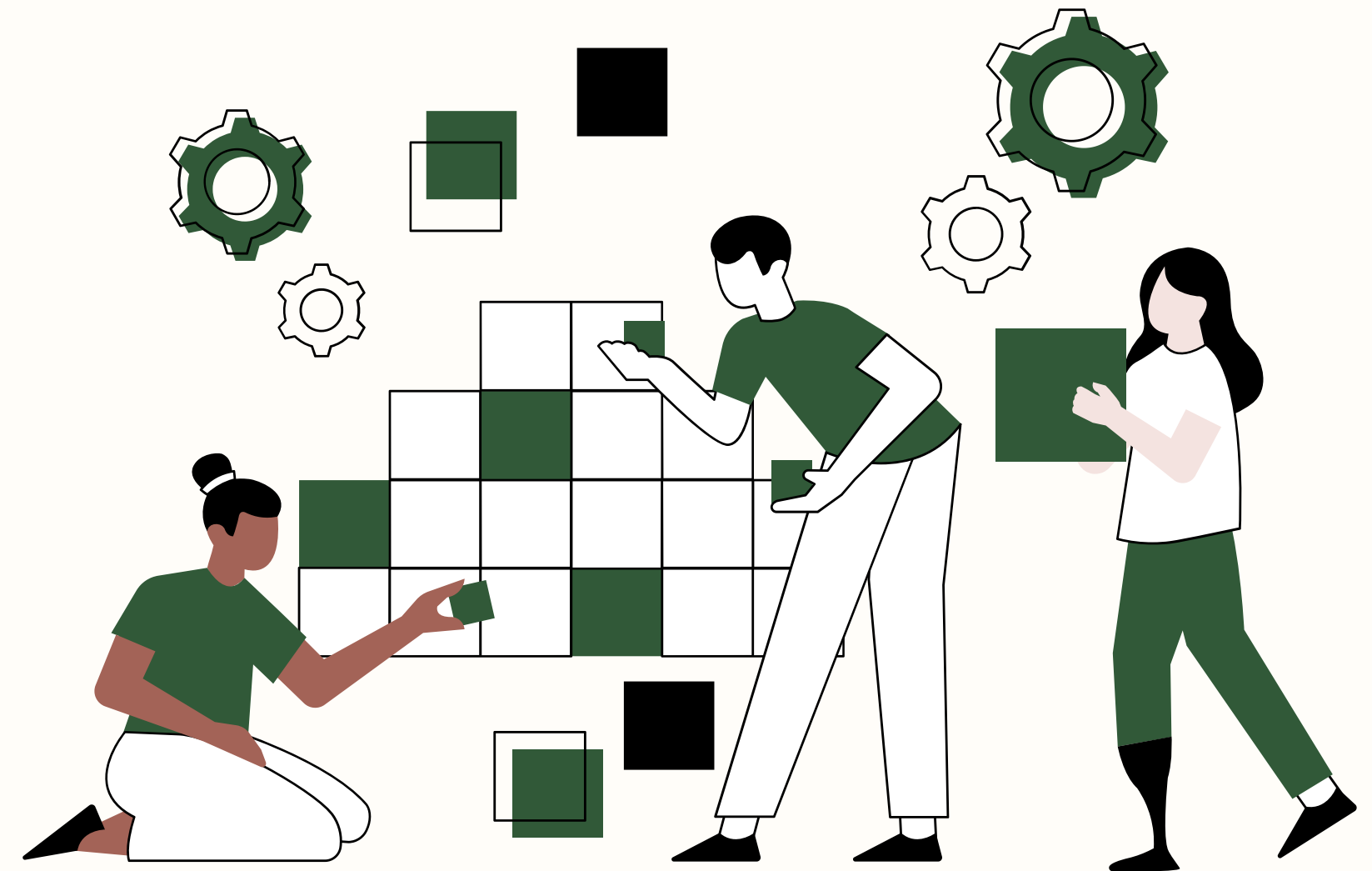(High / Medium / Low) priority level.
