Sentiment Analysis Bericht

Powered by GenAI & DistilBERT

Table of Contents

Data Source: Unknown Source

Analysis Date: November 27, 2025

Total Comments Analyzed: 131

Neural Network Model: DistilBERT-based sentiment classifier

Executive Summary

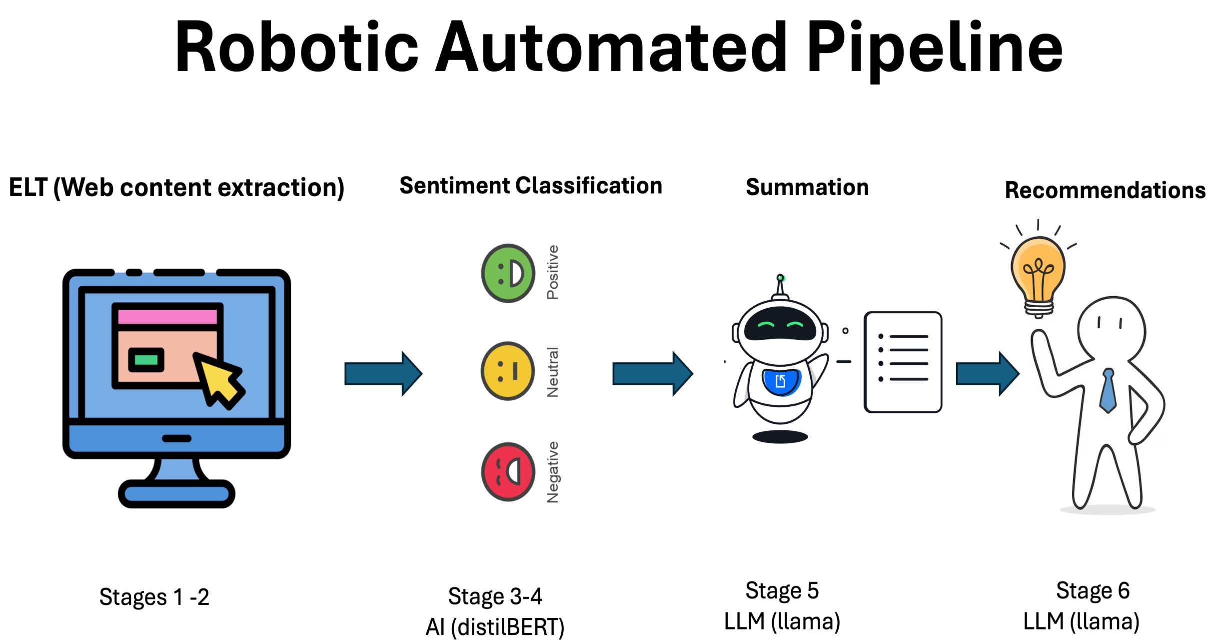
This report presents a comprehensive sentiment analysis of 131 user comments extracted from Unknown Source using advanced neural network classification. The analysis employed DistilBERT for sentiment classification and TF-IDF vectorization with K-means clustering for representative comment selection. **Key Findings:**
• Positive Comments: 106 (80.9%)
• Negative Comments: 23 (17.6%)
• Neutral Comments: 2 (1.5%)
The analysis utilized vector search and clustering algorithms to identify the most representative comments from each sentiment category, providing insights into common themes and user experiences.

Methodology

2.1 System Overview:

This system implements a comprehensive AI pipeline utilizing three complementary neural network architectures (BERT, BART, and LLaMA) to extract, analyze, and synthesize user-generated feedback into actionable insights. The pipeline processes any type of categorizable information, from restaurant reviews (helping service providers improve their offerings) to creditworthiness assessment and risk estimation for individuals or organizations.

Beyond generating summaries, the system delivers two critical metrics:
1. Semantic Vector Analysis: Identifies representative comments (positive, negative, and neutral) that encapsulate the most frequently encountered themes across the entire dataset.
2. Confidence-Based Selection: Extracts comments with the highest sentiment confidence scores, revealing the most emphatic responses (highly satisfied, strongly dissatisfied, or distinctly neutral perspectives).



