From Rants to Riches: A Review-Reading Pipeline for Multilingual Sentiment Analysis Using Neural Networks and Transformers

HEC MONTRÉAL

Inès Ben Hamza, Inès Battah, Xun Lu

Abstract

User-generated data: a goldmine for marketing insights.

This project introduces a review-reading pipeline that leverages Sentiment Analysis, Topic Modeling, Transformers and Neural Networks to analyze multilingual customer feedback, classifying sentiment as positive, negative, or neutral.

In the **first part** of the project, we investigate multilingual sentiment analysis and topic modeling for consumer reviews, focusing on insights across English, French, and Spanish datasets using state-of-the-art transformer models to classify sentiments into positive, neutral, and negative categories. Subsequently, we apply topic modelling to gain deeper insights into customer feedback on our product.

In the **second part**, we used Neural Networks to classify customer feedback. Tokenized review text is processed with LSTM layers to capture sequential patterns, and metadata like language and topics is integrated. By building our sentiment analysis model from scratch, we aimed to tailor it for nuanced, multilingual, domain-specific sentiment classification.

Objectives

1. **Leverage transformers** to enhance sentiment classification accuracy and handle complex, multilingual datasets.
2. **Use topic modeling to uncover customer frustrations** by analyzing themes in feedback, providing actionable insights to improve products and address key pain points.
3. **Develop a multilingual review-reading pipeline** to classify customer feedback into positive, neutral, and negative sentiments across English, French, and Spanish datasets.

Methodology

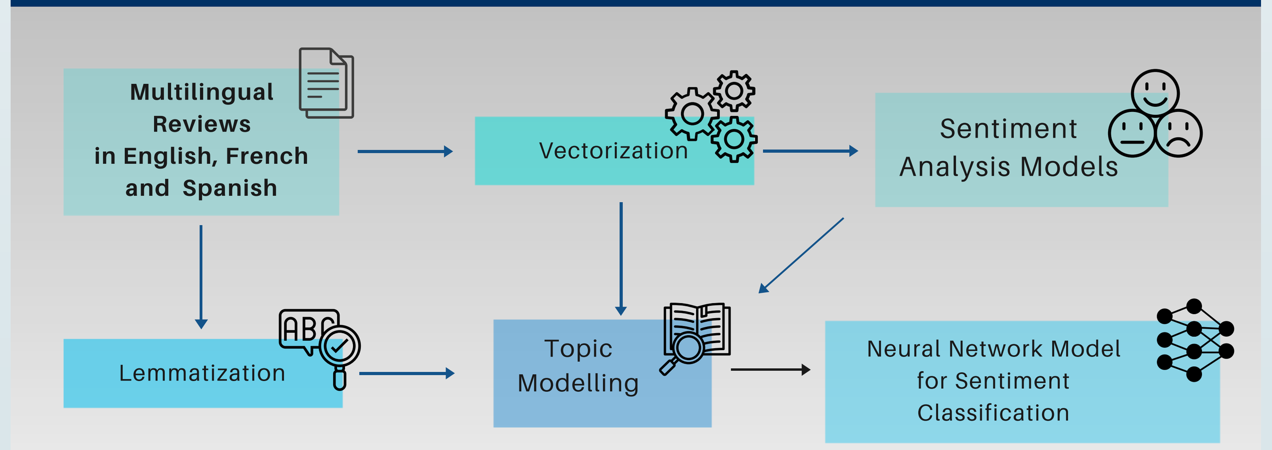

Figure 1: Multimodal & Multilingual Pipeline for Sentiment Review Reading

- **Data Preprocessing:** Reviews in English, French, and Spanish were tokenized, lemmatized, and cleaned. Non-Roman script reviews were excluded, and the dataset was balanced across sentiment categories.
- **Topic Modeling:** BERTopic, LDA, LSA, and NMF were leveraged to extract topics across languages.
- **Sentiment Analysis Models:** VADER served as a baseline for English reviews, while multilingual BERT-based models compared for more accurate sentiment classification.
- **Neural Network Sentiment Classification:** multimodal LSTM-based network processed text sequences (via embeddings) and structured features (language, topic) for improved sentiment predictions.

