Tabel Review

|  |  |
| --- | --- |
| **Bagian** | Review |
| Masalah | Keberadaan email spam pada system email. Email spam tersebut dapat berisi malware, dalam bentuk skrip atau file yang dapat dieksekusi, atau berisi tautan terselubung yang mengarah ke situs web phishing |
| Solusi | Menggunakan algoritma SVM-NB(Support Vector Machine based Naive Bayes) dalam upaya menklasifikasikan email. SVM digunakan untuk memangkas sampel yang buruk dan Naive Bayes dijadikan model untuk mengklasifikasikan email spam. |
| Metode | Konten dari dataset DATAMALL pertama-tama akan diproses oleh modul Text Participle Modul ini memecah kalimat bahasa Mandarin reguler menjadi beberapa unit yang bermakna seperti kata atau frasa, sesuai aturan tertentu. Unit yang dihasilkan dari modul participle teks kemudian akan diberi label oleh modul Text Annotation. Modul Text Annotation menandai setiap kata / segmen berdasarkan sifatnya, misalnya: Kata benda, kata kerja, kata sifat, atau sesuatu yang lain. Selain itu, modul ini akan menghapus kata-kata yang kurang penting, misalnya preposisi dan kata kunci tambahan. Setelah semua kata dianotasi, vektor dapat dibangun untuk memuat semua segmen yang dihasilkan dari modul Text Annotation. Selanjutnya, Modul Feature Selector memilih beberapa dimensi dari vektor untuk membentuk vektor fitur. Vektor fitur sebagian besar mencerminkan makna konten dalam email. Setelah itu, modul feature vector weight calculation akan mengidentifikasi skema berdasarkan bobot yang sesuai. Sebagai contoh, istilah istilah frekuensi dokumen terbalik (TF-IDF) metode dapat digunakan untuk bobot istilah tertentu. Metode IF-IDF pada dasarnya memberi bobot pada setiap istilah, berdasarkan frekuensi kemunculannya. Setelah preprocessing, vektor fitur akan digunakan untuk mewakili email dan dimasukkan ke dalam modul klasifikasi. Modul klasifikasi menentukan apakah vektor fitur input menunjukkan email spam atau non-spam. |
| Pengujian |  |
| Hasil |  |