AI Internship Challenge – Text Classification Task

Goal
Build a small AI/ML project to demonstrate your ability to work with data, preprocess it, train a simple model, and evaluate results.

Dataset

Choose any **publicly available text classification dataset** (e.g., news, spam detection, sentiment analysis).

Requirements:

- At least **2–5 categories** of text.

- Enough samples to train and test a model (minimum a few thousand entries).

- The dataset must be publicly accessible.

Include the dataset link/source in your `README.md`.

Tasks / Features

Data Preparation

- Load and inspect the dataset.

- Perform basic preprocessing (clean text, remove stopwords, lowercase, etc.).

Exploratory Analysis

- Show simple statistics (e.g., number of samples per category, most frequent words).

Model Training

- Train a simple classifier (e.g., Logistic Regression, Naive Bayes, or a small Neural Network)

- Split the dataset into training and test sets

- Evaluate model accuracy (optionally include precision, recall, F1-score)

Prediction Script

- Provide a function or script where a user can input a sentence and the model predicts its category

Bonus

- Use TF-IDF or embeddings instead of raw bag-of-words

- Compare at least two models and explain which performs better

- Create a small visualization (e.g., confusion matrix)

- Wrap the prediction script in a simple Flask or FastAPI endpoint

Submission Instructions

- Push your code to a public GitHub repository

- Include a short README.md with instructions and explanations

- After completing, send the repository link to **office@linkplus-it.com**

- Any commits with timestamps after sending the confirmation email will **not** be considered