**TUGAS KLASIFIKASI**



**Disusun Oleh :**

**Hermawan Adi N. (14.01.53.0005)**

**UNIVERSITAS STIKUBANK SEMARANG**

**FALKUTAS TEKNOLOGI INFORMASI**

**TEKNIK INFORMATIKA**

**2018**

**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang Permasalahan**

Banyak instansi yang bergerak dalam penyaluran informasi masyarakat atau berita yang pada awalnya menyampaikan berita melalui media Televisi, Surat Kabar, Majalah atau Radio sudah mulai menggunakan sistem berbasis web untuk menyampaikan beritanya secara up to date.

Pada umumnya berita yang disampaikan dalam portal tersebut terdiri dari beberapa kategori seperti berita politik, olahraga, ekonomi dan lain sebagainya. Namun, dalam membagi berita ke dalam kategori-kategori tersebut untuk saat ini masih dilakukan secara manual. Hal ini sangat merepotkan apabila berita yang ingin diunggah berjumlah banyak. Oleh karena itu perlu adanya sistem yang bisa mengklasifikasikan berita secara otomatis.

Text mining merupakan metode klasifikasi yang merupakan variasi dari data mining berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar.

**1.2 Rumusan Masalah**

Berdasarkan latar belakang permasalahan yang ada di atas, maka adapun rumusan permasalahannya adalah bagaimana mengklasifikasikan berita secara otomatis.

**BAB II**

**TINJAUAN PUSTAKA**

Pada bab ini akan dipaparkan penjelasan tentang tinjauan pustaka yang dipakai dalam penjelasan mengenai Klasifikasi suatu dokumen beserta implementasinya menggunakan bahasa R.

* 1. **Klasifikasi**

Klasifikasi adalah salah satu teknik machine learning. Teknik ini termasuk ke dalam tipe supervised learning. Istilah dari klasifikasi didapat dari tujuan utama teknik ini untuk memprediksikan suatu kategori dari input data.

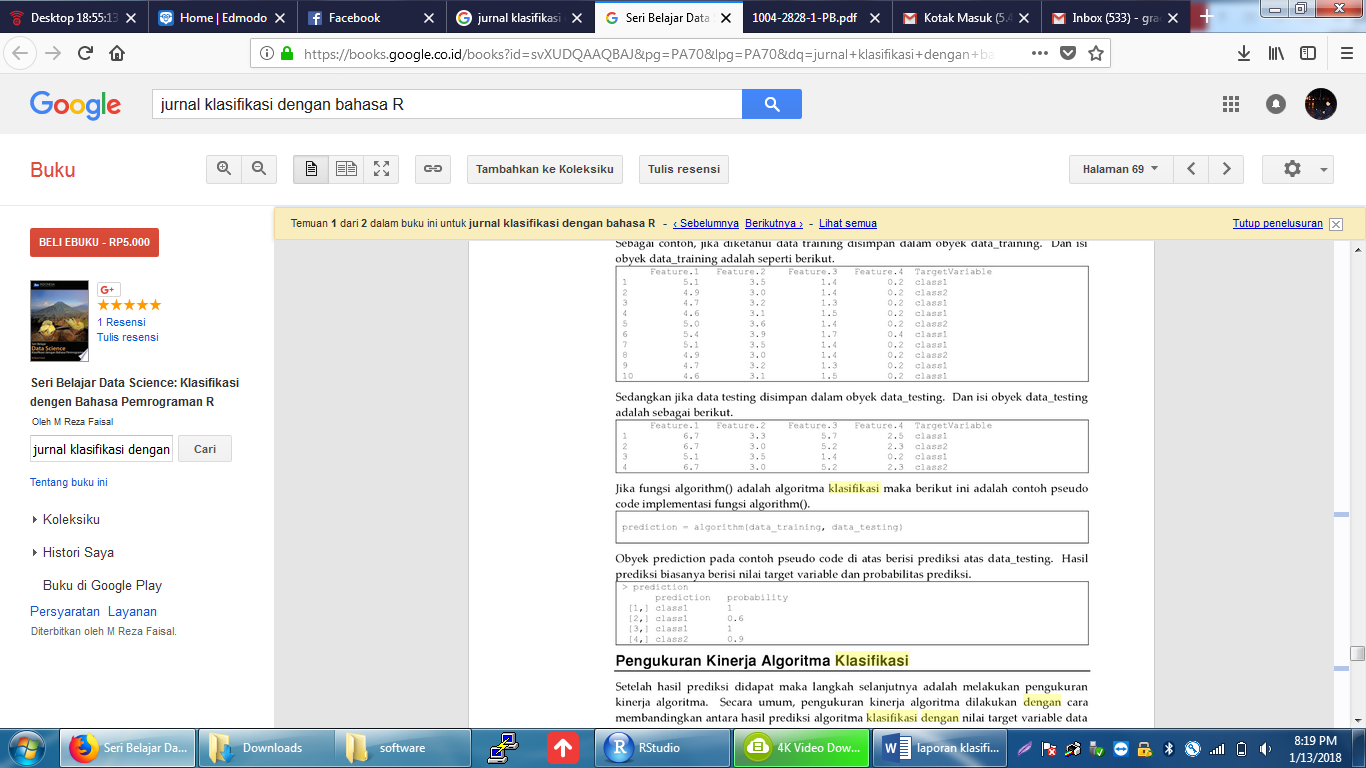
Sebelum proses prediksi dilakukan, langkah pertama yang dilakukan adalah proses pembelajaran. Proses pembelajaran memerlukan sebuah data, data yang digunakan adalah data latih / data training. Untuk data yang digunakan pada saat proses prediksi disebut dengan data uji / data testing.

