Assalamualaikum Wr Wb yang terhormat pak Randy sebagai dosen fasilitator, sebelumnya mohon izin perkenalkan nama saya Raditya Rinandyaswara dengan dosen pembimbing 1, bu Yuita Arum Sari S.Kom M.Kom, dan dosen pembimbing 2 pak Muhammad Tanzil Furqon. Mohon izin untuk mempresentasikan penelitian saya dengan judul Analisis Sentimen mengenai kuliah daring di twitter menggunakan metode naïve bayes dan pembentukan stopword dengan term based random sampling.

**==========================**

PENDAHULUAN

**==========================**

Pada akhir tahun 2019 dunia dikejutkan dengan virus corona yang menyerang sistem pernapasan manusia. Sehingga pemerintah mengadakan new normal untuk mempercepat covid-19 yang salah satu efek sampingnya adalah terjadinya kuliah daring. Media sosial menjadi salah satu sarana masyarakat menuliskan opininya mengenai kuliah daring. Oleh karena itu analisis sentimen diperlukan untuk melihat bagaimana masyarakat menanggapi protokol baru tersebut. Peneliti akan membangun sistem analisis sentimen menggunakan Naive Bayes serta Term Based Random Sampling untuk pembuatan daftar stopword.

**==========================**

Rumusan Masalah

1. Bagaimana pengaruh parameter *X*, *Y* dan *L* pada *Term Based Random Sampling* terhadap hasil pembentukan *stopword*?
2. Bagaimana hasil evaluasi dari pembentukan *stopword Term Based Random Sampling* pada analisis sentimen dengan *Naïve Bayes*?
3. Bagaimana perbandingan hasil evaluasi dari pembentukan *stopword* dengan *Term Based Random Sampling* dengan *stopword* Tala pada analisis sentimen dengan *Naïve* *Bayes*?

Batasan Masalah

1. Hanya menggunakan opini pengguna Twitter mengenai Kuliah Daring.
2. Algoritma yang digunakan hanya Naïve Bayes Classifier tidak membandingkan dengan algoritma lain.
3. Hasil klasifikasi sentimen hanya dibagi menjadi tiga kelas yaitu positif, netral, dan negatif.
4. Tweet yang merupakan data hanya tweet yang berbahasa Indonesia.
5. Jumlah data yang digunakan sebanyak 300 data.
6. Sistem yang dibuat hanya dapat menangani data yang seimbang setiap kelasnya.

**==========================**

**DASAR TEORI**

**==========================**

Adapun penelitian sebelumnya yang mendukung yaitu dari

Rahman dengan judul Pengelompokan Berita Online Menggunakan Multinomial Naïve Bayes

* Metode Klasifkasi dengan Multinomial Naïve Bayes
* Stopword yang digunakan adalah Stopword Tala
* Pembobotan dengan TF - IDF
* Penelitian difokuskan terhadap pembobotan kata

Hasil akurasi menggunakan Multinomial Naïve Bayes TFIDF menunjukan hasil akurasi akhir sebesar 94,29% yaitu lebih baik dibandingkan Multinomial Naïve Bayes DF-Threshold yaitu 92,38%.

Sa’rony dengan judul Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan Algoritme Term Based Random Sampling dan Metode Klasifikasi Naive Bayes.

* Metode Klasifkasi dengan Multinomial Naive Bayes
* Stopword yang digunakan adalah Stopword Term Based Random Sampling
* Pembobotan dengan Raw TF
* Kelas sentimen positif dan negatif
* Penelitian difokuskan terhadap persentase stoplist atau parameter L

Hasil akurasi menggunakan Multinomial Naïve Bayes dengan stoplist 20 persen mendapatkan macroaverage akurasi sebesar 94%.

Sehingga melihat penelitian sebelumnya, penelitian ini ditentukan menggunakan Klasifikasi Multinomial Naive Bayes, pembobotan TF-IDF, pembuatan stopword dengan Term Based Random Sampling, dengan kelas dibagi menjadi negatif, netral, dan positif dan penelitian difokuskan terhadap pengaruh parameter X, Y, dan L

**==========================**

Term Based Random Sampling adalah suatu metode yang dapat digunakan untuk menghasilkan daftar stopword secara otomatis menggunakan rumus teori Kullback Leibler dalam melakukan perhitungan untuk mengukur nilai kepentingan dari suatu term.

**BACA RUMUS**

**==========================**

Pilih term acak dari keseluruhan term, lalu

Ambil dokumen yang mengandung term tersebut, setelah itu,

Cari term dari seluruh dokumen sampel atau dokumen yang diambil, lalu

Hitung bobot tiap term menggunakan Kullback-Leibler, lalu

Normalisasi bobot menggunakan MinMax agar di dalam range 0 hingga 1, lalu

Ambil sejumlah X term (dimana X adalah parameter) yang diurutkan dari bobot terendah, lalu

Lakukan proses 1 hingga 7 sebanyak Y kali (dimana Y adalah parameter) , lalu

Setelah itu hitung rata-rata keseluruhan bobot yang didapatkan setiap term, lalu

Ambil sejumlah L term (dimana L adalah parameter yang menentukan jumlah persentase stopword yang ingin digunakan).

**==========================**

Naïve Bayes yaitu algoritma supervised dengan metode probabilistik dan statistik, dalam ilustrasi dibawah ditunjukan ketika ada data baru, akan dihitung menggunakan prior setiap kelas dikalikan dengan likelihood setiap kelas dan akan dipilih probabilitas tertinggi dan akan dilabelkan sebagai kelas tersebut.

