

Executive Brief — Product Review Sentiment Analyzer

Project Overview

This project delivers a **full-stack NLP solution** that classifies customer reviews as ‘**negative, neutral, or positive**’. It combines a lightweight **TF-IDF + Logistic Regression baseline** with a modern **Transformer (DistilBERT)** model, wrapped in a user-friendly **Streamlit application** featuring **explainable AI (XAI)** through LIME and SHAP.

Business Problem

Companies often drown in unstructured product feedback (e.g., Amazon/Yelp reviews). Manual review is costly and inconsistent.

Goal: Automatically analyze sentiment, surface pain points, and build trust with stakeholders by providing transparent model explanations.

Technical Solution

- **Data pipeline:** Kaggle Amazon Reviews → cleaned & labeled (1–2 = neg, 3 = neu, 4–5 = pos).
- **Models:**
 - Baseline: TF-IDF + Logistic Regression.
 - Optimized: XGBoost on TF-IDF.
 - Advanced: DistilBERT fine-tuned with Hugging Face Transformers.
- **Explainability:**
 - LIME → word-level local explanations.
 - SHAP → global feature importances (XGB).
 - Confidence gauge → quick user trust indicator.
- **Deployment:**
 - Streamlit app with side-by-side baseline vs transformer.
 - Batch CSV analysis & docs gallery.
 - MLflow for experiment tracking.

Key Results

Model	Val F1	Test F1
TF-IDF + Logistic Reg	~0.72	~0.71
XGBoost (TF-IDF)	~0.63	~0.64
DistilBERT Transformer	~0.75	~0.73

- **Transformer consistently outperformed the baseline.**
- **Explanations show baseline mislabels where transformer succeeds.**

✅ Impact

- **For Business Analysts:** Instant insights from raw reviews.
- **For Customers:** Transparent reasoning behind sentiment labels.
- **For Engineers:** Reproducible end-to-end ML pipeline (data → training → explainability → deployment).

🚀 Next Steps

- Add multi-language support.
- Integrate with live review ingestion (API/streaming).
- Extend to recommendation systems (e.g., “customers disliked packaging issues”).