* 1. **Pembagian Data**

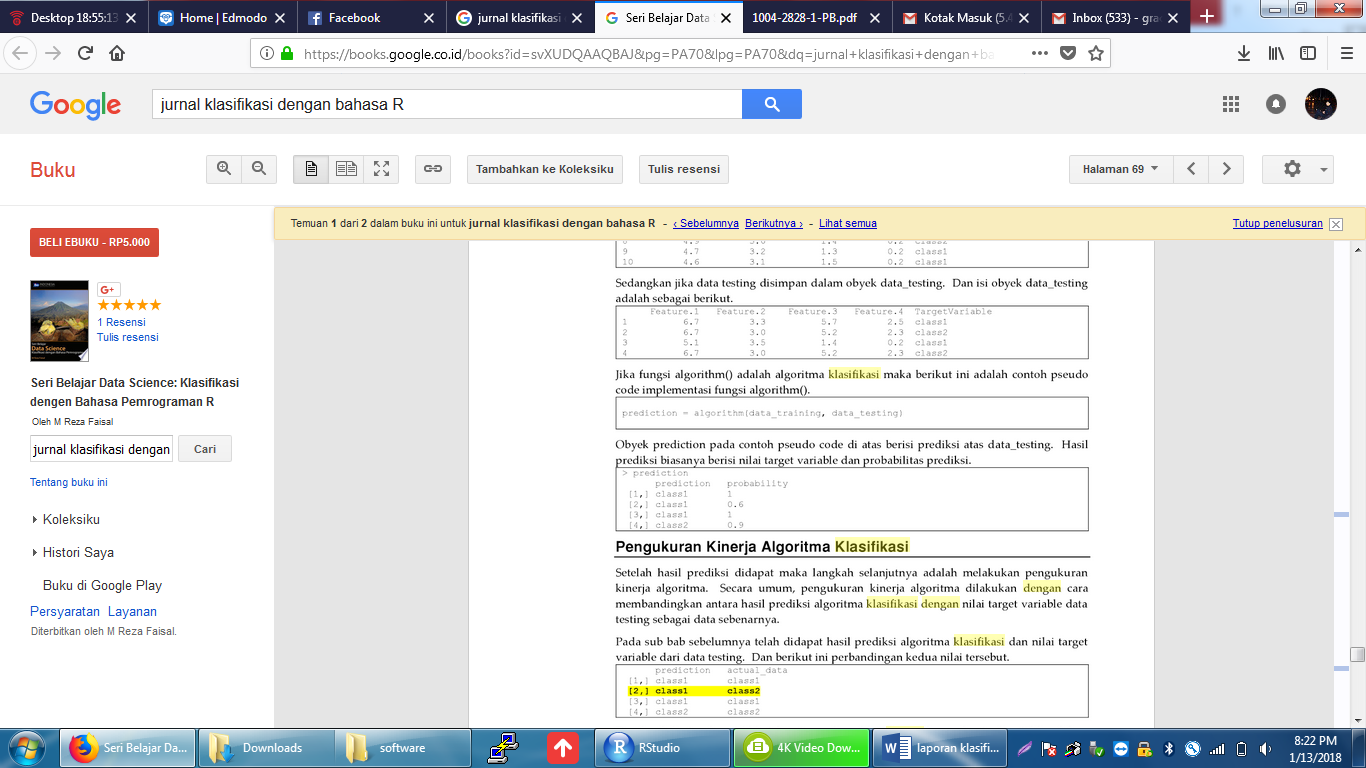
Algoritma klasifikasi ini dapat melakukan suatu prediksi setelah proses training yang dilakukan terlebih dahulu. Proses training sendiri memerlukan data yang telah tersedia. Setelah proses pembelajaran selesai maka data baru bisa diprediksi.

