3107- JAWAHAR ENGINEERING COLLEGE

Subject Title- AI 101- Artificial Intelligence
Project Title-Building a Smarter AI-Powered Spam Classifier: Phase-1

S.No.	NaanMudhalvan ID	Position	Name	Department
		Faculty Mentor	V. Nivaskumar	
1.	au310721106001	Team Head	Junia Susheela Shalom	ECE
2.	au310721205001	Team member	M. Hariragavan	IT
3.	au310721205002	Team Member	S. Muthunivas Pandi	IT

Project Title-Building a Smarter AI-Powered Spam Classifier

Aim- The objective of this project is to develop a machine learning model that can accurately distinguish between spam and non-spam messages in emails or text-messages based on a set of features such as pattern and probability of different words occurring in spam and ham mail.

Phases of creating an AI Powered Spam Classifier:

1. Data Collection:

> Download a dataset containing labeled examples of spam and non-spam messages from Kaggle.
> Upload the csv file into your Jupyter notebook for further analysis.

2. Data Preprocessing

The text is cleaned and preprocessed. This involves the following:
> Removing special characters.
> Converting text to lowercase.
> Tokenizing the text to individual words.
> Removing stop words and punctuation.
> Lemmatization that involves grouping together different inflected forms of the same word.

3. Feature Extraction

> The tokenized words are converted to numerical features using techniques like TF-IDF (Term Frequency –Inverse Frequency Document Frequency)
> It involves removing specific noisy and less informative terms to enhance the performance of the classifier and decrease feature space dimensionality.

4. Model Selection

> We can experiment with various machine learning algorithms such as Naïve Bayes, Support Vector Machines and more advanced techniques like deep learning using neural networks.
> For this project we implement the Naïve Bayes algorithm.
Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes Theorem and used for solving classification problems.
> It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

5. Evaluation

> The model's performance is measured using metrics like accuracy, precision, recall, and F1-score, Area under Curve, Confusion Matrix and Mean Square Error.
> Model Evaluation is important to assess the efficacy of a model during initial research phases, and it also helps in model monitoring.

6. Iterative Improvement

The model and the experiment are fine-tuned with hyperparameters to improve its accuracy.
The model can be improved by the following:
> Using more training data.
>Reducing or increasing model complexity.
> Applying regularization methods, like Ridge and Lasso regularization.
> In case of Neural networks, adding more dropout layers and early stopping.
> Training the model for more epochs.