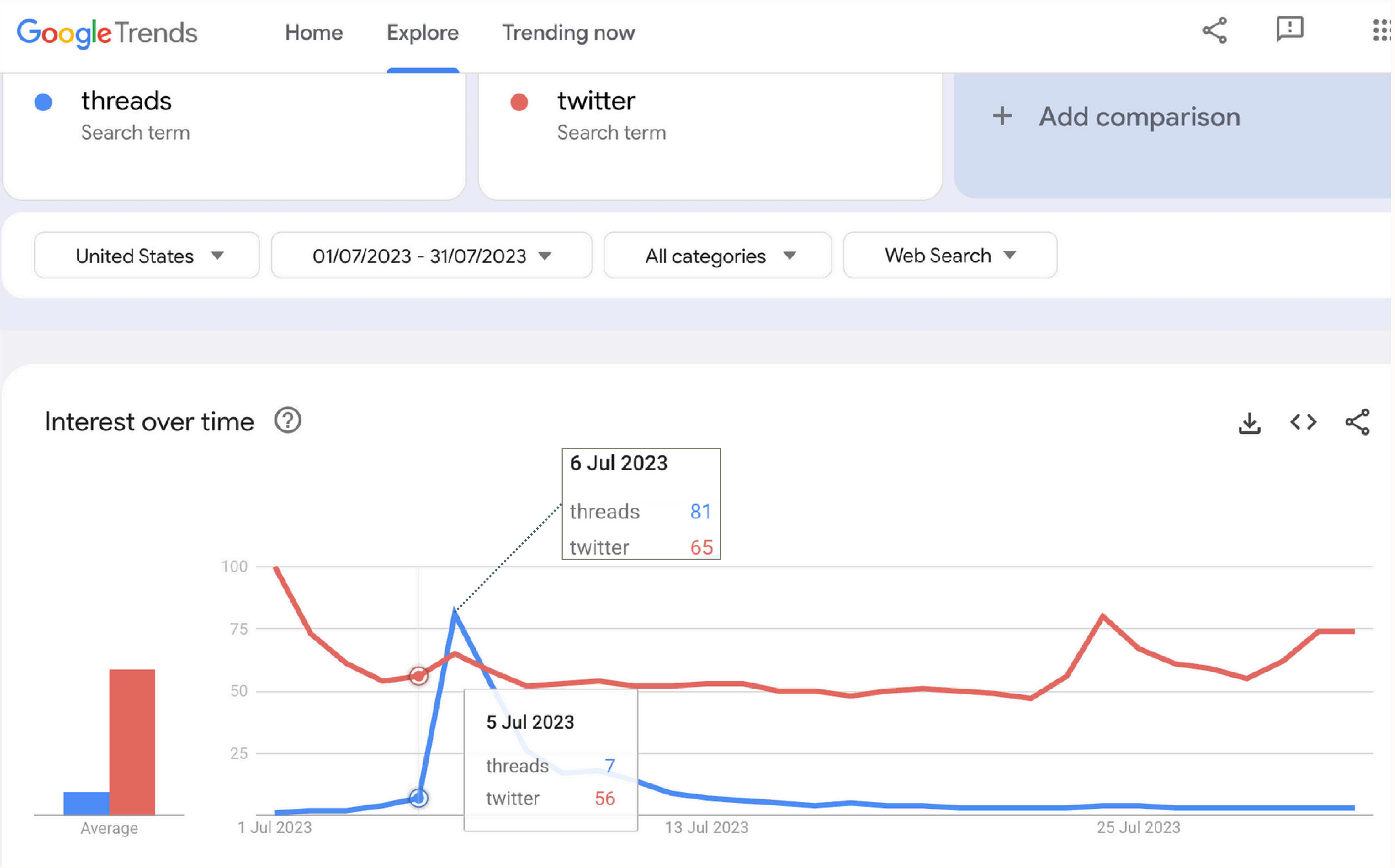
Comparative Analysis: Threads and Twitter Reviews

Natural Language Processing for Data Science_DATS_6312
Group Corpus Crew

Sai Rachana Kandikattu, Snehitha Tadapaneni, Haeyeon Jeong

List of Contents



https://trends.google.com/trends/explore?date=2023-07-01%202023-07-31&geo=US&q=threads,twitter&hl=en-GB

Nov 2022

- Elon Musk acquires Twitter

Jan–May 2023

- Major layoffs at Twitter (50–80% cuts)

June 2023

- Backend + API restructuring
- System instability & glitches

July 2023

- **Threads launches (July 5)**
- **Twitter rebrands to X**
- **Post-viewing limits introduced**

Research Goal

Goal
Understand how users perceived both platforms using NLP pipeline

Research Questions

- Q1: How does user sentiment differ between Threads and Twitter?
- Q2: How reliable are VADER pseudo-labels vs human annotations?
- Q3: How do TF-IDF + Logistic Regression and DistilBERT perform?
- Q4: How do LDA, NMF, and BERTopic differ in topic coherence and insights?

