Objective

Improving the interpretability of sentiment evaluation models in Natural Language Processing (NLP) is the intention of this thesis. By often including polarised terms to textual datasets that mimic sentiment modifications withinside the actual world, this observe seeks to recognize the results and analyse how they have an effect on sentiment evaluation algorithms primarily based totally on neural networks. In order to make sure that those models are relevant in high-stakes situations, the very last intention is to promote transparency and accept as true with in them.

Methodology

This study introduces a novel approach to sentiment analysis by treating each sentence as a sequence of sentiment elements over time. It employs advanced neural network architectures, including Long Short-Term Memory (LSTM) and Bidirectional LSTM models, along with the VADER Sentiment Analysis toolkit, to assess the impact of sentiment augmentation techniques. In order to create datasets for evaluation research, polarised terms are injected into sentences and dispersed in line with opportunity algorithms.Individual forecasts have been interpreted the use of strategies like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), which highlighted the effect of precise input characteristics.

Key Findings

The consequences of the assessments tested a outstanding development withinside the model's interpretation and normal efficacy.With a check accuracy of 96.61% and an Area Under the Curve (AUC) rating of 0.9996, the stepped forward model tested nearly the most brilliant prediction capabilities. Additionally, interpretation strategies looked at how injected terms affected sentiment predictions, imparting beneficial facts regarding the model's behavior. The assessment highlights the need for transparency AI to hold accountability and transparency. Additionally, research confirmed how nicely sentiment enhancement works with linguistic headaches like sarcasm and negation. Understanding the decision-making methods of neural networks became made simpler through the take a look at's identity of critical phrases and their respective weights in predictions.

Implications

This study applies to addressing numerous essential elements inclusive of healthcare, economic risk assessment, and legal report analysis. The discoveries enhance transparency, which facilitates create artificial intelligence that isn't simply green however moreover morally and legally sound. The research reaffirms that with a view to following legal norms just like the General Data Protection Regulation (GDPR), modern neural community models need to be blended with sturdy interpretability frameworks. The basis for destiny research in explainable artificial intelligence (XAI) is laid through this thesis which gives thoughts for combining local and global interpretation techniques. The precise emphasis on moral artificial intelligence behavior and adherence to legal necessities emphasizes the research's social significance.