

Implementation of Support Vector Machine on Text-based GERD Detection by using Drug Review Content

Asty Nabilah ¹Izzaturrahmah, Fhira Nhita, Isman Kurniawan

School of Computing, Telkom University, Bandung
astynabilah@students.telkomuniversity.ac.id, ²pembimbing1@telkomuniversity.ac.id,
ismankrn@telkomuniversity.ac.id

Abstract

GERD or Gastroesophageal Reflux Disease is a situation when the reflux of stomach contents leads to unpleasant symptoms and/or complications. The prevalence range of GERD is approximately 18.1% to 27.8% in North America, 8.8% to 25.9% in Europe, 2.5% to 7.8% in East Asia, 8.7% to 33.1% in the Middle East, 11.6% in Australia, and 23.0% in South America. The numbers may seem small, but GERD will lead to several complications including esophagitis, peptic stricture, and Barrett's esophagus if left untreated. The most common diagnostic test for the assessment of GERD along with its possible complications is the upper gastrointestinal endoscopy, or esophagogastroduodenoscopy (EGD). However, endoscopy has several risks. Disease detection using machine learning can be done and is needed due to the increment in medical data, new detection, and diagnostic modalities being developed. One of the machine learning algorithms often used in text classification is Support Vector Machine (SVM). This research applies SVM to do text-based classification, classifying data into two classes, namely "GERD patient" and "not GERD patient," using drug review data. The best model has 91.32% accuracy, 91% f1-score, and 91.32% AUC score with unigram as the n-gram range, and RBF with C is 1000, and gamma auto as the SVM kernel.

Keywords: GERD, SVM, GERD detection, disease detection