Dataset Overview

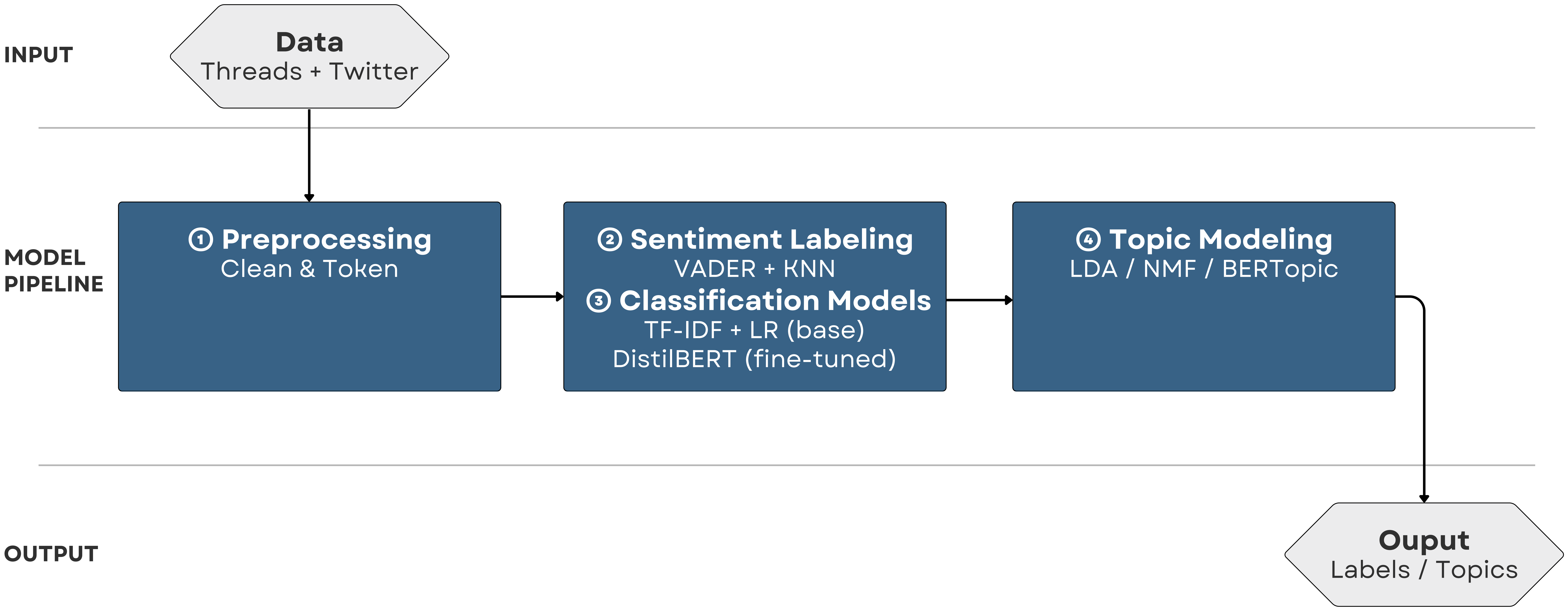
Threads	
Rows	32,910
Filelds	source
	review_description
	rating
	review_date

Twitter	
Rows	34,788
Filelds	review_id
	pseudo_author_id
	author_name
	review_text
	review_rating
	review_likes
	author_app_version
	reivew_time_stamp

Shared characteristics

- Soruce: Kaggle
- Data from July 2023
- Short, noisy text (misspellings, emoji)
- Many duplicates
- Multilingual content

Methodology



Data Cleaning Pipeline

01 Duplicate Removal

- Raw text duplicates
- Long-review duplicates (>15 words)

