Text Sentiment Annotation Workflow

Simulated Real-World Data Annotation Project

---

1. Project Overview

This project simulates a real-world text annotation workflow used in AI/ML data preparation pipelines.

The objective of this project was to:

- Design structured annotation guidelines
- Manually label raw text data
- Measure annotation consistency
- Evaluate inter-annotator agreement
- Train a basic machine learning model
- Observe the impact of labeled data on model performance

This project demonstrates understanding of structured annotation processes and quality validation methods commonly used in data annotation roles.

---

2. Dataset Description

- Total Samples: 51 text reviews
- Domain: Product and service feedback
- Type: Short textual sentiment data
- Labels: Positive, Negative, Neutral

Class Distribution:

- Positive: 22
- Negative: 18
- Neutral: 11

The dataset was intentionally kept small to simulate a controlled annotation workflow.

---

3. Annotation Guidelines

Structured guidelines were created before labeling to ensure consistency and reduce ambiguity.

Labels Defined:

- **Positive** → Clearly expresses satisfaction or approval
- **Negative** → Clearly expresses dissatisfaction or criticism
- **Neutral** → Factual, mixed, or emotionally balanced statement

Annotation Rules:

1. Label based on overall tone.
2. If both positive and negative sentiment appear, assign the dominant tone.
3. Do not assume meaning beyond written content.
4. Maintain consistency across similar examples.
5. Avoid personal bias.

These guidelines simulate real-world annotation documentation used in data labeling projects.

---

4. Annotation Process

- Raw reviews were manually labeled according to the defined guidelines.
- A second annotator column was simulated to measure agreement.
- Minor disagreements were introduced intentionally to test consistency metrics.

This mimics a real annotation environment where multiple annotators work on the same dataset.

---

5. Quality Evaluation – Inter-Annotator Agreement

Cohen's Kappa Score was calculated to measure annotation reliability.

Result:

Cohen's Kappa Score: **0.88**

Interpretation:

A Kappa score between 0.81–1.00 indicates *almost perfect agreement*.

This suggests:

- The annotation guidelines were clear.
- Labeling was consistent.
- Minimal ambiguity in interpretation.

This metric is commonly used in annotation projects to ensure data quality before model training.

---

6. Machine Learning Evaluation

To demonstrate how labeled data impacts AI systems:

- Text data was vectorized using CountVectorizer.
- Logistic Regression classifier was trained.
- Model performance was evaluated using accuracy and classification report.

Results:

- Accuracy: 54%
- Moderate performance due to:
    - Small dataset size
    - Class imbalance (Neutral underrepresented)
    - Limited training samples

Insight:

The objective was not model optimization, but to demonstrate how annotated datasets feed into downstream ML pipelines.

---

7. Key Findings

1. Structured annotation guidelines significantly improve consistency.
2. Cohen’s Kappa is an effective reliability metric in annotation workflows.
3. Even small datasets can demonstrate the importance of labeled data in ML.
4. Annotation quality directly affects model learning outcomes.

---

8. Tools Used

- Python
- Pandas
- Scikit-learn
- Jupyter Notebook / Google Colab

---

9. Learning Outcome

Through this project, I gained practical understanding of:

- Designing annotation guidelines
- Manual text labeling
- Inter-annotator agreement measurement
- Quality validation processes
- The relationship between annotation and AI model performance

This project reflects foundational skills required in structured data annotation roles.
