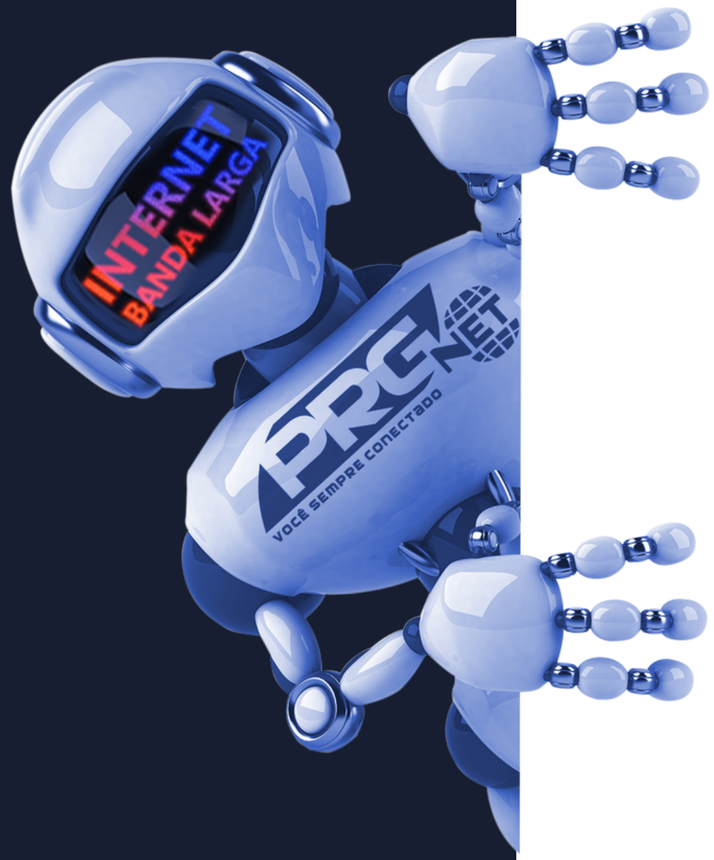
ΔDAMA SCIENCE AND TECHNOLOGY UNIVERSITY

School of Electrical Engineering and Computing

Department of Electronics and Communication Engineering

Natural Language processing (CSE5321)

Spam Message Detection



CONTENTS

01 INTRODUCTION

What is Spam Message ?

Spam is any kind of unwanted, unsolicited digital communication that gets sent out in bulk.

Often spam is sent via email, but it can also be distributed via text messages, phone calls, or social media.

01 INTRODUCTION

Spam Detection A machine-learning model with a set of examples of spam and ham messages and let it find the relevant patterns that separate the two different categories.

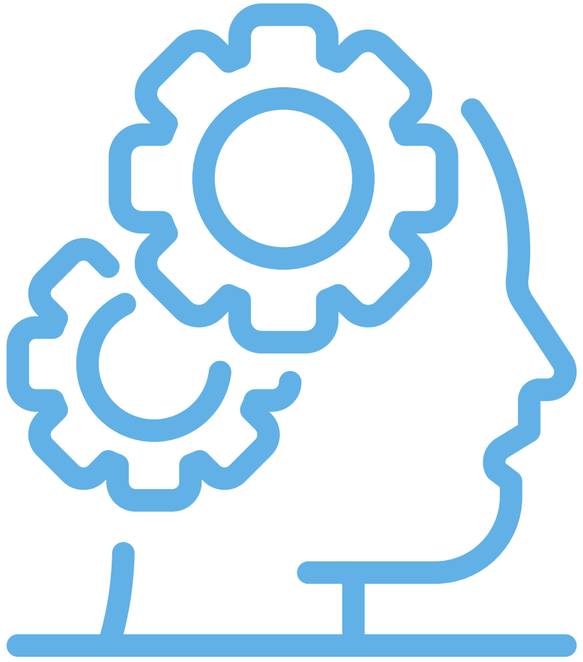
02 APPROCH

A general approach to detecting a spam message is using supervised learning.

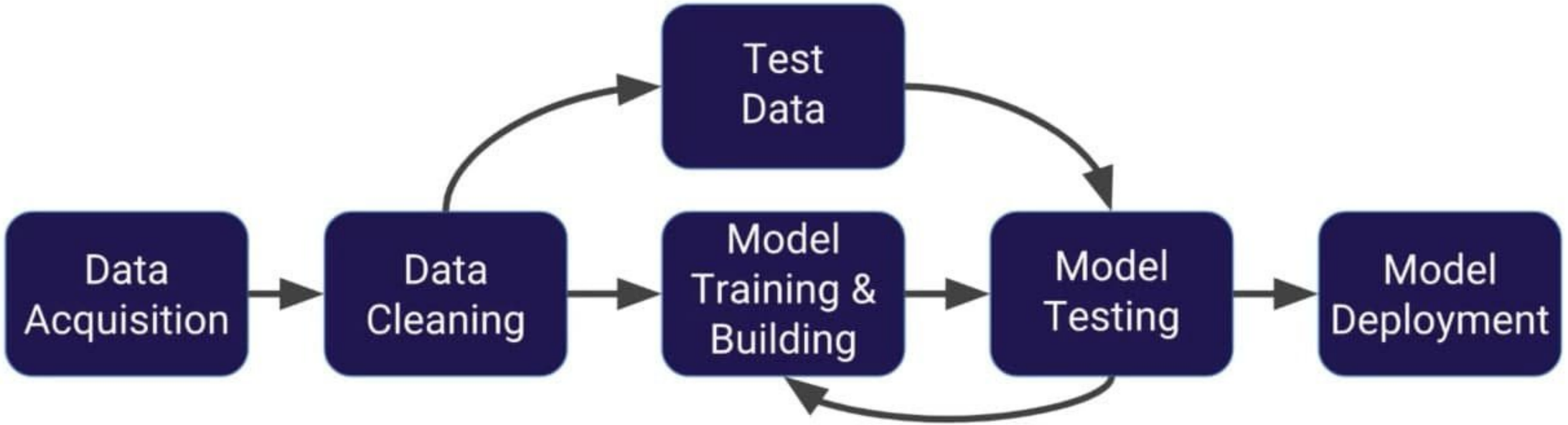
Supervised Learning An algorithm trained on input data that has been labeled for a particular output.

The algorithm learns by comparing its actual output with the correct output to find errors.

Text classification refers to labeling sentences or documents, such as email spam classification and sentiment analysis.



Overall Process



02 APPROCH

Data acquisition is the process of sampling signals that measure real world physical conditions and converting the resulting samples into digital numeric values that can be manipulated by a computer.

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

A data set (or dataset) is a collection of data.

02 APPROCH

Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples. In supervised learning, a machine learning algorithm builds a model by examining many examples and attempting to find a model that minimizes loss

Splitting up into Training and Test sets are common best practices. This allows to tune various parameters of the algorithm without making judgements that specifically conform to training data.

Model Training Algorithms

Linear SVC method applies a linear kernel function to perform classification and it performs well with a large number of samples.

Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable.

Model Training Algorithms

Model testing is referred to as the process where the performance of a fully trained model is evaluated on a testing set.

A metric is a quantitative measurement of data, in relation to what you are actually measuring. Your data point may be just a number. Common Metrics are Accuracy, recall, precision,fl-score.

The application of a model for prediction using a new data refer to as **Deployment**.

03 APPLICATION

- Messaging Platforms
- Mailing Service

Thank you!