BUILDING A SMARTER AI POWERED SPAM CLASSIFIER DEFINITION AND DESIGN TECHNIQUES.

DEFINITION:

It involves a combination of advanced techniques and careful design considerations.

DESIGN TECHNIQUES:

I.PROBLEM DOMAIN:

1. Clearly define what constitutes spam for your specific description.
2. Identify the objectives such as minimizing false positives and negatives.

II. DATA COLLECTION AND PREPROCESSING:

1. Gather a diverse and representative dataset of both spam and non-spam examples.

2. Preprocess the data which may include text normalization toxenization and removing step words.

III. FEATURE ENGINEERING:

1. Extract relevant features such as word frequencies, N-grams, embedding.
2. Consider additional features like sender information, timestamps, and user behavior.

IV. MODEL SELECTION:

1. Naive Bayes.
2. Support Vector Machines.
3. Decision Trees.
4. Neural Networks.

V. HANDLING IMBALANCED DATA:

1. Address class imbalance issues by using techniques like oversampling, under sampling, or synthesize data generation.

VI. TEXT REPRESENTATION:

1. Experiment with different text representations such as word embedding, contextual embedding.

VII.REAL TIME UPDATES:

1. Implement mechanisms for retraining the model periodically.

VIII.THRESHOLD TRAINING:

1. Adjust the classification threshold to balance between precision and recall based on specific needs of application.