Final Project Technical Documentation

Project Title: NewsBot Intelligence System 2.0

Student Name: Sharise Griggs

Group Name: TeamSharise

Course: ITAI-2373 – Natural Language Processing

Date: August 3, 2024

🔍 Project Overview

NewsBot 2.0 is a production-ready news analysis platform built using NLP techniques. It helps users understand the topic, category, and emotional tone of any article. My goal was to create a smart, interactive tool that reflects what I've learned this semester from cleaning text to deploying web apps.

The system consists of:
1. A News Category Classifier
2. A Topic & Sentiment Analyzer

Both are supported by Flask apps, trained models, and thoughtful UX.

📓 Notebook_02.ipynb – News Category Classification

- Dataset: BBC News
- Preprocessing: Removed noise, used TF-IDF for vectorization
- Models: Trained Logistic Regression and Multinomial Naive Bayes
- Evaluation: Accuracy scores + confusion matrix
- Chosen Model: Logistic Regression
- Deployment: Saved model and vectorizer as .pkl files

Bonus Features:
- Created WordClouds per category
- Built a responsive Flask app
- Added confidence score + emoji-based predictions

📓 Notebook_03.ipynb – Topic Modeling & Sentiment Analysis

• Used LDA (Latent Dirichlet Allocation) to discover hidden topics
• Preprocessed the text and vectorized with CountVectorizer
• Set n_components=5 to capture high-level topic clusters
• Displayed top words per topic (saved in reflection journal)

🧠 Sample Topics from Model:
Topic 0: film, awards, music, star
Topic 1: technology, mobile, digital, phone
Topic 2: government, election, party, minister
Topic 3: game, england, win, players
Topic 4: market, company, economy, firm

• Sentiment: Used TextBlob for polarity + subjectivity scores
• Visuals: Used pyLDAvis to show topic clusters
• Final Output: A second Flask app that returns dominant topic + sentiment tone

🌐 Web Applications

1️⃣ Web App: News Category Classifier
• Predicts the type of news (tech, business, sports, etc.)
• Uses Logistic Regression + TF-IDF vectorizer
• Real-time predictions with confidence score
• Polished frontend (index_bonus.html)
• Emoji-based feedback and live input support

2️⃣ Web App: Topic + Sentiment Analyzer
• Returns the dominant topic from LDA
• Shows sentiment (positive, neutral, negative)
• Simpler UI but functional and fully integrated
• Uses saved models for fast results

💻 GitHub Repo Structure (Ready to Submit)

```
newsbot-flask/
├── app_classifier.py
├── app_sentiment_topic.py
├── models/
│   ├── lda_model.pkl
```

```
│   ├── vectorizer.pkl
│   └── classifier.pkl
├── templates/
│   └── index.html
├── static/ (optional)
├── notebooks/
│   ├── Notebook_02.ipynb
│   └── Notebook_03.ipynb
├── README.md
```