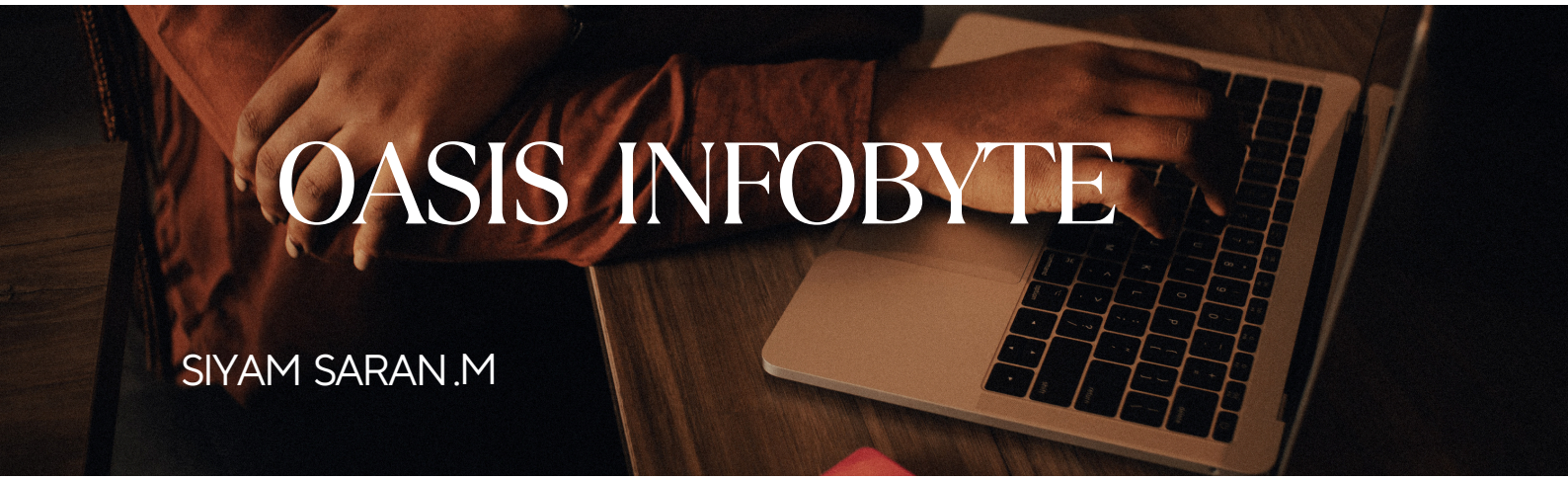

Task Title:	E-MAIL SPAM
Date:	Aug 30, 2024
Data Set:	spamdetect

DESCRIPTION

Creating an email spam detection model involves developing a program that can automatically classify incoming emails as "spam" or "not spam" based on their content and other characteristics.

Objective:

The goal of this program is to accurately classify emails as spam or non-spam (also known as "ham") using machine learning techniques. The model will analyze the content of the emails and other metadata to identify patterns commonly associated with spam emails.

- Data Collection
- Preprocessing
- Model Training
- Model Evaluation

Challenges:

- **Evolving Nature of Spam:** Spammers constantly change tactics, making it challenging to keep the model up-to-date.
- **Imbalanced Data:** Spam emails are often less frequent than non-spam, leading to imbalanced datasets that can bias the model.
- **False Positives:** Minimizing false positives is crucial as misclassifying important emails as spam can lead to significant inconvenience for users.

	FEATURES
DATA COLLECTION	The model requires a labeled dataset containing a large number of emails marked as either spam or non-spam.
DATA PREPROCESSING	The email text is cleaned by removing punctuation, stop words, and special characters.
MODEL EVALUATION	The selected algorithm is trained on the labeled dataset using the extracted features. The dataset is typically split into training and testing sets to evaluate the model's performance.
EVALUATION	The model evaluation metrics such as accuracy calculation, precision etc are evaluated.
TEST	Test the model with the 20% testing data