Models

Sentiment Analysis

We chose to use a **wide range of multilingual models** for different tasks to assess the sentiments of our reviews:

This product is awful and too expensive!	Ce produit est horrible et trop cher!	Este producto es horrible y demasiado caro!
BERT-Base 1 star, Score: 0.933	1 star, Score: 0.71	1 star, Score: 0.91
Fine-Tuned BERT-Base 1 star, Score: 0.94	5 stars, Score: 0.57	1 star, Score: 0.96
XLM-Roberta negative, Score: 0.95	negative, Score: 0.95	negative, Score: 0.95

Topic Modeling

Aspect	NMF	LSA	LDA	BERTopic
Methodology	Non-negative matrix factorization.	Singular Value Decomposition (SVD).	Probabilistic topic modeling.	Clustering-based using embeddings.
Input	TF-IDF matrix.	Document-term matrix.	Document-term matrix.	Transformer embeddings + TF-IDF
Output	Non-negative topic-word and doc-topic matrices.	Reduced matrix capturing latent semantics.	Topic distributions for docs and words.	Topic representations with embeddings and keywords

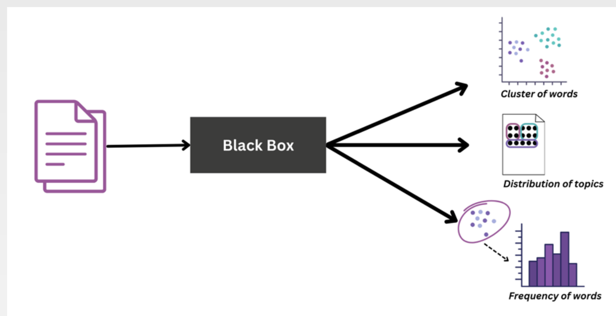

Figure 2: Average Topic Modeling Process

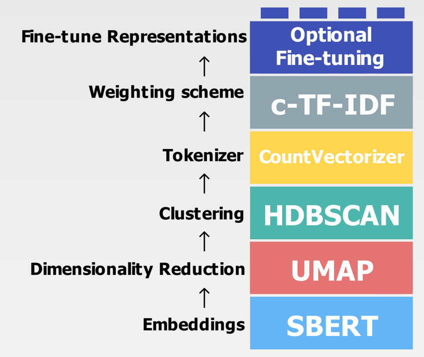

Figure 3: Overview of BERTopic's Architecture

Multilingual & Multimodal Neural Network

Model	Test Accuracy (%)	Test Loss
Activation Functions		
ReLU	**68.59**	**0.77**
Tanh	67.80	0.76
Leaky ReLU	67.41	0.77
ELU	68.35	0.76
Custom Weights & Preprocessing		
Custom Weights	65.00	0.74
Lemmatized Reviews	48.19	0.97
# of Hidden Layers (with ReLU)		
3	67.00	0.77
4	**68.59**	**0.77**
5	66.32	0.78