02 Brand Normalization

- "twt", "twit", "x" → “twitter”
- "treads", "threds" → “threads”

03 Text Normalization

- Lowercase
- Remove URLs, emojis, special characters
- Whitespace normalization

04 Stop-words & Lemmatization

- Remove English stopwords
- Retain negations (“not”, “never”)
- Lemmatize words

05 Minimal Language Filtering

- langdetect: only remove frequently misdetected languages
- Keep English + “unknown” safely

Final dataset sizes

- Threads: 29,646
- Twitter: 29,611

Sentiment Labeling (VADER + KNN)

Pseudo-label Generation

- VADER compound score ∈ [−1, +1]
- Fixed thresholds (+0.5/−0.5)
- Optimal Thresholds using KNN distribution analysis:
  Positive: ≥ +0.25
  Neutral: between −0.25 and +0.25
  Negative: ≤ −0.25

Manual Evaluation (~3,000 human-labeled samples)

- Threads: Accuracy 0.54
- Twitter: Accuracy 0.60

Sentiment Comparison Insights

01 Sentiment Distribution (VADER labels)

Sentiment	Threads	Twitter
Positive	~16k	~11k
Negative	~9k	~10k
Neutral	~4k	~9k

- Threads more positive overall
- Twitter more neutral → requesting help etc
- Negative similar, indicating comparable dissatisfaction

02 VADER vs Observed Insights

Threads

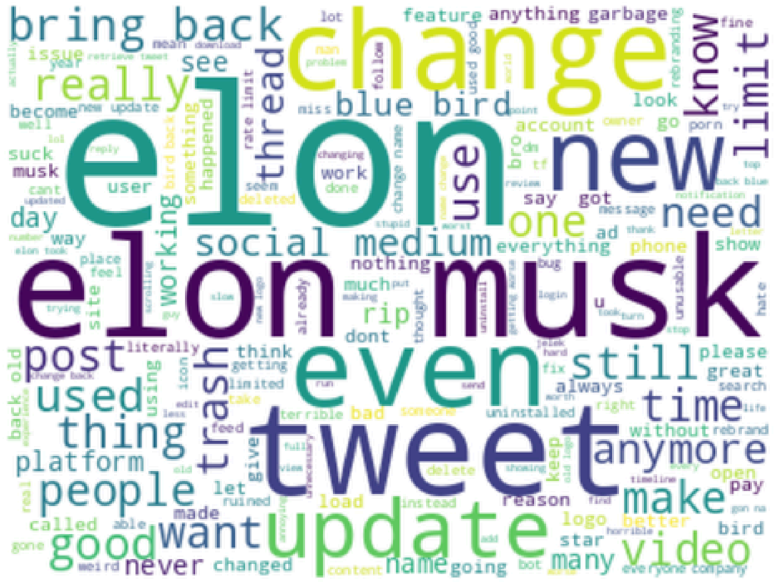
- Many reviews which were complaints about app glitches, bugs were classified as neutral due to absence of explicit negative words.

Twitter

- Threads reviews had explicit negative words which helped VADER classify the reviews better

Key Insights:
VADER Struggles with mixed reviews as there are both positive and negative words in the reviews.

Sentiment Analysis (Threads vs Twitter)

<table>
<tr><th></th><th>Positive</th><th>Neutral</th><th>Negative</th></tr>
<tr><td>Threads</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>Twitter</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
</table>

Threads: Frequent comparisons to Twitter appear in both positive and negative contexts.
Twitter: Elon Musk dominates the conversation, reflecting recent ownership changes.

