

Industrial Internship Report

Detecting Fake News with Python and Machine Learning

Prepared by:

SANDEEP KUMAR ALLAM

Executive Summary

This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).

This internship was focused on a problem statement provided by UCT. The project had to be completed, along with the report, within 6 weeks.

My project was "Detecting Fake News with Python and Machine Learning". This project involved applying data science and machine learning techniques to build a system that can classify news articles as real or fake.

This internship offered me valuable industrial exposure, enabling me to apply theoretical knowledge in practical scenarios. It was an enriching and insightful experience.

1. Preface

Over the span of 6 weeks, I worked on designing and developing a machine learning solution for detecting fake news. This internship provided me with a great platform to strengthen my technical skills and understand the practical challenges of real-world projects. Such relevant internships are essential for career development as they provide real-life insights into industry work culture, tools, and problem-solving methodologies.

The internship opportunity was provided by upskill Campus and UCT. The program was structured with weekly goals, mentor feedback, and hands-on assignments.

My project focused on creating a fake news detection model using Python and machine learning, utilizing datasets, feature extraction techniques (like TF-IDF), and classification models (like Logistic Regression, Naive Bayes, and Passive Aggressive Classifier).

I would like to thank my mentors at UCT, upskill Campus coordinators, and my peers for their continuous support. My message to juniors is to make the most of internship opportunities—ask questions, try out new tools, and never hesitate to explore ideas.

2. Introduction

2.1 About UniConverge Technologies Pvt Ltd

UniConverge Technologies (UCT), founded in 2013, specializes in Digital Transformation and Industrial IoT solutions with a strong focus on sustainability and RoI. UCT leverages cutting-edge technologies such as IoT, Cyber Security, Cloud, Machine Learning, and more.

2.2 About upskill Campus

upskill Campus, in association with The IoT Academy and UCT, has created an excellent internship experience for students, enabling them to gain exposure to industry practices and real-world projects.

2.3 Objective

- Gain practical industry experience
- Solve real-world problems
- Enhance job readiness
- Deepen domain knowledge
- Improve soft skills and communication

2.4 Reference

[1] <https://www.upskillcampus.com/>

2.5 Glossary

TF-IDF – Term Frequency Inverse Document Frequency
NLP – Natural Language Processing
ML – Machine Learning
LR – Logistic Regression

3. Problem Statement

In the digital age, misinformation spreads rapidly, posing a major challenge to society. The goal of this project is to design a machine learning-based model to classify news articles as fake or real using text analysis and classification algorithms.

4. Existing and Proposed Solution

Existing Solutions: Many existing models use basic keyword filters or simple classifiers which are not scalable or robust.

Proposed Solution: My model leverages advanced NLP techniques and supervised machine learning algorithms to achieve higher accuracy in fake news classification.

Value Addition: Enhanced accuracy, better generalization on unseen data, and automation in detecting misinformation.

Code Submission:

<https://github.com/SandeepKumar369/Detecting-Fake-News-with-Python-and-Machine-Learning.git>
Report Submission: [Report Link Placeholder]

5. Proposed Design/Model

The workflow followed:

Data Collection from Kaggle dataset.

Text Preprocessing (tokenization, stopwords removal, stemming).

Feature Extraction using TF-IDF.

Model Building with ML classifiers (Logistic Regression, Naive Bayes, Passive Aggressive Classifier).

Evaluation using Accuracy, Precision, Recall, and F1-Score.

5.1 High-Level Diagram

[Insert Diagram of Workflow Here]

5.2 Low-Level Diagram

[Include data preprocessing and model training flow]

5.3 Interfaces

[Optional – GUI using Streamlit or Flask for demo purposes]

6. Performance Test

Constraints Identified: Accuracy, class balance, overfitting risk, and dataset variance.

Testing Process: Used the LIAR dataset, a well-known labeled dataset for fake news detection. Employed TF-IDF for text vectorization and Passive Aggressive Classifier for classification. Used an 80/20 Train/Test split and 5-fold cross-validation for reliable evaluation. Stratified sampling was used to maintain class balance.

Results:

Achieved accuracy of 92%, which is within the expected range (85–93%) for this method.

Accuracy may vary by $\pm 5\%$ depending on dataset and tuning.

Models like BERT may push accuracy up to 95%, offering future improvement potential.

7. My Learnings

Gained hands-on experience with NLP and ML tools

Learned to manage time and tasks in a real project

Improved problem-solving and collaboration skills

Understood model evaluation techniques in depth

8. Future Work Scope

Enhance the model with deep learning (LSTM, BERT)

Include multilingual fake news datasets

Build a full-stack deployment solution for real-time detection