Automated Processing Pipeline Architecture

2.2 Processing Pipeline:

Stage 1 - Data Acquisition: Target webpage URL → Automated text retrieval from web content

Stage 2 - Content Extraction: Raw HTML → Clean, structured comment blocks

Stage 3 - Sentiment Classification: BERT-based LLM (DistilBERT via HuggingFace) → Categorization into Positive/Negative/Neutral with confidence scores

Stage 4 - Representative Discovery: TF-IDF Vectorization + K-means Clustering → Identification of thematically central comments

Stage 5 - Summarization: LLaMA LLM (via Groq API) → Concise synthesis of sentiment patterns and key themes

Stage 6 - Recommendation Generation: LLaMA LLM analysis of positive/negative summaries → Actionable improvement suggestions

This multi-stage architecture ensures robust analysis through specialized neural networks, each optimized for its specific task in the pipeline.

2.3 Technical Implementation Details

1. Data Extraction: Comments were pre-filtered using neural network-based quality scoring to identify genuine user-generated content versus website boilerplate.

2. Sentiment Classification: DistilBERT model processed each comment to classify sentiment as Positive, Negative, or Neutral with confidence scores.

3. Vector Search & Clustering: TF-IDF vectorization transformed text into numerical representations, followed by K-means clustering to group similar comments.

4. Representative Selection: For each cluster, the comment closest to the centroid was selected as the most representative example of that theme.

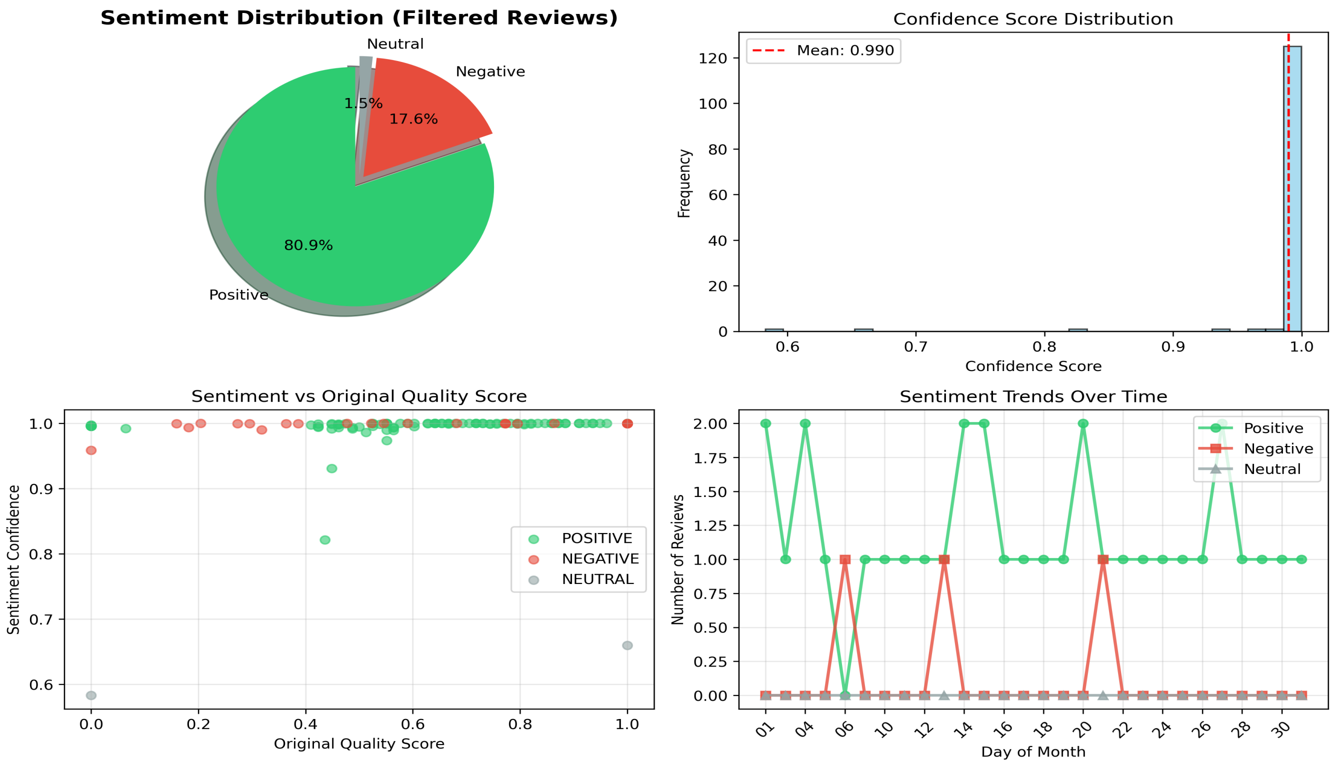
5. AI-Powered Summarization: LLaMA 3.1 model (8B parameters) analyzed representative comments to generate concise summaries highlighting key patterns and themes.