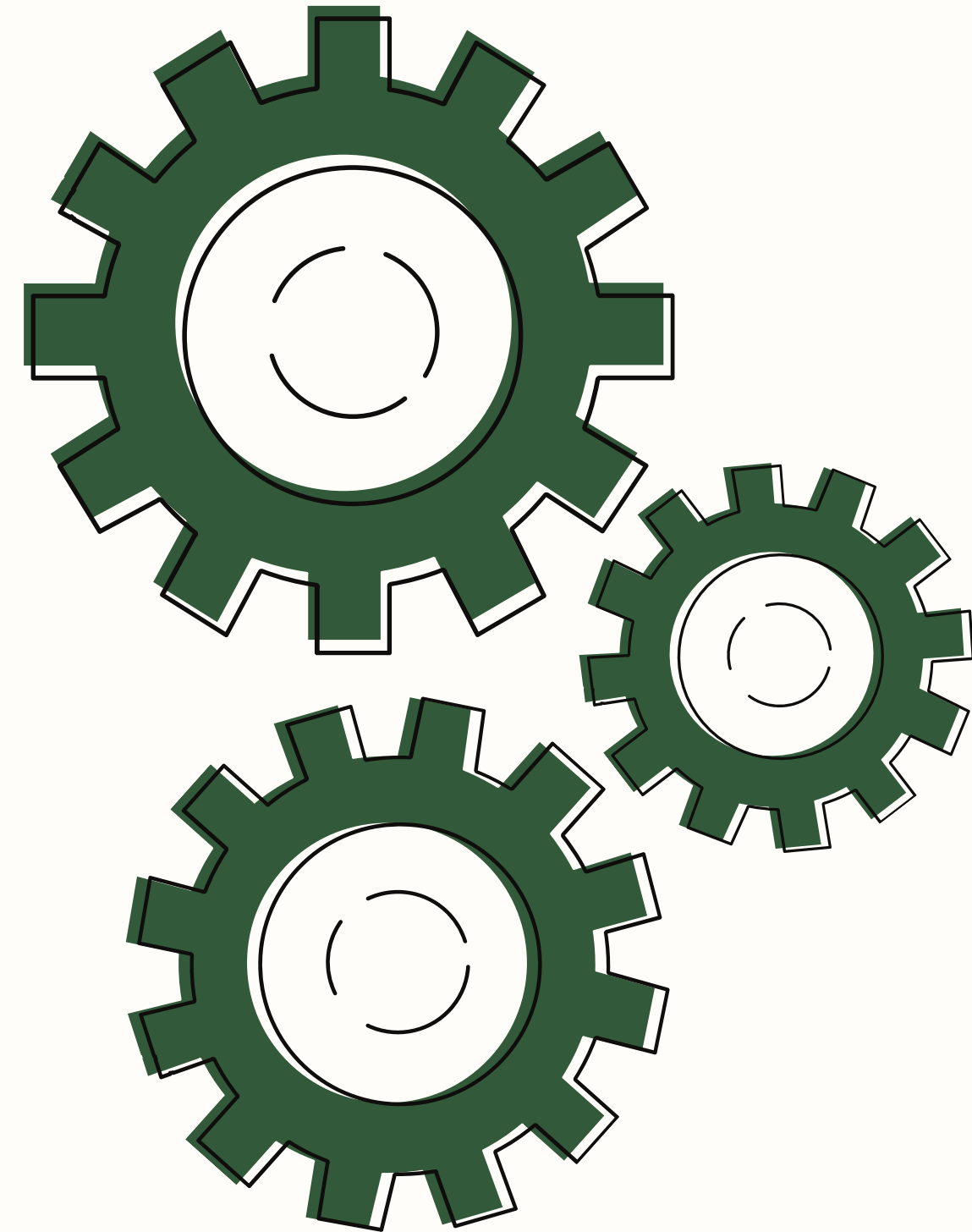
Group 2 -
Vaishnavi
Anirudh
Manikanta
Rajnandani