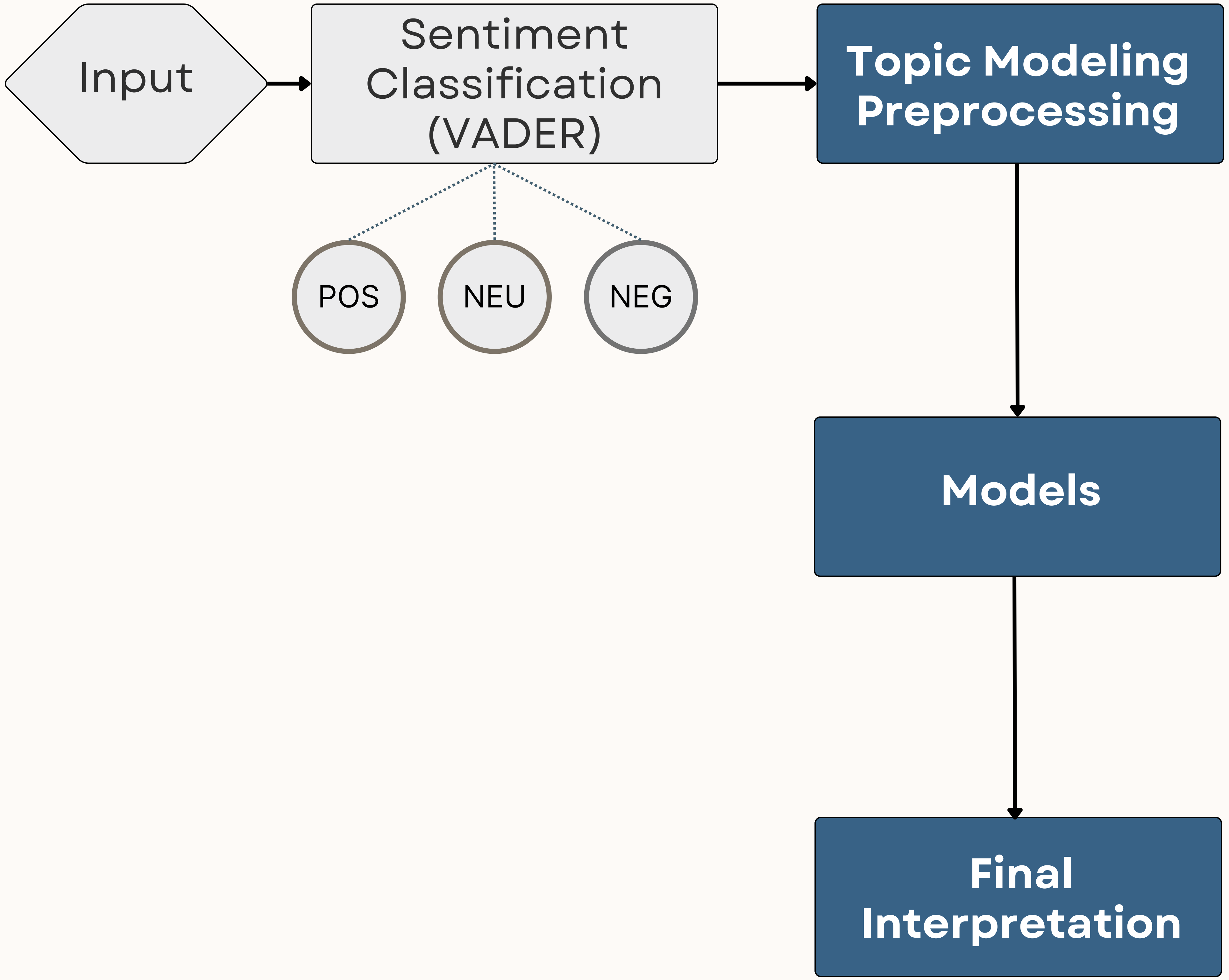
Sentiment Classification Models

Baseline: TF-IDF + Logistic Regression	Advanced: Fine-tuned DistilBERT
• Threads: Acc 0.74, F1 0.69 • Twitter: Acc 0.84, F1 0.83	• Fine-tuned on cleaned reviews using VADER labels. • Threads: Acc 0.86, F1 0.82 • Twitter: Acc 0.93, F1 0.92

Key insight:

- **DistilBERT outperforms classical models by capturing semantic, contextual, and positional information**
- **Better performance on Twitter compared to Threads due to less noisy-labels**

Topic Modeling Approach



Additional Preprocessing

- *Bigram detection*
- *Bag-of-Words/ TF-IDF / Embeddings (depending on model)*

- *LDA – probabilistic, bag-of-words baseline*
- *NMF – TF-IDF matrix factorization*
- *BERTopic – Sentence-BERT + UMAP + HDBSCAN + c-TF-IDF (Modern embedding based approach)*

