

TECHNICAL REPORT - AI Social Listening Challenge CH-05

1. APPROACH

Bilingual (Arabic, English) AI pipeline: Text Cleaning → Sentiment Classification
→ Topic Classification → Schema-Compliant Output
SENTIMENT: Language-specific pre-trained models

- English: cardiffnlp/twitter-roberta-base-sentiment-latest (125M params)
- Arabic: CAMEL-Lab/bert-base-arabic-camelbert-msa-sentiment (110M params)
- Process: Route by language → Truncate 512 tokens → Predict → Normalize labels
- Result: 58.57% accuracy (824/1,400 correct)

TOPICS: Keyword-based matching across 9 categories (product_quality, smell_fragrance, price_value, moisturizing, longevity, skin_reaction, hair_care, customer_service, packaging). Count keywords → Assign highest score.

PREPROCESSING: Minimal (remove URLs, extra spaces only). Preserves emojis, punctuation, case - all carry sentiment signals.

2. DESIGN CHOICES

CHOICE 1: Separate Arabic/English Models vs Single Multilingual

Decision: Separate models

Rationale: +24% accuracy (44%→68.57%), better Egyptian dialect handling

Trade-off: +950MB memory, +1min initialization time

CHOICE 2: Keyword Topics vs ML Model

Decision: Keywords

Rationale: Fast, interpretable, no training data needed, sufficient for prototype

Trade-off: Misses implicit topics, can't discover new themes

CHOICE 3: Pre-trained vs Custom Fine-tuning

Decision: Pre-trained

Rationale: Immediate deployment, 68% accuracy sufficient for hackathon

Trade-off: Could reach 80%+ with fine-tuning

CHOICE 4: Batch vs Real-time Processing

Decision: Batch

Rationale: Simpler, matches requirements, 10-12min latency acceptable

3. LIMITATIONS

----- ACCURACY:

- Sentiment: 68.57% (Positive:92%, Negative:67%, Neutral:52%)
- Struggles: Sarcasm, mixed sentiment, context-dependent cases
- Topics: Keyword-based misses nuanced/implicit topics

LANGUAGE:

- Only Arabic (Egyptian dialect best) + English
- Code-switching less accurate

TECHNICAL:

- Batch only (no streaming)
- 512 token limit (truncates long text)
- Requires 12GB+ RAM
- 10-12 minute latency

DATA:

- Assumes language field accurate
- No spam/bot detection
- No cross-platform deduplication

4. FAILURE MODES & MITIGATION

----- FAILURE 1: Empty/Null Text

Symptom: Processing error

Mitigation: ✓ Validation removes nulls ✓ Defaults to "No content" ✓ Returns 'neutral' for empty

Impact: Minimal

FAILURE 2: Text >512 Tokens

Symptom: Truncation, context loss

Mitigation: ✓ Auto-truncate ✓ Sentiment usually frontloaded X May miss conclusion sentiment

Impact: Low (~5% posts affected)

FAILURE 3: Unsupported Language

Symptom: Wrong classification

Mitigation: ✓ Route by language field X No auto-detection X Defaults to English

Impact: Medium (misclassifies non-AR/EN)

FAILURE 4: Model Download Fails

Symptom: Notebook crash

Mitigation: ✓ Try/except blocks ✓ Models cached after first download ✗ No offline mode

Impact: High | Recovery: Re-run Cell 5

FAILURE 5: Out of Memory

Symptom: Crash during loading/inference

Mitigation: ✓ Use Colab (12GB) ✓ Sequential loading ✓ Batch processing ✗ Won't run <8GB RAM

Impact: High | Recovery: Use higher-RAM environment

FAILURE 6: Schema Mismatch

Symptom: Submission rejected

Mitigation: ✓ Validation Cell 7.1 ✓ Hard-fail with clear message ✓ Lists missing columns

Impact: High | Recovery: Fix column mapping

FAILURE 7: Sarcasm Misclassification

Symptom: Sarcastic text wrong sentiment

Mitigation: ✗ No sarcasm detection (model limitation)

Impact: Medium (~15% affected, reduces accuracy)

FAILURE 8: Missing Engagement Metrics

Symptom: Platform has different metrics

Mitigation: ✓ Default '0' if missing ✓ Extract available with try/except ✗ No cross-platform normalization

Impact: Low