A computer with email icons on the screen

Description automatically generated

faculty of computer and information science

Ain Shams University

زيد هاني محمد صلاح الدين 2022170175

أحمد عمرو الحسيني أمين 2022170029

أحمد محمد مجاهد 2022170038

ارساني عادل قسطندي تاوضروس 2022170050

احمد محمد عبدالرءوف ابراهيم 2022170036

على اشرف ابراهيم سيد 2022170256

*OVERVIEW*

Classify emails as "spam" or "not spam" based on text

features like word frequencies, presence of spam keywords, and sender

information. SVM excels at handling high-dimensional data often

present in text classification tasks.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1- Preprocessing:

First thing we explore through the dataset to see its feature,

call function head() and tail() to see data at each row, seeing basic information about each column , for numerical columns we need description for them so we call function describe() , to know ( count , mean std, min , max…..etc).

Then we need to check nulls if exist and manipulate with them by dropping cells which contains nulls or replace its value with median value if numeric or mod if categorial.

After that we need to drop duplicates rows if exist because it makes reduncency in the data.