Corning Future Innovator Program2025

Corning Digital and IT Centre India

Instructions

- Form a team of **two** – comprising of undergraduate students of any engineering disciplines. You are not eligible to enter the competition individually.
- The problem is designed to encourage interdisciplinary thinking, robust solutioning and trending digital technology.
- Comprehend the problem statement and submit the abstract in PDF.
- Teams shortlisted for the final round will be emailed by **Friday, July 11, 2025**.
- All the necessary information is provided about the problem. In case, additional information is required, make suitable assumptions, and clearly state them.

By participating, you are agreeing to the following terms and conditions:

- Any information provided to Corning shall be considered non-confidential. You shall not share any information protected under any law, patent, confidentiality, or other contract, etc.
- Nothing provided by you shall prohibit or restrict Corning's right to develop, make, use, market, license, or distribute products and services. You acknowledge that Corning may already possess or have developed products or services similar to or competitive with those provided by you.

Problem 1 – Detecting Fake News

Background

In the current digital age, the pervasiveness of fake news has grown significantly, creating a global challenge that compromises the integrity of information, misguides public opinion, and can potentially incite real-world harm. The issue is further exacerbated by the speed and ease with which these false narratives can be disseminated across various social media platforms. Consequently, it is of utmost importance to address this issue in order to safeguard the accuracy and trustworthiness of news.

The objective of this project is to develop a reliable algorithm capable of detecting fake news. This algorithm should be able to analyze and categorize news articles or posts as either 'real' or 'fake' based on various features such as the credibility of the source, the language used, the consistency of the information with other reputable sources, among others.

The creation of such an algorithm would be a significant step towards mitigating the effects of fake news, by providing individuals and organizations with a tool to verify the authenticity of news content, thereby fostering a more informed and discerning public.

Problem

By leveraging natural language processing, machine learning, and deep learning techniques, the algorithm should be able to discern patterns and inconsistencies indicative of fake news. Furthermore, it should continually learn and adapt to the evolving nature of fake news in order to maintain high accuracy over time.

Combat misinformation by classifying news articles as real or fake. Possible skillsets/methodologies to use: NLP, TF-IDF or word embeddngs, Naïve Bayes, deep learning models.

Resources

Please refer data set Fake News Dataset on Kaggle.

Helper Code

Use Pandas for data manipulation, sklearn for machine learning, and tf-idf for text feature extraction. PassiveAggressiveClassifier.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.metrics import accuracy_score, confusion_matrix

# Load the data
df = pd.read_csv('news.csv')

# Get the labels
labels = df.label
```

```
# Split the dataset
x_train, x_test, y_train, y_test = train_test_split(df['text'], labels,
test_size=0.2, random_state=7)

# Initialize a TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer(stop_words='english', max_df=0.7)

# Fit and transform train set, transform test set
tfidf_train = tfidf_vectorizer.fit_transform(x_train)
tfidf_test = tfidf_vectorizer.transform(x_test)

# Initialize a PassiveAggressiveClassifier
pac = PassiveAggressiveClassifier(max_iter=50)
pac.fit(tfidf_train, y_train)

# Predict on the test set and calculate accuracy
y_pred = pac.predict(tfidf_test)
score = accuracy_score(y_test, y_pred)
print(f'Accuracy: {round(score*100,2)}%')

# Build confusion matrix
confusion_matrix(y_test, y_pred, labels=['FAKE','REAL'])
```

Solution Evaluation Metric

- Split the data into training data and data for evaluation.
- Create a model to detect fake news.
- Prove that the model detects fake news by training the model and then evaluating the model against the evaluation data. Log findings.
- Report out steps taken along the way, examples of fake news versus real news in power point format.