**INFO 511 Project Proposal**

1. **Focus Area :**

This project’s main focus area will be Data Visualization.

It will involve Data Collection, Data Cleaning, EDA, Fake News Detection(Applying ML models), Sentiment Analysis and Visualization & Insights.

1. **Team details :**

This is a team project.

Team members : 1. Tanmay Nalawade – email id :- tanmaynalawade@arizona.edu

Github username :- Tanmay-Nalawade

2. Neha Chaudhari - email id :- nehaschaudhari02@arizona.edu

Github username :- nehachaudhari02

1. **Research question :**

**Question :-** How can machine learning techniques be applied to classify and predict the credibility of news articles in real-time?

This question focuses on using machine learning techniques to assess the credibility of news articles in real-time. It goes beyond just detecting fake news by also evaluating the trustworthiness of the sources and identifying patterns of reliable versus unreliable reporting.

1. **Framework, theory & Policies :**
2. **Framework :**

The framework for this project will rely on Machine Learning and Natural Language Processing (NLP) techniques to analyze and classify news articles. Methodologies like logistic regression, Random Forest, LSTM, etc. might be used.

1. **Theory :**

The project will be guided by theories media bias shapes public perception, along with machine learning and NLP methods for text classification and sentiment analysis.

1. **Policy :**

This project will adhere to ethical guidelines, ensuring the use of publicly available & anonymized data. Efforts will be made to minimize bias in the dataset and ensure fairness in the machine learning models. We will maintain transparency and accountability will be maintained throughout, with clear documentation of methods and results.

1. **Citation Style :**

The citation style used throughout the project will be APA(American Psychological Association)

1. **Dataset Discription :**

For this project we will be collecting data by scraping news articles from platforms like BBC, CNN, NY Times, and Twitter for real-time analysis. We might utilize datasets such as the LIAR Dataset and the Kaggle Fake News Dataset, which contain labelled news articles for fake news detection. These datasets will help in training and evaluating machine learning models for fake news classification

1. **Work Plan and Timeline Milestones :** The following is the timeline for the project :

* Data Collection and Cleaning : Completed by February 24th, 2025
* Exploratory Data Analysis : Completed by March 6th, 2025
* Model Development : Completed by March 18th, 2025
* Design of Visualizations : Completed by March 25th, 2025
* Evaluation and Model Tuning : Completed by March 30th, 2025
* Final Report and Documentation : Completed by April 10th, 2025
* Final Presentation : Completed by April 15th, 2025