Research Task -

Bag of Words

We compare sentences with vocabulary and based on that we predict 0,1 or 2. For example, in the sentences "I love NLP" and "I love AI", the vocabulary is {I, love, NLP, AI}. Comparing each sentence with this vocabulary, "I love NLP" becomes [1, 1, 1, 0] and "I love AI" becomes [1, 1, 0, 1]. For each sentence or document, we count how many times each word from the vocabulary appears in it. If the word is present once, we write 1; if it appears twice, we write 2; and if it is not present, we write 0.

TF-IDF -- Term Frequency – Inverse Document Frequency

Term Frequency measures how often a word appears in a single document compared to the total number of words in that document. For example, in the sentence *"I love NLP, NLP is great"* there are 6 words in total, and the word NLP appears 2 times. So the TF(NLP) = 2/6 = 0.33.

Inverse Document Frequency tells how important a word is across all documents. If a word appears in many documents (like “the” or “is”), it is considered less useful and gets a low score. If a word appears in only a few documents (like “NLP”), it is considered special and gets a high score.

Word Embeddings   
  
Word Embeddings can convert words into numbers in such a way that computers can understand their meanings and relationships, and not just their counts. For example, girl is close to boy(gender relationship) ; apple is close to mango (fruits relationship).

One-Hot Encoding

It is very similar to Bag of words. The only difference is that in Bag of Words, we use to compare entire sentences and then write vectors, but here we compare only a single word. For example, if we have eg = {cat, dog, cow}  
cat → [1, 0, 0]  
dog → [0, 1, 0]  
cow → [0, 0, 1]

Event Task   
  
**1. Event Description**

AI Quest is a unique and interactive event designed to immerse participants in the world of Artificial Intelligence and Machine Learning. Unlike traditional hackathons or coding competitions, this event focuses on **problem-solving with real-world data** in a short, gamified format.

Participants will receive a surprise dataset and challenge statement only at the start, ensuring a fair competition. They will work in teams to explore the data, build predictive models, and present their insights creatively. The event combines **technical accuracy, innovation, and storytelling**, making it equally engaging for beginners and seasoned AI enthusiasts.

**2. Flow of the Event**

**Step 1: Kickoff Briefing (10–15 minutes)**

* Teams are introduced to the challenge theme (e.g., healthcare, entertainment, environment, finance).
* A surprise dataset is shared (examples: predicting air quality, movie box office ratings, fake news detection, or stroke prediction).
* Rules, judging criteria, and submission guidelines are explained.

**Step 2: Data Sprint (45 minutes – 1 hour)**

* Teams begin exploring the dataset.
* Tasks:
  + Cleaning missing values,
  + Visualizing patterns (graphs, plots, dashboards),
  + Identifying key features..

**Step 3: Model Building (1–1.5 hours)**

* Teams develop and train an ML model using tools of their choice: Python (Scikit-learn, TensorFlow, PyTorch) or even low-code/no-code ML platforms for beginners.
* Focus: Achieving good performance (accuracy, precision, recall, F1 score, etc.).
* Live leaderboard (optional): Teams submit their model predictions, and scores update in real-time on a screen

**Step 4: Pitch & Demo (5 minutes per team)**

* Each team presents:
  1. **Their approach** (data cleaning, features used, algorithm chosen).
  2. **Model results** (metrics, accuracy).
  3. **Key insights** from the data.
  4. **Creative twist**: A fun visualization, meme, or real-world application idea.
* Audience voting can be added for “Most Creative Presentation.”

**Step 5: Judging & Awards**

* **Judging Criteria**:
  + *50%*: Technical performance (model accuracy, relevance of approach).
  + *30%*: Creativity (unique insights, engaging visualizations, innovative ideas).
  + *20%*: Presentation & clarity.
* Awards could include:
  + Best Model Performance
  + Most Creative Insight
  + Best Team Presentation