**KILL Project**

**Email spam dection**

**Year : 2nd year**

**Branch: CSD dayschoolars**

**Team number : 8**

**Team details :**

* + - **P.Ajay**
    - **P. Sai Kumar**
    - **R.Ravichandra**
    - **K.Mohan**
    - **S.Trilok**

**Problem statement:**

**Email spam detection**

**Spam emails are a significant issue for email service providers and users, leading to wasted time, security risks, and phishing attacks. The goal of this project is to develop a machine learning-based solution to automatically detect and filter spam emails with high accuracy, improving the user experience and reducing the risk of malicious attacks.**

**Domain :**

* + **NLP**
  + **Mechine learning**

**Technologies used :**

* **Python**
* **Scikit-learn**
* **TensorFlow**
* **NLTK (Natural Language Toolkit)**
* **Pandas, NumPy**
* **Jupyter Notebook**

**Approach to Do Project:**

* **Data Collection: Collected email data from open-source datasets like the Enron email dataset.**
* **Data Preprocessing:**

**Cleaned the text data (removing stop words, punctuation, and special characters).**

**Tokenized and converted text data into numerical format using TF-IDF and word embeddings.**

* **Model Selection and Training:**

**Trained various models including Logistic Regression, Naïve Bayes, and Random Forest.**

**Used deep learning models like LSTM for sequence-based analysis.**

* **Evaluation:**

**Evaluated models using metrics like accuracy, precision, recall, and F1-score.**

**Fine-tuned hyperparameters to improve performance.**

* **Deployment:**

**Integrated the best-performing model into an email filtering system.**

**Outcome:**

* **Achieved a classification accuracy of 96% using an LSTM model.**
* **Reduced false positives and improved spam detection efficiency.**
* **Enhanced user experience by minimizing the number of spam emails in the inbox.**

**Conclusion:**

**The project successfully developed a machine learning-based spam detection system with high accuracy and minimal false positives. The use of deep learning models significantly improved the system’s ability to identify spam emails accurately. Future improvements could include real-time detection and adapting the model to new spam pattrns**