Topic Model Comparison & Selection

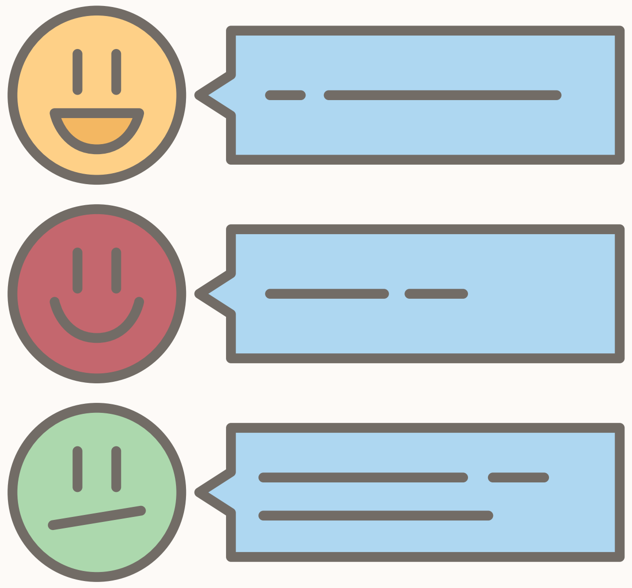
01 Topic Modeling Methods Comparison

Model	Threads	Twitter
LDA	~0.52	~0.50
NMF	~0.46	~0.43
BERTopic	0.53	0.52

(Coherence Score)

02 Best Topic Model Architecture: BERTopic

- Highest coherence on both datasets
- Robust for short, noisy social-media reviews
- Captures deeper semantic patterns



Key insight:

