**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING  
PROJECT REPORT**

**TITLE:** MACHINE LEARNING MODEL TO RECOGNIZE TABLE STRUCTURE AND INTERACT WITH CONTENT

**Team-20**  
**2320040001- Varshitha Reddy**  
**2320040026- G.V. Naga Jahnavi**  
**2320040130- A. Vaishnavi**

**PROBLEM STATEMENT:**  
To develop a machine learning model that can accurately recognize, extract, and interact with the structure of tables from various document formats, including PDFs.

**EXPLANATION:**  
With the rapid pace of digital transformation, organizations are collecting and processing vast amounts of data daily. Tabular data is a powerful way to represent information, found across various documents such as historical records, scientific journals, receipts, and financial reports. Automating the extraction and interaction with this data using artificial intelligence is essential for efficient and accurate information retrieval.

Traditional methods like text extraction from PDFs have been employed to identify and extract text and tables. However, with advancements in data science and machine learning, especially deep learning, there is a significant opportunity to enhance these processes. Our project leverages these advancements to not only recognize and extract tabular data from PDFs but also to enable users to interact with the extracted content by asking questions and receiving accurate responses. We use Streamlit to build and deploy this interactive model, making it user-friendly and accessible.

**ALGORITHM:**

* **Step 1: Table Detection**  
  The Table Detection step utilizes Camelot and machine learning models to detect tables within PDFs. This step ensures that all tabular data within the documents is identified for further processing.
* **Step 2: Table Recognition**  
  The Table Recognition step employs Camelot and machine learning models to discern the structure of the tables, including columns, rows, and individual cells. This recognition is crucial for understanding the data within the tables.
* **Step 3: Table Extraction and Interaction**  
  The Table Extraction and Interaction step not only extracts tables from the documents but also integrates them into a vector store, allowing users to ask questions related to the document's content. The model processes these questions, searches the relevant sections of the document, and provides detailed, context-aware responses. Additionally, users can request specific tables, which will be displayed directly in the application.