

Naïve Bayes Classifier Model for Detecting Spam Mails

Shrawan Kumar¹ · Kavita Gupta² · Manya Gupta¹

Received: 22 March 2023 / Revised: 23 May 2023 / Accepted: 29 May 2023 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

In this paper, the machine learning algorithm Naive Bayes Classifier is applied to the Kaggle spam mails dataset to classify the emails in our inbox as spam or ham. The dataset is made up of two main attributes: type and text. The target variable "Type" has two factors: ham and spam. The text variable contains the text messages that will be classified as spam or ham. The results are obtained by employing two different Laplace values. It is up to the decision maker to select error tolerance in ham and spam messages derived from two different Laplace values. Computing software R is used for data analysis.

Keywords Naïve Bayes Classifier · Machine learning · Predictive analytics · Artificial intelligence · Supervised machine learning

JEL Classification C38 · C65

1 Introduction

Data science is the study of data with the goal of gaining important business insights. It is a multidisciplinary method for analysing massive volumes of data that integrates ideas and techniques from the domains of mathematics, statistics, artificial intelligence, and computer engineering. Through data analysis, one can answer questions like what happened, why it happened, what will happen in future, and what can be done with

Kavita Gupta gupta_kavita31@yahoo.com
Shrawan Kumar shrawan.kmc@gmail.com
Manya Gupta manyagupta020403@gmail.com

¹ Department of Statistics, Kirori Mal College, University of Delhi, Delhi, India

² Department of Mathematics, Kirori Mal College, University of Delhi, Delhi, India