



# Oasis Info Bytes : Data Science Internship

## **E-mail Spam Detection with Machine Learning**

**Description:** The goal of this project is to develop a model that can accurately detect spam email messages. Utilizing a dataset that includes the subject lines of emails, each entry is classified as either spam or non-spam based on the content. To achieve this, I have employed the Logistic Regression algorithm, a widely-used method in supervised learning, to train the model on identifying patterns that distinguish spam emails from legitimate ones.

**Dataset :** [Email spam](#)

### **Key Features:**

- **Dataset :** Collecting the required dataset is a crucial initial step in any data analysis project. For my project, I utilized a dataset from Kaggle to ensure comprehensive and high-quality data. Since the dataset contains only two features, it's straightforward and easy to understand, making data visualization unnecessary in this case.
- **Feature Extraction :** Feature extraction involves identifying dependent and independent variable x and y.
- **Model Implementation:** Implement the model using Logistic Regression from the library sklearn
- **Model Training :** Split the data into training and testing data. Training data(80%) is used to train the model. Test data(20%) is used to evaluate the accuracy of the

trained model

- **Evaluate the model :** Calculate the statistics features such as accuracy, precision, recall, F1 score
- **Prediction :** With the help of testing data we can predict the outcomes for new values using the evaluated model.