Objective

- Experiment with sentence embeddings for customer support tickets.

- Analyze ticket messages to generate semantic vector representations.

- Categorize tickets into High, Medium, Low priority.

- Foundation for Knowledge Categorization module in the AI Knowledge Engine.

Why This Dataset?

Dataset Used: Tobi-Bueck/customer-support-tickets (Hugging Face)

Reasons for Choosing This Dataset:

- Real-world relevance: Contains authentic customer support tickets.
- Rich text data: Includes subjects, messages, and categories, perfect for embedding experiments.
- Structured yet diverse: Allows testing of both text processing and priority categorization.
- Open-source & accessible: Easy to load and integrate in Colab for prototyping.

Model Used

Model: all-MiniLM-L6-v2 from Sentence-Transformers

Purpose:

- Convert text tickets into 384-dimensional embeddings.
- Capture semantic meaning for similarity and categorization.
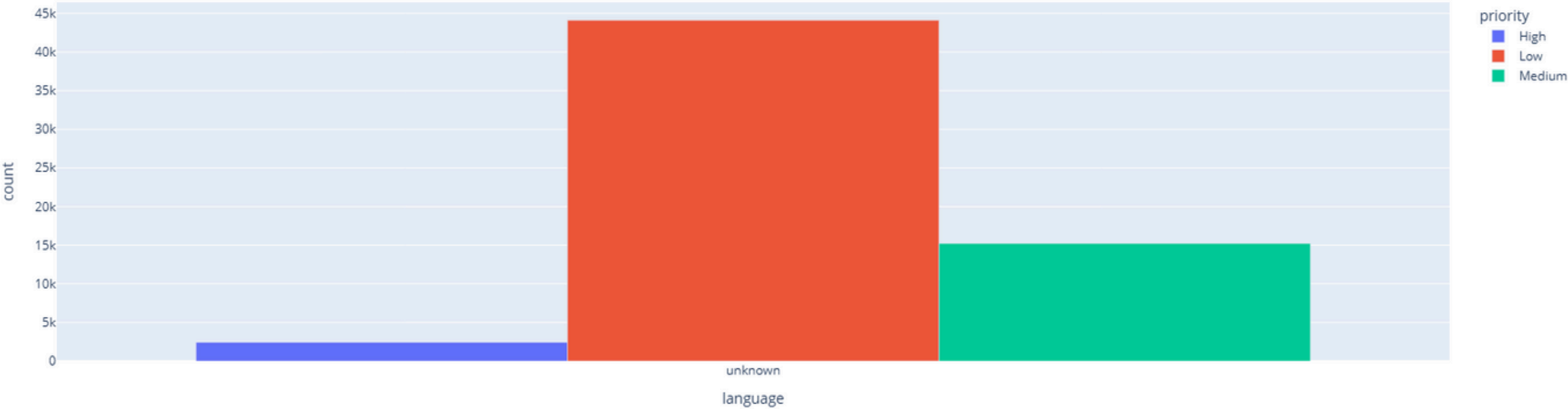
Platform: Google Colab

Implementation Steps

1. Loaded dataset via Hugging Face datasets library.

2. Converted ticket messages to embeddings using SentenceTransformer.

3. Applied rule-based priority tagging:

   - High: error, fail, crash, urgent
   - Medium: delay, slow, issue, confusion
   - Low: general/informational queries

4. Applied language detection using FastText (lid.176.bin):

   - Detected language for each ticket message.
   - Categorized tickets by language for analytics & multilingual support.