- **BERTopic captures the sharpest platform-specific themes.**

Topic Modeling: Best Model Architecture

01 Pipeline Architecture

1. Sentence-BERT Embeddings

– semantic representation of each review

2. UMAP Dimensionality Reduction

– preserves structure, compresses vectors into smaller space

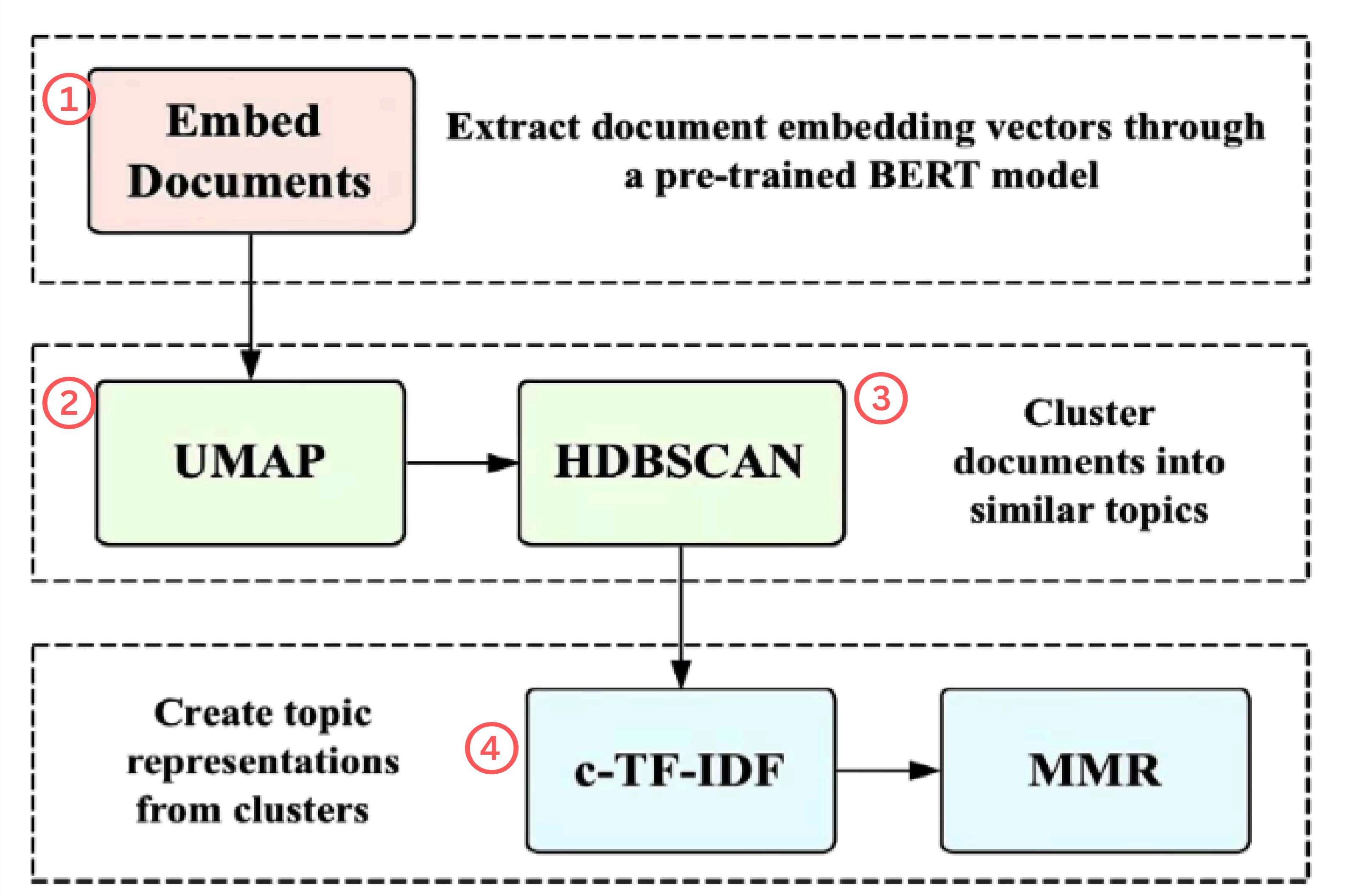
3. HDBSCAN Clustering

– finds dense groups of similar reviews without us specifying the number of clusters

4. c-TF-IDF Topic Extraction

– identifies the most representative words for each cluster

02 BERTopic Architecture



Topic Modeling Insights

Topic Modeling Analysis (BERTopic)

Threads (New-platform stability)	Twitter (Identity & Leadership)
• Positive → "good app", simple praise • Neutral → copy/clone discussions • Negative → crashes, missing features, bugs	• Positive → free speech, improvements • Neutral → logo/name changes • Negative → Musk decisions, rate limits, rebranding anger

Key insight:
Threads is criticized for what it lacks, while Twitter is criticized for what it has become.

Topic Modeling Insights: Threads

Threads (New-platform stability)

Positive Sentiment
General positive impressions - good, nice, best, awesome
Positive Twitter comparisons - twitter, better, app
Positive Experience - app, good, nice
Copy/clone comments (positive or sarcastic) - copy, copycat, original

Neutral Sentiment
Platform comparisons or suggestions - twitter, instagram, account
Functionality & bug mentions - app, crashing, glitching, bug
Leadership mentions - elon, musk, zuck
Copying discussions - copy, copied, clone

Negative Sentiment
Lacking aspects - app, account, not
Bad experience - bad, boring, useless
Leadership/platform direction critique - elon, musk, zuck, hate
Accusing Threads of cheating- cheating,competition, not ,fine

Topic Modeling Insights: Twitter

Twitter (Identity & leadership)

Positive Sentiment
How Amazing twitter is compared to X - app, good, best
Simple praise - good, nice, great, excellent
Free speech appreciation of X - speech, free, freedom, elon
Account help request - account, help, please, number

Neutral Sentiment
Rebranding/name-change remarks: twitter, app,
App quality over time - trash, used, better
Elon mentions (neutral tone) - elon, musk
Twitter Logo change - Bird, Blue, Black

Negative Sentiment
Platform decline due to updates - account, app, updates
Disappointment with Elon - elon, ruined, sucks, worst
Bad experiences - bad, worse, sucks, worst
Strong ideological/political frustration - racism, hate, biased

Conclusion

Conclusion

1. Technical Summary

- **VADER** worked but was less accurate on Threads due to noisy labels.
- **DistilBERT** performed better by capturing context and emotion more effectively.
- **BERTopic** produced the clearest platform-specific topics, outperforming LDA and NMF for short, noisy reviews.

2. Key Insights / Recommendations

- **Threads reviews** focused on stability, missing features, and comparisons to Instagram/Twitter.
  →New products must deliver a stable core experience first.
- **Twitter reviews** centered on leadership changes(Elon Musk), rebranding, bugs, and rate-limit frustration.
  →Major updates or rebranding require careful, transparent communication to maintain user trust.

Limitations

- VADER struggles with mixed or nuanced sentiment.
- Pseudo-labels can introduce noise into model training.

Future work

- Explore Top2Vec for topic extraction.
- Investigate multi-label classification for reviews expressing multiple themes.

Thank You

Group Corpus Crew:
Sai Rachana Kandikattu, Snehitha Tadapaneni, Haeyeon Jeong