**Project Proposal**

**Title: Fake Review Detector**

**Group Members:**

**Muhammad Azfar Waqas (UW-23-CY-BS-013)**

**Ibrar Ul Hassan Shami (UW-23-CY-BS-018)**

**Saad Ali (UW-23-CY-BS-050)**

**Table of Contents**

[1. Project Overview: 3](#_Toc187174422)

[2. Objectives: 3](#_Toc187174423)

[2.1. Detect Fake Reviews: 3](#_Toc187174424)

[2.2. Improve User Trust: 3](#_Toc187174425)

[2.3. Contribute to Research: 3](#_Toc187174426)

[3. Key Features: 3](#_Toc187174427)

[3.1. Data Preprocessing: 3](#_Toc187174428)

[3.2. Feature Engineering: 3](#_Toc187174429)

[3.3. Model Building: 3](#_Toc187174430)

[3.4. Model Persistence: 3](#_Toc187174431)

[4. Requirements: 3](#_Toc187174432)

[4.1. Libraries and Tools: 3](#_Toc187174433)

[4.2. Dataset: 4](#_Toc187174434)

[4.3. Hardware/Software: 4](#_Toc187174435)

[5. Development Phases: 4](#_Toc187174436)

[5.1. Data Preprocessing: 4](#_Toc187174437)

[6. FAQs: 5](#_Toc187174438)

# Project Overview:

This project aims to develop an AI-based solution to identify fake reviews using Natural Language Processing (NLP) techniques. Fake reviews can mislead consumers and harm businesses, making detection essential for maintaining trust in online platforms. The proposed system will preprocess review text, extract features, and employ machine learning to classify reviews as genuine or fake.

# Objectives:

## **Detect Fake Reviews:**

Identify fraudulent reviews among genuine ones using NLP techniques.

## **Improve User Trust:**

Help online platforms build trust with their users by filtering out deceptive reviews.

## **Contribute to Research:**

Provide insights into text classification and machine learning techniques for fake review detection.

# Key Features:

## Data Preprocessing:

* + Handle missing values and format inconsistencies.
  + Perform text cleaning, including removing HTML tags, punctuation, and numbers.
  + Apply lemmatization and stop word removal.

## Feature Engineering:

* + Extract text features using TF-IDF (Term Frequency-Inverse Document Frequency).
  + Incorporate metadata like "Verified Purchase" as an additional feature.

## Model Building:

* + Train a Support Vector Machine (SVM) classifier for binary classification.
  + Evaluate the model's accuracy and other performance metrics.

## Model Persistence:

* + Save the trained model for future use in the pickle library.

# Requirements:

## Libraries and Tools:

* + Python 3.7+
  + pandas, numpy for data handling.
  + nltk, nlppreprocess for text preprocessing.
  + scikit-learn for machine learning and evaluation.
  + pickle for model serialization.

## Dataset:

* + A labelled dataset of reviews indicates whether they are fake or genuine.
  + Example datasets can be picked from Kaggle, UCI repository

## Hardware/Software:

* + Python IDE (VS Code, PyCharm, or Jupyter Notebook).
  + Sufficient computational resources to handle data preprocessing and model training.

# Development Phases:

## Data Preprocessing:

* + Load and clean the dataset.
  + Apply text preprocessing techniques such as removing HTML tags, lemmatization, and stop word removal.
  1. **Feature Engineering**:
  + Perform TF-IDF vectorization on the preprocessed text.
  + Add metadata features like "Verified Purchase."
  1. **Model Training**:
  + Split data into training and test sets.
  + Train an SVM classifier.
  + Evaluate the model's accuracy, precision, recall, and F1 score.
  1. **Deployment**:
  + Save the trained model.
  + Provide an interface for user input to classify reviews in real-time.

# FAQs:

1. **Why are you developing this project and What will this project give to the industry?**

This project is being developed to address the growing issue of fake reviews, which undermine consumer trust and negatively impact businesses. The primary motive is to leverage artificial intelligence (AI) and NLP techniques to provide a reliable solution for detecting deceptive reviews. By creating this system, we aim to empower e-commerce platforms, review aggregators, and other stakeholders with tools to ensure transparency and authenticity in online reviews. This project seeks to contribute to the industry by demonstrating how AI can solve real-world problems effectively.

1. **How does this project benefit the end user?**

The Fake Review Detection system benefits end users by:

* **Enhancing Trust**: It ensures that the reviews they read are genuine, helping them make informed purchasing decisions.
* **Saving Time and Money**: Users can avoid low-quality products or services promoted through deceptive reviews.
* **Improving User Experience**: By filtering out fake reviews, the system makes online platforms more reliable and user-friendly.