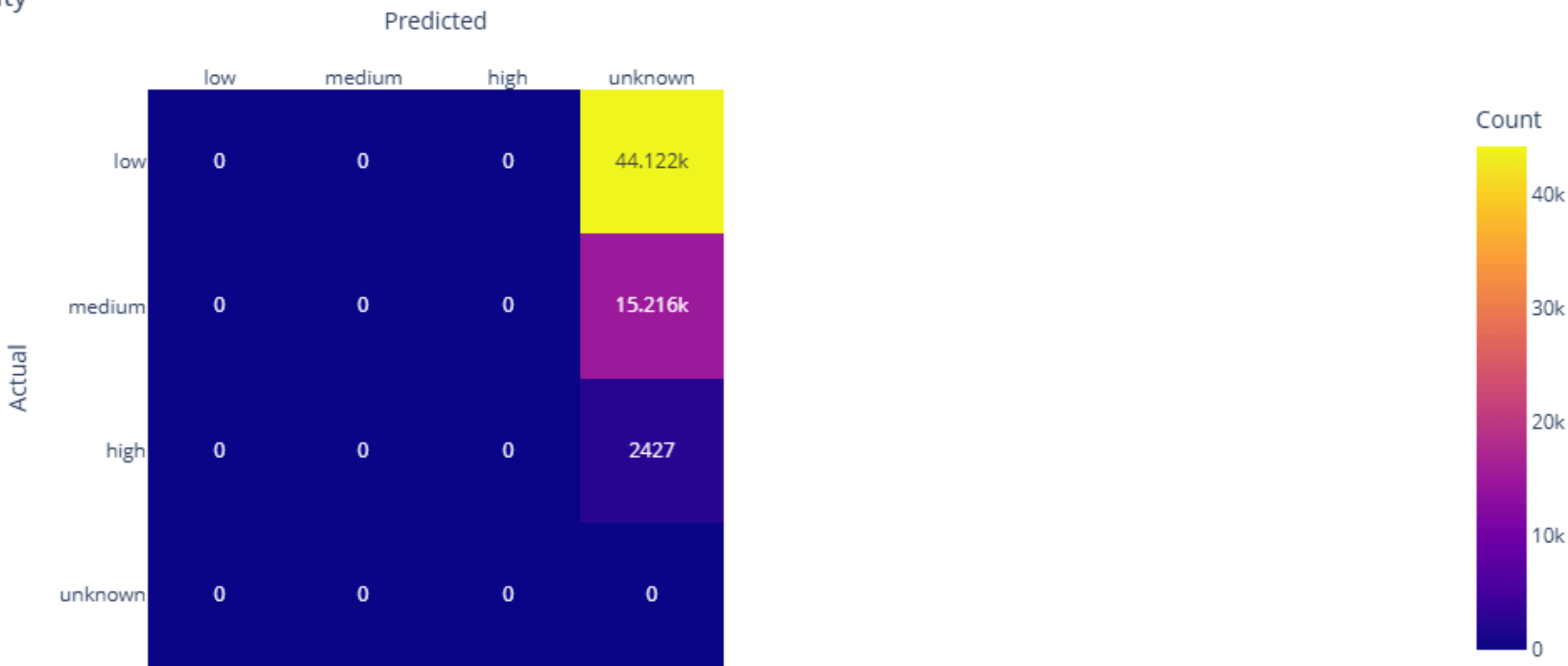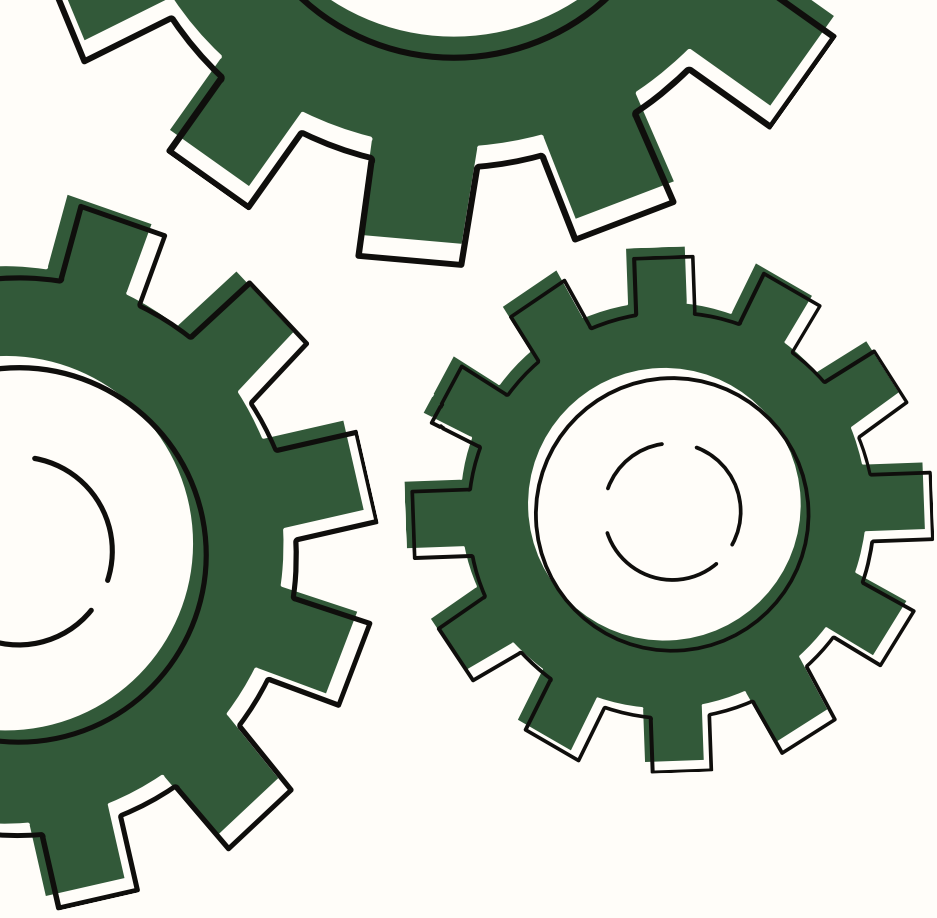5. Visualized ticket priority distribution and language distribution using Matplotlib.

Visualization / Results

Thank you!