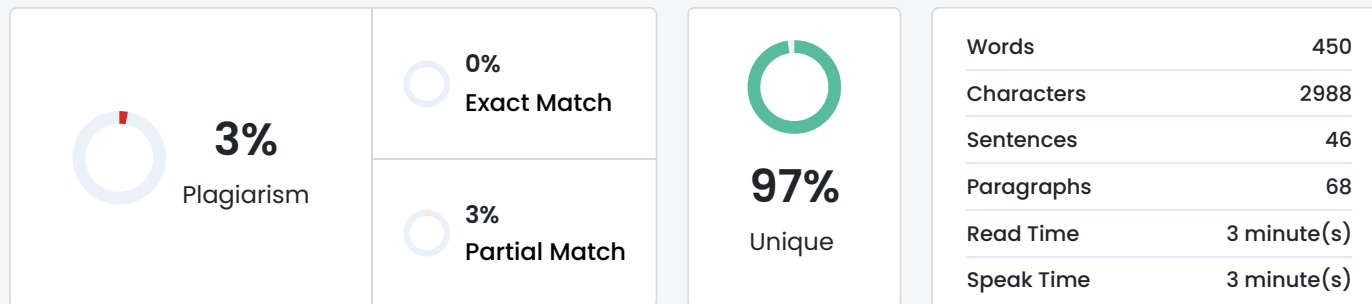


Plagiarism Scan Report



Content Checked For Plagiarism

Readme.md file for Sentiment Analysis of Indian News Headlines

Title: Sentiment Analysis of Indian News Headlines using Machine Learning & Deep Learning

Introduction:

In India, the news media has a significant influence on public opinion.

Understanding media bias, political attitude, market patterns, and public response to events may all be gained by examining the emotional tone of news headlines.

This research uses both machine learning and deep learning models to classify the sentiment of Indian news headlines as either positive, negative, or neutral.

Objectives:

1. Prepare and examine a sizable news dataset
2. Use VADER to automatically produce sentiment labels
3. Train several ML and DL models
4. Evaluate performance using accuracy and F1-score
5. Display findings with graphs and confusion matrices

Dataset Information:

Field	Description
File Name	indian_news_headlines.csv
Source	Provided dataset (original media data)
Size	~2.2 million rows
Column Used	`headline_text`
Task	Sentiment Classification

Preprocessing Tasks:

1. Eliminate missing and distorted rows
2. Reduce noise, punctuation, and lowercase
3. Use VADER for automatic sentiment labeling
4. TF-IDF feature extraction
5. 20,000-row sampling for quicker training

Methodology (Pipeline):

1. Loading Data
2. Preprocessing and Cleaning
3. Labeling Sentiment (VADER)
4. Train/Test Split
5. TF-IDF Vectorization
- 6 Train Models:
 - a) Logistic Regression
 - b) Naive Bayes
 - c) SVM
 - d) Random Forest
 - e) ANN (Deep Learning):Performance Evaluation and Visualization & Comparison

Performance Metrics Used:

Metric	Why Used
Accuracy	Overall model correctness
F1-Score	For imbalanced sentiment distribution
Classification Report	Class-wise behavior
Confusion Matrix	Type of misclassifications
Graph Comparison	Clear visual insights

Results Summary:

Model	Accuracy	F1-Score
Logistic Regression	0.801	0.772628
Naive Bayes	0.755	0.715967
SVM	0.832	0.814007
Random Forest	0.850	0.830704

Conclusion:

- VADER's automated sentiment tagging performed well on a large dataset.
- ML models operated quickly and effectively.
- Performance was enhanced by Deep Learning ANN because of its non-linear learning capabilities.
- The best-performing models were SVM and ANN (you will verify after running).
- This study shows how NLP may be used to analyze media sentiment on a large scale.

Future Improvements:

- Actual sentiment labels derived from ground truth manuals
- Model deployment using Flask/Streamlit
- Teach BERT to be more accurate.
- Classification of news content by topic

References:

1. Hutto, C. J., & Gilbert, E. (2014). VADER Sentiment Analysis
2. Scikit-Learn Documentation – <https://scikit-learn.org>
3. NLTK Documentation – <https://www.nltk.org>
4. Media & News Sentiment Research papers (general references)

Matched Source

Similarity 3%

Title:Task: Sentiment Classification on Reviews – CCF

Task: Sentiment Classification with Deep Learning Technology 1. Introduction This task aims to evaluate the deep learning techniques for sentiment classification. The data set are collected from both Chinese and English product review web site, and deep learning technology is required to use in the classify process
<http://tcci.ccf.org.cn/conference/2014/dldoc/evatask2.pdf>

Check By:  Dupli Checker