* 1. **Implementasi Algoritma Klasifikasi**

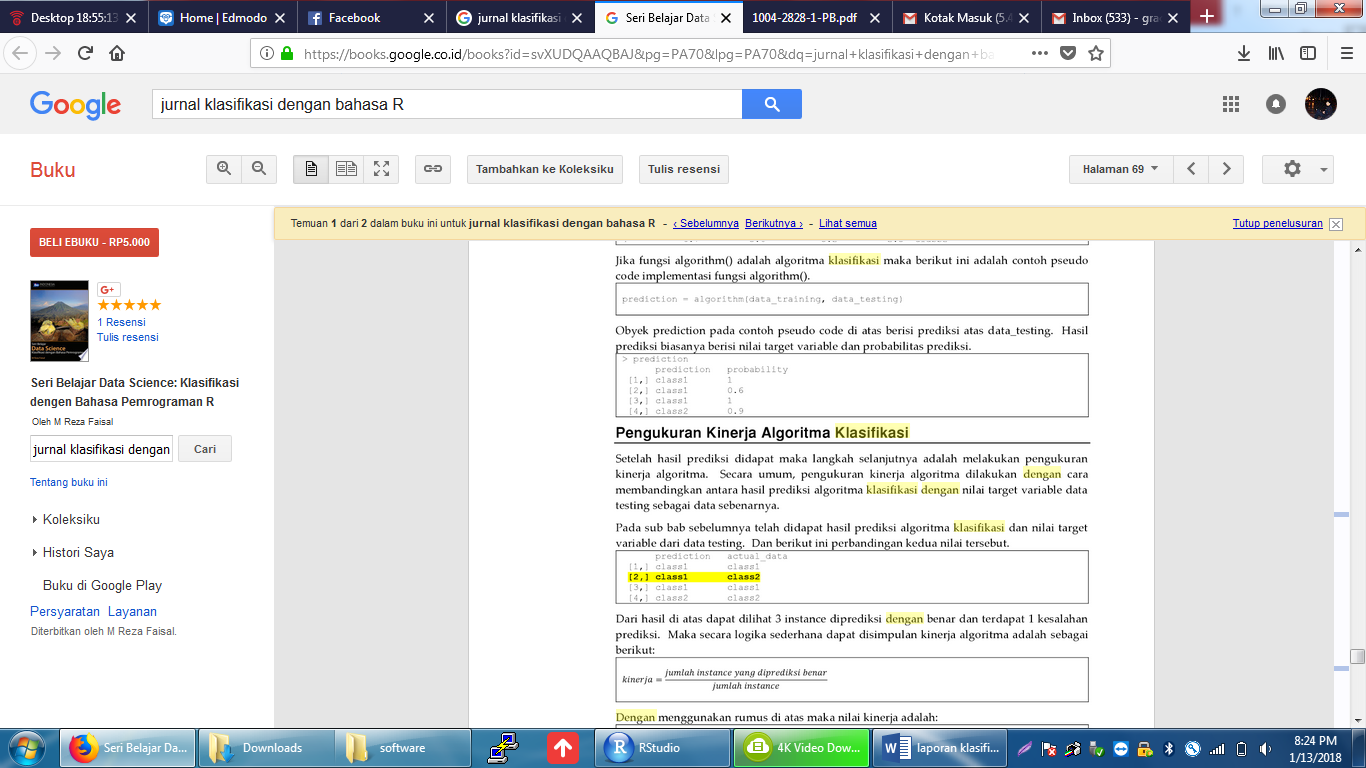
Setelah data dibagi menjadi 2 bagian yaitu data training dan data testing, maka tahap selanjutnya adalah implementasi algoritma klasifikasi. Pada tahap ini, data training akan digunakan oleh algoritma untuk belajar. Setelah proses pembelajaran selesai, maka data testing akan digunakan oleh algoritma klasifikasi. Output dari algoritma itu sendiri adalah hasil prediksi.