**==========================**

**BACA PPT**

**==========================**

**METODOLOGI PENELITIAN**

**==========================**

Berikut adalah 6 dari 300 data yang digunakan yang sudah dilabelkan oleh 5 pakar yang terbagi menjadi 3 kelas.

Data ini diambil dari Twitter dengan kata kunci “Kuliah Daring” yang diambil dari bulan April 2020 hingga Oktober 2020. Data yang dikumpulkan dilakukan proses normalisasi secara manual terlebih dahulu seperti perbaikan singkatan kata, kata tidak baku, dan kata typo.

**==========================**

**BACA DIAGRAM**

**==========================**

**PERANCANGAN PENGUJIAN**

**==========================**

Berikut adalah tabel pengujian untuk menguji pengaruh parameter X, Y, L yang setiap parameternya mencoba kombinasi 10, 20, 30, 40, 50 dengan 10 fold cross validation.

**==========================**

Berikut adalah tabel pengujian untuk menguji pengaruh stopword TBRS jika dibandingkan dengan tanpa proses stopword dengan 10 fold cross validation.

**==========================**

Berikut adalah tabel pengujian untuk menguji pengaruh stopword TBRS jika dibandingkan dengan stopword Tala dengan 10 fold cross validation.

**==========================**

**EVALUASI**

**==========================**

Berikut adalah tabel 10 kombinasi tertinggi dari 125 kombinasi parameter X, Y, dan L dan didapatkan *macroaverage accuracy* tertinggi disaat X bernilai 10, Y bernilai 10, dan L bernilai 40 yaitu 75,8%.

**==========================**

Dapat dilihat dalam diagram bahwa ketika nilai X bernilai 10, hasil evaluasi lebih stabil tinggi dibandingkan dengan disaat X bernilai 50.

**==========================**

Sama halnya dengan parameter X bahwa ketika nilai Y bernilai 10, hasil evaluasi lebih stabil tinggi dibandingkan dengan disaat Y bernilai 50.

**==========================**

Dapat dilihat dalam diagram bahwa ketika nilai L bernilai 10, hasil evaluasi lebih stabil tinggi dibandingkan dengan disaat L bernilai 50.

**==========================**

Selanjutnya adalah pengaruh Term Based Random Sampling yang dibandingkan jika tidak menggunakan proses stopword removal. Didapatkan bahwa Term based random sampling mendapatkan evaluasi sedikit lebih baik dibandingkan dengan tanpa proses stopword dengan akurasi lebih besar 0,5%. Hal ini terjadi karena proses stopword removal dapat menghapus kata-kata yang tidak relevan yang dapat menghindari kesalahan klasifikasi karena terdapat kata stopword yang lebih banyak muncul di label tertentu.

**==========================**

Selanjutnya adalah pengaruh Term Based Random Sampling yang dibandingkan jika menggunakan stopword Tala. Didapatkan bahwa Term based random sampling mendapatkan evaluasi lebih baik dibandingkan dengan stopword Tala dengan akurasi lebih besar 2%. Hal ini terjadi karena stopword yang dihasilkan oleh Term Based Random Sampling adalah stopword yang dibuat sesuai berdasarkan data latih sehingga stopword yang dihasilkan lebih tepat sesuai dengan yang sistem perlukan,,,,,berbeda hal nya Tala yang bersifat statis yang daftar stopwordnya sudah tersedia.

**==========================**

Selanjutnya adalah kesimpulan dan saran

**==========================**

Dalam pencarian kombinasi parameter X, Y, dan L terbaik dari 125 kombinasi didapatkan kombinasi terbaik yakni X sebesar 10, Y sebesar 10, dan L sebesar 40 untuk analisis sentimen dengan Naïve Bayes yang mendapatkan macroaverage accuracy sebesar 75,8%, macroaverage precision sebesar 65,8%, macroaverage recall sebesar 63,6%, dan macroaverage f-measure sebesar 64,7%. Dapat disimpulkan bahwa semakin besar nilai X, Y, dan L maka semakin tinggi kemungkinannya untuk accuracy, precision, recall, dan f- measure turun.

Pengunaan metode pembentukan stopword Term Based Random Sampling untuk analisis sentimen dengan Naïve Bayes dapat diterapkan dengan baik, hal ini dapat dilihat dengan meningkatnya akurasi sistem yang dilakukan sebanyak 10-fold ketika menggunakan stopword Term Based Random Sampling sebesar 0,5% jika dibandingkan dengan tidak menggunakan proses stopword removal.

Berdasarkan pengujian perbandingan antara Tala dan stopword Term Based Random Sampling mendapatkan rata-rata akurasi dari 10 fold, stopword Term Based Random Sampling memiliki akurasi sebesar 75,8% sedangkan jika menggunakan stopword Tala adalah sebesar 73,8%. Penggunaan stopword Term Based Random Sampling terbukti dapat meningkatkan akurasi pada analisis sentimen dengan Naïve Bayes sebesar 2%.

**==========================**

Adapun saran dari penelitian ini adalah

Penelitian yang dilakukan masih memiliki banyak kekurangan yang perlu diperbaiki. Adapun saran yang dapat diberikan untuk penelitian berikutnya adalah pada tahap *preprocessing* sebaiknya dilakukan proses normalisasi kata secara otomatis untuk dapat meningkatkan akurasi sistem serta pemilihan data yang lebih baik.

**==========================**

Selanjutnya adalah demo, untuk mempermudah menjalankan algoritme, saya mengimplementasikan algoritme saya kedalam aplikasi android

Aku lebih baik kuliah daring kalau virus corona masih meningkat seperti ini, memang lebih aman kalau dirumah saja