Table 2: Comparison of Model Performance

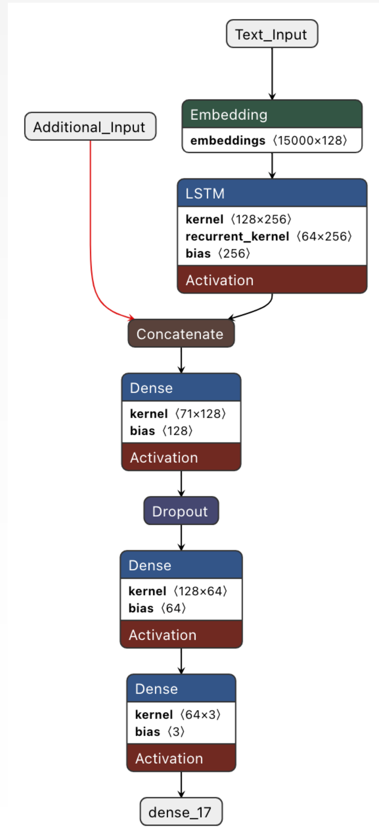

Figure 4: Architecture of final Neural Network

Results

Sentiment Analysis

Language	BERT-base	Fine-tuned BERT-base	XLM-Roberta
French	Accuracy : 0.77	0.74	0.68
	F1 score : 0.72	0.67	0.53
	F1 score weighted : 0.77	0.73	0.62
Spanish	Accuracy 0.76	0.73	0.67
	F1 score macro : 0.71	0.65	0.59
	F1 score weighted: 0.76	0.71	0.66
English	Accuracy 0.76	0.75	0.67
	F1 score macro : 0.72	0.68	0.57
	F1 score weighted : 0.76	0.74	0.65

Table 3: Classification Results for Sentiment Models across Languages

Topic Modeling

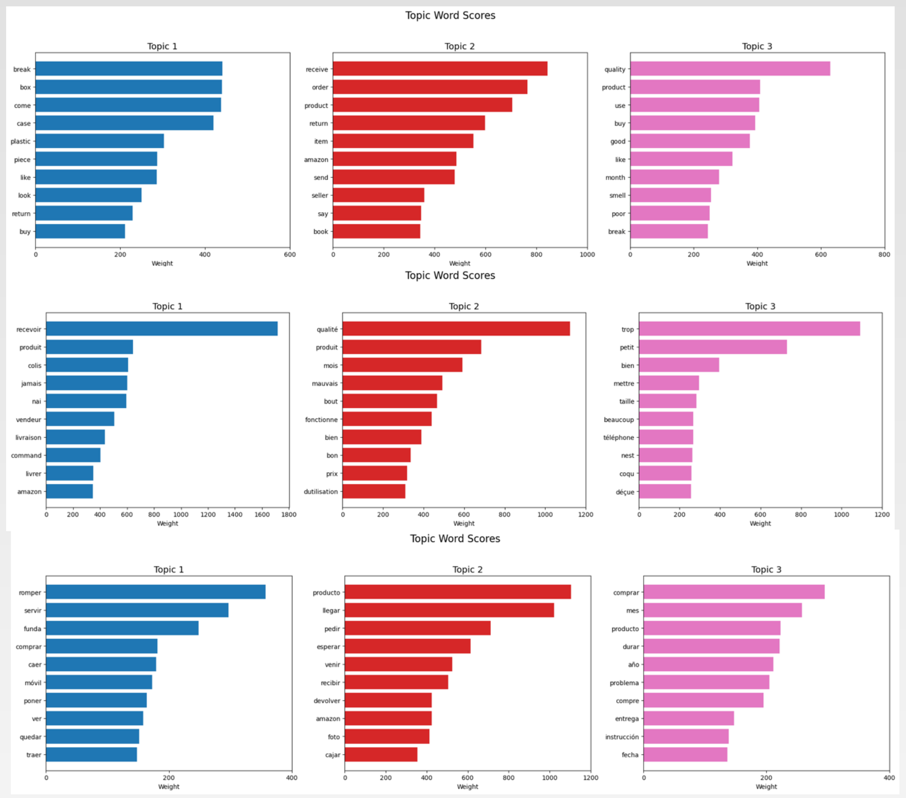

Figure 5: LDA Top 3 Topics in BERT-Predicted Negative Reviews (1- and 2-Star Ratings)

Multilingual & Multimodal Neural Network

Comparison of Classification Performance of NN with VADER Baseline

Model	Sentiment	F1-Score
Baseline (Vader, English only)	Negative	0.56
	Neutral	0.18
	Positive	0.69
Neural Network (Multimodal & Multilingual)	Negative	0.70
	Neutral	0.65
	Positive	0.72
Overall Accuracy: Baseline = 0.56, Neural Network = 0.69		

