

NAME: KSHITIJ VINOD SALI

CLASS & DIV: TE - A

Roll No.: 35059

ML

UA-04

Title: Assignment on Improving Performance of Classifier Models.

Problem Statement: A SMS unsolicited mail (every now and then known as cell smartphone junk mail) is any junk message brought to a cellular phone as textual content messaging via the Short Message Service (SMS). Use probabilistic approach (Naive Bayes Classifier/ Bayesian Network) to implement SMS spam filtering system. SMS messages are categorized as SPAM or HAM using features like length of message, word depend, unique keywords etc. Download dataset from: <https://archive.ics.uci.edu/ml/datasets/sms+spam+collection>.

This dataset is composed by just one text file, while each line has the correct class followed by the raw message. a) Apply data pre-processing (Label Encoding, Data Transformation...) techniques if necessary b) Perform data-preparation (Train-Test-Split). c) Apply at least two Machine Learning Algorithm and evaluate models. d) Apply cross-validation & evaluate models & compare performance. e) Apply hyper parameter tuning & evaluate models & compare performance.

Course Outcome:

C314448.1: Implement different supervised & unsuper-

vised, learning algorithm.

C314448.2: Evaluate performance of machine learning algorithms for real-world applications.

Pre-requisites: Python, Discrete Structure.

Practical Assignment Objective: Understand the basics fundamental elements of machine learning to work on machine learning classification algorithm.

Requirements: Python programming, Jupiter Notebook, Google colab.

Theory:

1) Explain classifier models

→ Classifier models are algorithms used to categorize data points into predefined labels or classes. These models learn from labeled training data & make predictions for new, unseen data and make predictions for new unseen data. Common classifier models include decision trees, support vector machines, k-nearest neighbors, Naïve Bayes.

They are used in various applications such as spam detection, medical diagnosis & image classification evaluating their performance through metrics like accuracy, precision, recall & F1-score.

2) What is a Naïve Bayes classifier?

→ The Naïve Bayes classifier is a probabilistic model based on Bayes' theorem, assuming that

features are conditionally independent given the class label. Despite this 'Naive' assumption it performs well in many real-world situations, especially for large datasets. It is commonly used in text classification tasks like spam detection & sentiment analysis due to its simplicity & efficiency & in handling categorical data.

3) How does the Naive Bayes classifier work?

→ Naive Bayes works by calculating the posterior probability of a class given the feature values, using Baye's theorem:

$$P(c|x) = \frac{P(x/c) \cdot P(c)}{P(x)}$$

The model assumes that features are independent of each other, given the class label. It multiplies the likelihoods of each feature & chooses the class with highest posterior probability. This makes it fast & effective for classification.

Algorithm:

- 1) Data preprocessing step.
- 2) Fitting Naive Bayes to the training set.
- 3) Predicting the test result.
- 4) Test accuracy of the result.
- 5) Visualizing the test set result.

Input: We are giving dataset SMS-SPAM-COLLECTION as an input to classifier model.

Output: Students are able to design classification model & predicting whether a student will get permission or not.

Inference: Hence, we have studied working of Naïve Bayes classifier algorithm in details.