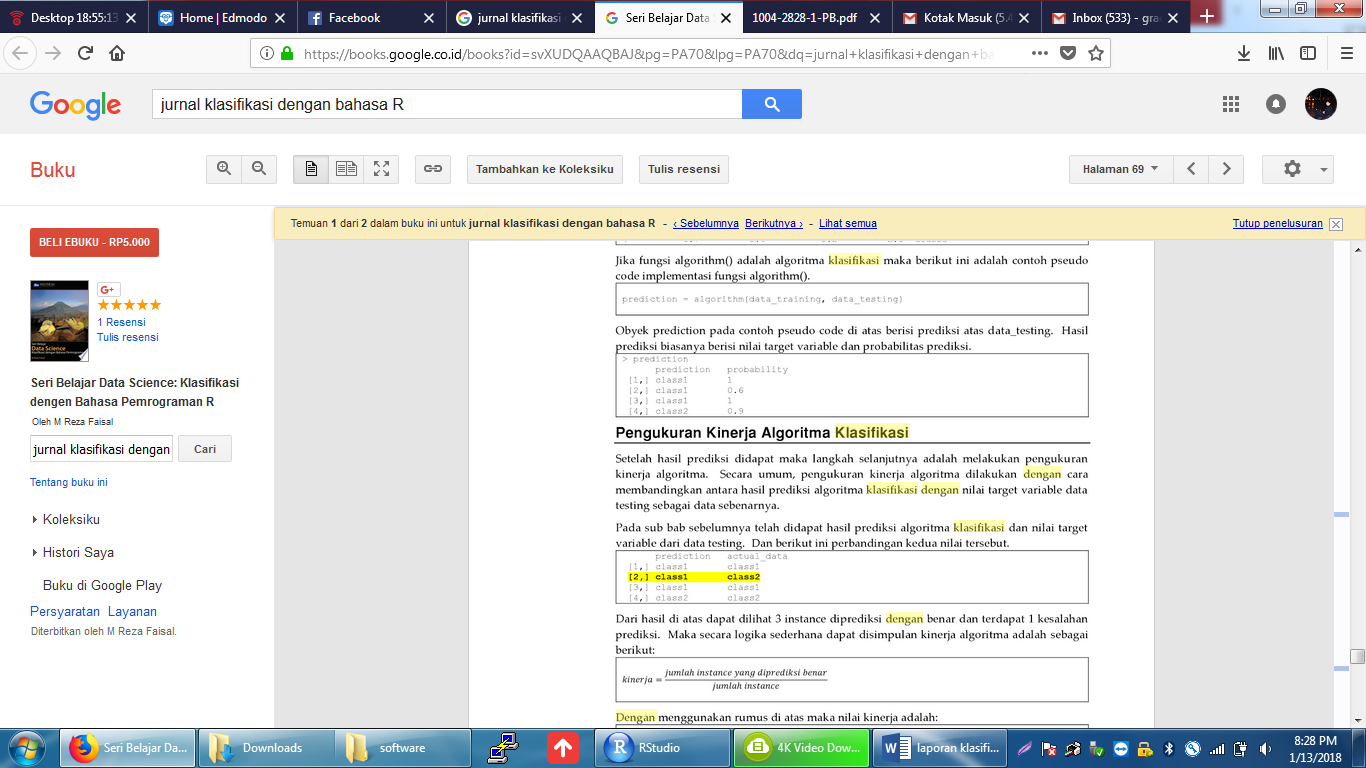
Sebagai contoh, apabila data training diketahui disimpan dalam obyek data\_training, dan isi obyek data\_training tersebut terdapat pada gambar di bawah ini :

Sedangkan jika data testing disimpan dalam obyek data\_testing, dan obyek data\_testing terdapat pada gambar di bawah ini :



Jika fungsi algortima () adalah algoritma klasifikasi, maka berikut ini adalah contoh pseudo code implementasi fungsi algoritma ().