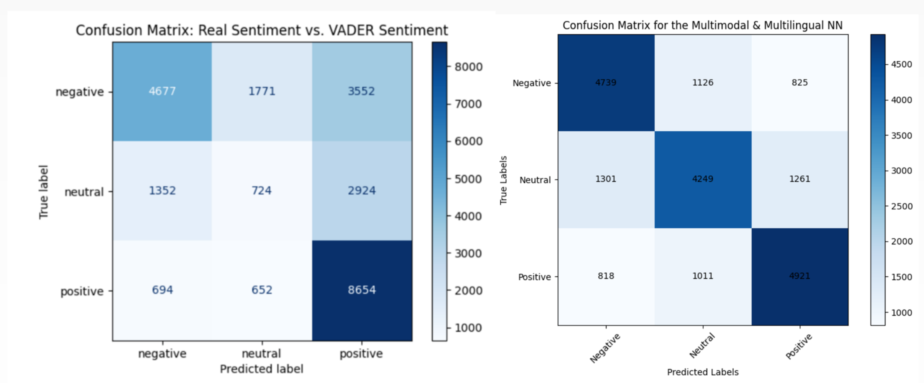

Figure 6: Confusion Matrices for Comparison of Model Performance

- **Better Neutral Sentiment Handling:** VADER struggles with nuanced neutral reviews, often misclassifying them as positive or negative, while our model outperforms by incorporating context and multimodal features.
- **LSTM for Context Understanding:** LSTM layer in our RNN captures sequential dependencies in text, improving the model's ability to understand subtle sentiment nuances.

Use case

This product is amazing and exceeded expectations, I'm really happy with it!	Ce produit est incroyable et a dépassé mes attentes, je suis vraiment content de l'avoir !	Este producto es increíble y superó mis expectativas, estoy muy feliz con él!
Prediction: **Positive** Sentiment Score: 0.84	Prediction: **Positive** Sentiment Score: 0.42	Prediction: **Positive** Sentiment Score: 0.90

Figure 7: Multilingual Sentiment Analysis with Neural Networks: Sample Predictions

Sentiment Analysis	Topic Modeling
Track overall customer satisfaction through reviews	Identify recurring complaints: "long shipping times" "poor customer service"
Highlight 5-star reviews and flag low-star reviews for further investigation	Identify specific areas for improvement: "battery life" , "screen durability" in electronics reviews

Figure 8: From Sentiment Analysis to Topic Modeling: Insights from Customer Reviews

Conclusion

Using sentiment analysis models, we extracted positive and negative reviews to perform topic modeling, gaining insights into recurring themes and analyzing customer feedback in our product reviews.

Sentiment Analysis & Topic Modeling Insights

- Extracted themes from positive and negative reviews in English, French, and Spanish.
- Identified common issues like poor product quality, functionality problems, delivery delays, and refund challenges across all languages.
- **Language-specific insights:**
  - **English:** Broken products, delivery inefficiencies, unclear return policies.
  - **French:** Undelivered packages, dissatisfaction with size and fit, courier inefficiencies.
  - **Spanish:** Fragile products, poor craftsmanship, need for simplified returns.

Neural Network Insight

Advantage Over VADER:
VADER struggles with nuanced expressions, neutral sentiment, and multilingual data. By leveraging LSTM layers for sequential text patterns and integrating metadata (e.g., language, topics), our model excels in context-aware sentiment classification.
Comparison to Transformer Models:
BERT-based techniques achieved higher accuracy in sentiment classification, particularly for positive and negative labels.
Takeaway:
While transformer-based models outperform in general sentiment classification, our custom NN complements them by offering actionable, domain-specific insights and better handling of neutral sentiment, making it an invaluable tool for business-focused analytics.

Acknowledgments

Grootendorst, Maarten. BERTopic Graph
GitHub, https://github.com/MaartenGr/BERTopic.

Hugging Face's **nlptown/bert-base-multilingual-uncased-sentiment**:
https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment

Hugging Face's **LiYuan/amazon-review-sentiment-analysis**:
https://huggingface.co/LiYuan/amazon-review-sentiment-analysis

Hugging Face's **cardiffnlp/twitter-xlm-roberta-base-sentiment**:
https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment
Kapadia, Kavita. "Introduction to Topic Modeling." GitHub, 2018,
https://github.com/kapadias/medium-articles/blob/master/natural-language-processing