**Action Plan**

**Project Title:**

Fake News Detection Using Natural Language Processing

**Objective:**

To develop a machine learning model that can accurately detect fake news articles using textual features extracted via NLP.

**Team Role:**

Individual (Research and Implementation)

**Tools and Libraries:**

* Python
* Scikit-learn
* Pandas, NumPy
* NLTK
* Matplotlib, Seaborn
* WordCloud
* Jupyter Notebook

**Timeline and Activities:**

| **Date** | **Task** |
| --- | --- |
| April 3 | Project initiation and dataset research |
| April 6 | Data cleaning, tokenization, stopword removal, lemmatization |
| April 10 | Exploratory Data Analysis – class balance, top words, visualizations |
| April 14 | Feature extraction using TF-IDF |
| April 17 | Initial model testing: Random Forest, Naive Bayes, Logistic Regression |
| April 20 | Observed Random Forest overfitting; poor real/fake classification |
| April 23 | Switched to Logistic Regression after evaluating generalization |
| April 25 | Final model evaluation and confusion matrix plotting |
| April 28 | Deliverable writing and preparation |
| May 3 | PPT and project finalization |

**Key Changes Made:**

* **Model Change:** Switched from Random Forest to Logistic Regression
* **Reason:** Random Forest showed high training accuracy but struggled to generalize well (especially on real news). Logistic Regression provided balanced and stable results.

**Final Deliverables:**

1. Research Report ✅
2. Updated Action Plan ✅
3. EDA & Data Cleaning Notebook (in progress)
4. Final Model & Evaluation Notebook (in progress)
5. Streamlit Web App (optional)
6. GitHub README (optional)

**Remarks:**

All tasks have been completed on time. The model now performs with high accuracy and low misclassification. Ready for integration into a web interface.

**End of Action Plan**