Prediksi obyek pada contoh pseudo code di atas berisi prediksi atas data\_testing. Hasil prediksi biasanya berisi nilai target variable dan probabilitas prediksi.

**BAB III**

**METODE PENELITIAN**

Penelitian ini dilakukan dengan cara melakukan survey dari kertas hasil penelitian yang terbit di jurnal nasional ataupun yang dipublikasikan dalam seminar nasional dengan menggunakan topik bahasan bidang klasterisasi, klasifikasi dan peringkasan teks berbahasa Indonesia. Metode – metode yang digunakan dalam proses penambangan data, baik pre-proses mapupun post-prosesnya terdapat pada kertas hasil penelitian tersebut. Survey juga melihat banyaknya data pengujian yang digunakan dengan tingkat akurasi yang didapatkan. Dalam survei ini dapat dilihat secara umum metode penulisan kertas yang dilakukan.

**BAB IV**

**HASIL DAN PEMBAHASAN**

Dalam bab ini, akan dilakukan pembahasan mengenai survey yang telah dilakukan dalam kertas yang sudah ada.

1. Pemrosesan Awal

Dalam pemrosesan awal ini dilakukan persiapan dokumen yang diap menjadi bahan penambangan data. Pada pemrosesan dokumen teks, bagian ini merupakan proses melakukan pembersihan, perapihan, pembakuan, koreksi, standarisasi bahkan modifikasi dari isi dokumen teks. Mempersiapkan data dari dokumen sumber dapat dilakukan dengan memperhatikan dalam beberapa permasalahan berikut :

1. Formulasi standar
2. Tipe data
3. Pembersihan data
4. Nilai penganti
5. Penyesuaian jumlah atribut
6. Standarisasi dan Pembersihan Data

Kertas yang disurvei tidak menyebutkan secara eksplisit metode standarisasi dokumen yang digunakan namun secara umum akan terdapat proses pemilihan sumber data.