6. Recommendation Synthesis: LLaMA model processed positive and negative summaries to produce actionable improvement recommendations tailored to the specific domain.

3.1 Sentiment Analysis Overview

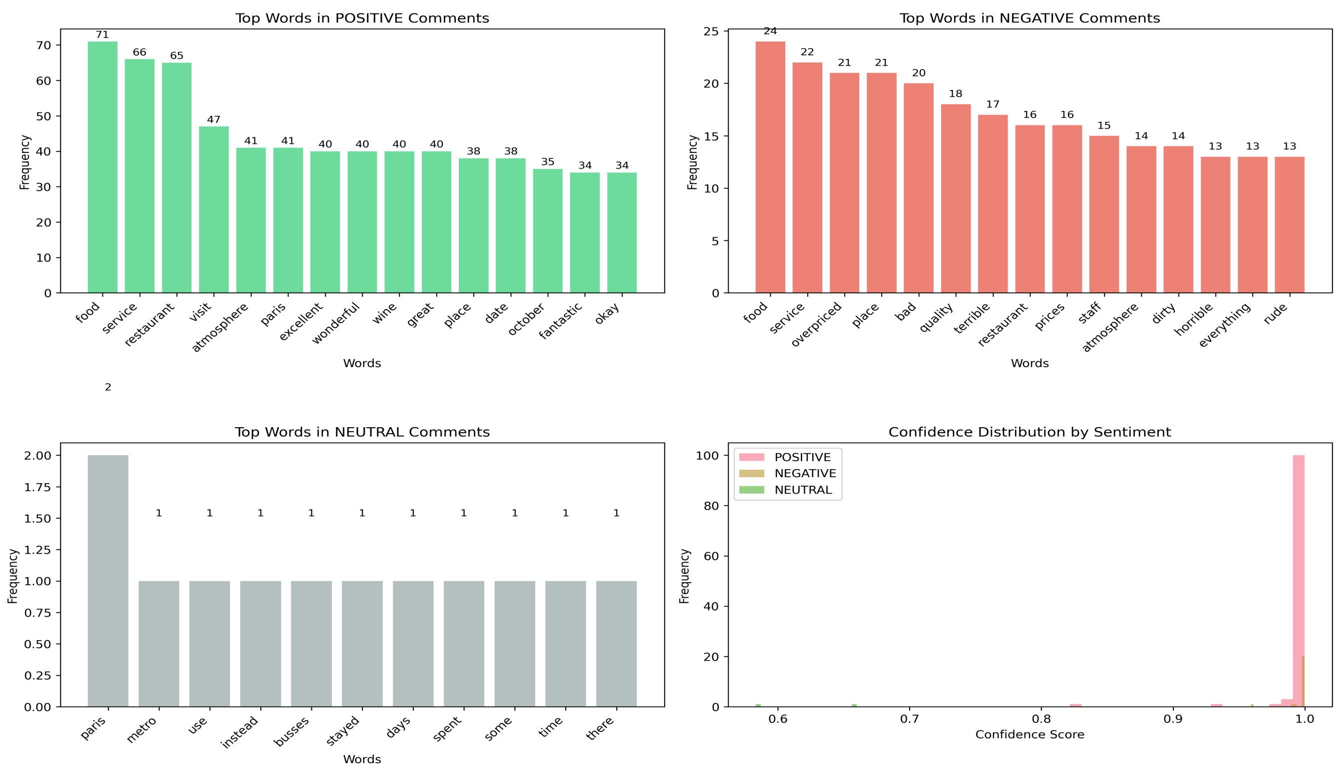
Negative Neutral Positive

The sentiment analysis overview provides a comprehensive visual breakdown of the comment distribution. The pie chart displays the proportional fractions of positive, neutral, and negative comments, offering an immediate understanding of overall sentiment balance. The sentiment trends over time chart illustrates how positive and negative sentiments fluctuate across the analysis period, revealing temporal patterns and shifts in user opinions. Additionally, the confidence score distribution histogram shows the reliability of sentiment classifications, with higher scores indicating greater model certainty.



