Multilingual Customer Sentiment Analysis for Slang and Code-Switching

1. Introduction

Customer sentiment analysis plays a vital role in understanding public perception of brands, products, and services across digital platforms. Social media users increasingly communicate using informal language, slang, sarcasm, and code-switching between multiple languages within a single sentence. Examples include mixtures such as Hinglish or Spanglish, which are highly expressive but linguistically complex.

Traditional sentiment analysis systems are typically trained on monolingual, formal text and perform poorly when exposed to multilingual slang-heavy content. Accurately interpreting such data is essential for businesses seeking real-time customer insights in diverse markets.

2. Problem Statement

Most existing sentiment analysis models fail to capture the nuanced meaning of code-switched language, informal expressions, and sarcasm. Literal translations often distort sentiment polarity, while sarcasm detection remains an unresolved challenge in multilingual settings. As a result, sentiment predictions become unreliable, leading to flawed decision-making based on inaccurate analytics.

There is a need for a robust sentiment analysis system capable of handling multilingual, slang-rich, and sarcastic social media text while preserving cultural and contextual meaning.

3. Methodology

The system will collect multilingual social media data containing code-switching patterns and slang expressions. Text preprocessing will include language identification at the token level and normalization of informal expressions.

Transformer-based language models pre-trained on multilingual corpora will be fine-tuned for sentiment classification and sarcasm detection. Contextual embeddings will be used to capture cross-language dependencies within the same sentence.

Evaluation will be conducted using sentiment classification accuracy, sarcasm detection precision, and robustness across language combinations. The final system will output sentiment scores along with confidence indicators.