1. Stemming

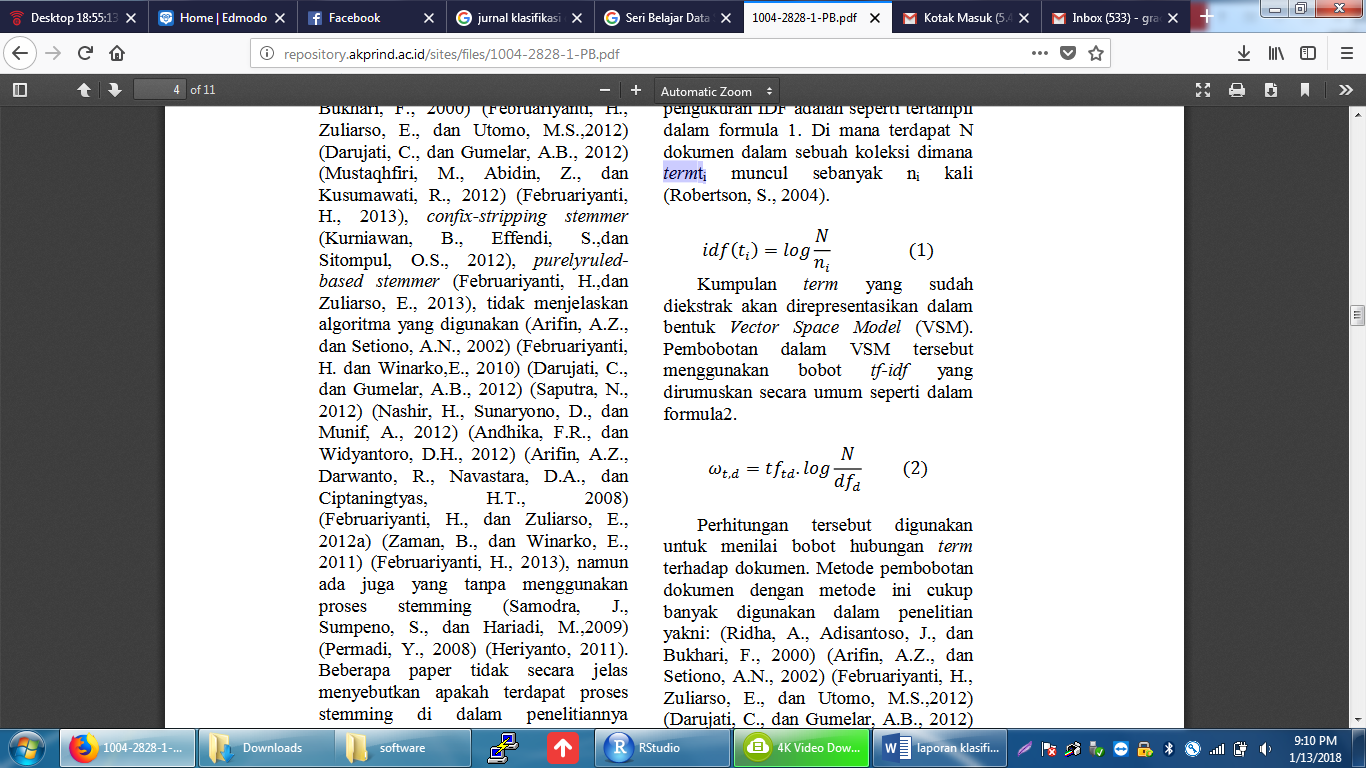
Stemming merupakan proses yang menyediakan pemetaan varian morfologi yang berbeda dari suatu katake akar katanya ( stem ).

Algoritma stemming adalah sebuah prosedur untuk mengurangi semua kata yang memiliki akar kata yang sama daam bentuk yang umum ( Lovins, J.B., 1968 ).

1. Perhitungan Bobot

Pada tahun 1972, Karen Sp”arck Jones mempublikasikan dalam Journal of Documentation sebuah paper dengan judul A Statistical interpretation of term specificity and its application in retrieval. Sebuahpengukuran dari kekhususan suatu term yang kemudian akan dikenal dengan nama inverse document frequency ( IDF ) ( Jones, K.S., 1972 ). Pengukuran tersebut didasarkan pada perhitungan frekuensi kemunculan suatu term dalam dokumen. Formula dasar dari pengukuran IDF adalah seperti tertampil dalam formula 1, di mana terdapat N dokumen dalam sebuah koleksi di mana termt*i* muncul sebanyak n*i*  kali ( Robertson, S., 2004 ).

Perhitungan tersebut digunakan untuk menilai bobot hubungan term terhadap dokumen.







Tabel 1. Hasil Data Pengujian klasifikasi

Pada pengujian tahap testing hal-hal yang dilakukan adalah dengan melakukan klasifikasi berita yang belum diketahui kategorinya. Berita yang dijadikan pengujian tahap testing berjumlah 10 berita untuk masing-masing kategori. Setelah melakukan pengujian proses testing maka berita-berita yang belum berkategori akan mendapatkan kategori yang sesuai.

**BAB V**

**KESIMPULAN**

Setelah melakukan studi literatur, perancangan, analisis, implementasi dan pengujian pengklasifikasian berita secara otomatis maka dapat disimpulkan ini sudah mampu melakukan proses klasifikasi data berita secara otomatis dan proses klasifikasi semakin akurat jika data latih yang digunakan dalam pembelajaran berjumlah banyak. Untuk penelitian berikutnya diharapkan sistem ini tidak hanya untuk mengklasifikasi berita melainkan bisa juga digunakan untuk mengklasifikasikan dokumen lain.

**DAFTAR PUSTAKA**

Bambang K, 2012. *Klasifikasi Konten Berita Dengan Metode Text Mining*, Universitas Sumatra Utara.

M Reza, TT. *Seri Belajar Data Science Klasifikasi dengan Bahasa Pemrograman R*, 68-69.