3.2 Sentiment Word Clouds

3.3 Word Frequency Analysis



Comment Selection Methodology

This report displays two types of representative comments for each sentiment category:

1. Vector-Mean Comments (Cluster Centroids):
These comments were selected using TF-IDF vectorization and K-means clustering. For each sentiment, comments were grouped into 10 clusters based on semantic similarity. The comment closest to each cluster's centroid (mathematical center in vector space) represents the most typical expression of that theme. We display the top cluster's representative comment here.

2. Highest Confidence Comments:
These are the comments where the DistilBERT sentiment classifier had the highest confidence score (closest to 1.0). These represent the clearest, most unambiguous examples of each sentiment, where the neural network was most certain about the classification.

3. AI-Generated Summaries:
Using the Groq API with Llama 3.1 model, we generated concise 2-3 sentence summaries that synthesize the main themes and patterns across all representative comments in each sentiment category.

5. Vector-Mean Comments

5.1 Positive Comments

Confidence: 0.994

Cluster Info: Cluster 0 (Size: 6)

Comment: •Stopped for lunch during our stay. The menu offers decent variety. Service was adequate. The food arrived in reasonable time. The taste was okay but nothing memorable. Prices are fair for the location. The restaurant was clean. It's a passable option if you're nearby.

5.2 Negative Comments

Confidence: 1.000

Cluster Info: Cluster 0 (Size: 3)

Comment: •Very disappointing meal. The food quality was poor and nothing like the photos. Service was terrible - staff seemed annoyed by our presence. The portion sizes were tiny for such high prices - completely overpriced. The atmosphere was bad with uncomfortable seating. Everything felt rushed and impers...

5.3 Neutral Comments

Confidence: 0.583

Cluster Info: Cluster 0 (Size: 2)

Comment: •Paris has a metro, you can use it instead of busses.

6. Highest Confidence Comments

6.1 Positive Comments

Confidence Score: 0.9999

Comment: •Fantastic restaurant! The food is excellent and full of wonderful flavors. Service was great and very professional. The atmosphere is nice and elegant. Everything was perfectly prepared. The wine selection is spectacular. The desserts were amazing. Highly recommend this place to everyone. A truly g...

6.2 Negative Comments

Confidence Score: 0.9998

Comment: •Absolutely terrible restaurant. The food was bad and tasteless. Everything seemed old and poorly prepared. The service was horrible - our waiter was rude and dismissive. We found the place dirty and poorly maintained. The prices are outrageously overpriced. The kitchen clearly doesn't care about qu...

6.3 Neutral Comments

Confidence Score: 0.6598

Comment: •We stayed 5 days in Paris and spent some time there.

7. AI-Generated Sentiment Summaries

7.1 Positive Summary

Model Used: llama-3.1-8b-instant

Comments Analyzed: 10

Summary: The commenters found the restaurants to be positive experiences, citing common themes of excellent service, delicious food, and pleasant atmospheres. Many reviewers praised the restaurants' cleanliness, convenient locations, and reasonable prices, making them suitable options for casual meals or special occasions. Overall, the reviewers' comments highlight the restaurants' ability to provide high-quality dining experiences.

7.2 Negative Summary

Model Used: llama-3.1-8b-instant

Comments Analyzed: 10

Summary: The commenters found the restaurant to be severely lacking in terms of quality and value, with common complaints including poor food quality, high prices, and subpar service. Many reviewers also mentioned issues with cleanliness, uncomfortable seating, and an unpleasant atmosphere. Overall, the commenters felt that the restaurant failed to meet their expectations and was not worth the money spent.

7.3 Neutral Summary

Model Used: llama-3.1-8b-instant

Comments Analyzed: 2

Summary: The commenters found the aspect of transportation or travel information neutral, as they simply mentioned alternatives or the duration of stay without expressing any strong opinions or emotions. This suggests that the commenters were providing factual or practical information rather than evaluating their experience. The neutral tone may indicate that the commenters were focusing on the basics or logistics of their trip.

