Text Analytics Process: Hands-On Practice

From Raw Reviews to Business Decisions

Today's Mission

You're a data analyst at a soap company. Customer reviews are piling up. Your boss wants answers by Monday. Let's use text analytics to find them.

Part 1: Understanding the Pipeline

Draw arrows connecting each step to its output:

STEPS:

- Collect reviews
- Clean the mess
- Find patterns
- Make decisions

OUTPUTS:

- Business recommendations
- Structured data
- Raw text chaos
- Statistical insights

Part 2: Collection Methods

Your company wants to analyze reviews. Match each method to its best use case:

Methods: A. API (official connection) B. Web scraping C. Data partnership D. Manual collection

Scenarios:

1. Amazon won't give you API access: _____
2. Walmart offers bulk data downloads: _____
3. You need 50 reviews for a pilot: _____
4. Twitter provides developer access: _____

Cost Reality Check:

- Most expensive method: _____
- Most reliable method: _____
- Fastest for small samples: _____

Part 3: Cleaning the Chaos

Here's a real review. Clean it step by step:

Raw Review: "OMG this soap is AMAZINGGGGG!!! 😍😍 Best purchse everrr... my skinnn feels so softttt #blessed"

Step 1 - Lower case:

Step 2 - Remove punctuation:

Step 3 - Fix spelling:

Step 4 - Remove emojis/hashtags:

Step 5 - Final cleaned version:

Why do we clean?

- Computers see "Amazing" and "amazing" as different: True / False
- Typos confuse algorithms: True / False
- Emojis always ruin analysis: True / False

Part 4: Building Your Rules

You're analyzing shampoo reviews. Create word lists for detection:

Aspect Detection Rules

```python
# Hair texture words
texture_words = ["smooth", "___", "___", "___", "___"]

# Scalp health words
scalp_words = ["dandruff", "___", "___", "___"]

# Value perception words
value_words = ["worth", "___", "___", "___"]

# Effectiveness words
effectiveness = ["works", "___", "___", "___"]
```

Sentiment Indicators

Positive indicators for shampoo:

1. ---
2. ---
3. ---

Negative indicators for shampoo:

1. ---
2. ---
3. ---

Urgent/Safety words to flag immediately:

1. ---
2. ---

Part 5: The Four Techniques in Action

Apply each technique to this review set:

Reviews:

1. "Love Dove soap but switched to Olay. Better price."
2. "Horrible rash after using. Went to doctor."

3. "Smells nice. Lathers well. Too expensive though."
4. "Best soap for sensitive skin. Highly recommend!"
5. "Packaging broke during shipping. Product fine."

Technique 1: Sentiment Analysis

Rate each review:

1. Positive / Negative / Mixed
2. Positive / Negative / Mixed
3. Positive / Negative / Mixed
4. Positive / Negative / Mixed
5. Positive / Negative / Mixed

Overall sentiment: ___% positive

Technique 2: Topic Modeling

What topics emerge? (List top 3)

1. _____
2. _____
3. _____

Technique 3: Entity Recognition

Find all mentioned entities:

- Brands: _____
- Medical: _____
- Product aspects: _____

Technique 4: Classification

Assign each review to a department:

1. Review 1 → Marketing / Operations / R&D / Safety
2. Review 2 → Marketing / Operations / R&D / Safety
3. Review 3 → Marketing / Operations / R&D / Safety
4. Review 4 → Marketing / Operations / R&D / Safety
5. Review 5 → Marketing / Operations / R&D / Safety

Part 6: Finding Business Patterns

You've analyzed 1,000 soap reviews. Here are your results:

Aspect Mentions:
- Dry skin: 423 times (89% negative)
- Scent: 234 times (91% positive)
- Price: 198 times (72% negative)
- Lather: 167 times (94% positive)
- Packaging: 43 times (53% negative)

Priority Matrix

Place each aspect in the right quadrant:

```
    High Mentions
        |
FIX NOW  |  PROMOTE
---------|----------
MONITOR  |  IGNORE
        |
    Low Mentions
```

Your placement:

- Fix Now: _____
- Promote: _____
- Monitor: _____
- Ignore: _____

Business Recommendations

Based on the data, write 3 specific actions:

1. **Engineering/R&D should:**

   _____

2. **Marketing should:**

   _____

3. **Operations should:**

   _____

Part 7: Calculate the ROI

Your analysis found that 40% of negative reviews mention dry skin.

Current situation:

- Monthly sales: 10,000 units
- Return rate: 15%
- Returns cost: $25 per unit

If fixing the formula reduces returns by 40%:

- New return rate: ____%
- Monthly returns saved: ____ units
- Monthly savings: $_____
- Annual savings: $_____

If the fix costs $100,000: Payback period: _____ months

Part 8: Build Your Detection Rules

Create a simple Python-style rule to flag urgent reviews:

```python
def needs_urgent_attention(review_text):
    urgent_words = [_____, _____, _____, _____]

    review_lower = review_text.lower()

    for word in urgent_words:
        if word in review_lower:
            return True
    return False
```

Test your rule on:
"Had an allergic reaction, needed medical attention"
Would flag: Yes / No

"Product quality went downhill recently"
Would flag: Yes / No

Part 9: Scaling Up

From 100 to 1 Million Reviews

At 100 reviews:

- Time to read manually: _____ hours
- Patterns you'd find: Few / Many
- Confidence in decisions: Low / High

At 10,000 reviews:

- Time to read manually: _____ hours
- Need automation? Yes / No
- Best approach: _____

At 1 million reviews (like Unilever):

- Only option: _____
- Processing time with AI: _____
- Cost of NOT automating: _____

Part 10: Real-World Application

Pick a real product you use. Design a mini analysis:

Product: _____

Where to get reviews:

- [ ] Amazon
- [ ] Company website
- [ ] Social media
- [ ] Reddit
- [ ] Other: _____

Aspects to track:

1. _____
2. _____
3. _____

Business questions to answer:

1. _____?
2. _____?

Expected insights:

- Biggest complaint will be: _____
- Biggest strength will be: _____

Action you'd recommend:

Quick Decision Tree

You found 500 reviews mentioning "sticky residue."

Decision process:

1. What % of total reviews? _____
2. Sentiment of these mentions? _____
3. Competitor mentioned as alternative? Yes / No
4. Trending up or down? _____

Final decision:

- ☐ Urgent fix needed
- ☐ Monitor for now
- ☐ Not significant

Reflection Questions

1. **What surprised you most about text analytics?**

   ---

2. **Which technique seems most useful for your career?**

   ---

3. **What product would you want to analyze?**

   ---

4. **One company that needs this badly:**

Your Week Challenge

1. **Monday:** Find 20 reviews for any product
2. **Tuesday:** Clean and structure them
3. **Wednesday:** Count aspect mentions
4. **Thursday:** Calculate sentiment
5. **Friday:** Present 3 recommendations

Product chosen: _____ **Platform:** _____

Remember

You just learned how Unilever analyzes 411 million reviews daily. The process is the same whether you have 50 reviews or 50 million. The only difference is the tools.

Start small. Think big. Build your portfolio.

This skill = $100K+ career.