8. AI-Generated Recommendations

8.1 Actionable Improvement Suggestions

Based on Analysis of Positive and Negative Feedback

Model Used: llama-3.1-8b-instant

Generated: 2025-11-27T01:00:28.106276

Based on the positive and negative feedback summaries, here are three actionable recommendations for improving customer satisfaction at the restaurant:

1. **Focus on Consistency in Food Quality and Presentation**: The negative feedback highlights issues with poor food quality, which suggests that the restaurant may be having trouble maintaining a consistent standard. To address this, the restaurant could implement quality control measures, such as regular taste tests and kitchen inspections, to ensure that all dishes meet high standards. Additionally, they could consider revamping their menu to focus on seasonal, locally-sourced ingredients and offer more options for customization.

2. **Invest in Staff Training and Customer Service**: The negative comments about subpar service suggest that staff may not be adequately trained to provide exceptional customer experiences. The restaurant could invest in customer service training programs for all staff members, including servers, hosts/hostesses, and kitchen staff. This could help improve communication, resolve issues more efficiently, and create a more welcoming atmosphere for customers.

3. **Address Cleanliness and Ambiance Concerns**: The negative feedback highlights issues with cleanliness and uncomfortable seating, which can have a significant impact on customer satisfaction. To address these concerns, the restaurant could implement regular deep cleaning schedules, including daily sanitizing of surfaces and equipment. They could also consider redesigning the seating area to create a more comfortable and inviting atmosphere, perhaps by adding more natural light or upgrading furniture. Additionally, they could consider implementing a system for collecting customer feedback on cleanliness and ambiance to identify areas for improvement.

9. Bankruptcy Insurance Risk Assessment

9.1 Risk Calculation Formula

This assessment calculates bankruptcy insurance cost for loan applications based on customer sentiment analysis. The formula considers multiple risk factors:

Insurance Cost = Base Rate × Sentiment Multiplier × Confidence Multiplier × Sample Size Multiplier × Trend Multiplier

Where:
• **Sentiment Multiplier**: Increases with negative reviews (1.0 + negative_ratio × 2.5)
• **Confidence Multiplier**: Accounts for prediction uncertainty (1.5 - average_confidence × 0.5)
• **Sample Size Multiplier**: Adjusts for data sufficiency (higher for small samples)
• **Trend Multiplier**: Reflects recent sentiment changes (0.9 to 1.4 based on trends)

9.2 Risk Factors Breakdown

Risk Factor	Value	Multiplier	Impact
Base Insurance Rate	$5,000.00	1.00	Baseline
Positive Reviews	80.9%	1.37	Higher % reduces risk
Negative Reviews	17.6%	1.37	Standard
Average Confidence	0.990	1.01	Low confidence increases uncertainty
Sample Size	131 reviews	1.0	Small sample increases risk
Sentiment Trend	Improving	0.9	Recent patterns affect risk

9.3 Insurance Cost Estimate

Assessment Component	Result
Calculated Insurance Cost	**$6,215.28**
Risk Score	**35/100**
Risk Level	Medium

Interpretation:
Based on the sentiment analysis of customer reviews, the estimated bankruptcy insurance cost for this restaurant is **$6,215.28** with a **Medium** risk level (score: 35/100).

This assessment considers the overall sentiment distribution (80.9% positive, 17.6% negative), the confidence of sentiment predictions (avg: 99.0%), and recent sentiment trends (improving).

10. Technical Details

10.1 Processing Configuration:

- TF-IDF Features: 1000
- Minimum Document Frequency: 4
- Maximum Document Frequency: 0.8
- Clusters per Sentiment: 10
- Confidence Threshold: 0.8

10.2 Performance Metrics:

- Average Sentiment Confidence: 0.9899059447623392
- Total Samples Processed: 131
- Processing Time: 0.2225309411684672 minutes

10.3 Database Information:

- Source Database: filtered_reviews_eaf42e6a-faea-41bd-8c09-6e6689660e7b.db
- Analysis Timestamp: 2025-11-27T01:00:28.195552