

MENGUKUR TINGKAT KEPUASAN MAHASISWA DALAM PEMBELAJARAN DENGAN NAÏVE BAYES

Amril Mutoi Siregar, Maulana Abdur Rofik

Buana Perjuangan University, Jl ronggowaluyo, karawang 41311, Indonesia

amrilmutoi@ubpkarawang.ac.id

f16.maulanaabdurrofik@mhs.ubpkarawang.ac.id

Abstract

Mengukur kepuasan pembelajaran adalah hak mutlak bagi mahasiswa sehingga kampus wajib memperbaiki layanan baik kualitas dan sarana dan prasarana. Karena keberhasilan suatu institusi adalah di ukur dari tingkat kepuasan. Sering jadi kendala adalah bagaimana mengukur kepuasan pelanggan dengan data teks berbentuk tidak terstruktur. Sehingga konsumen kesulitan mendapatkan hasil yang baik. Mahasiswa memiliki sentiment yang berbeda beda terhadap baik pelayanan, proses pembelajaran dan sarana prasarana. Untuk mempermudah mengukur kepuasan pelanggan. Penelitian ini menerapkan teknik analisis sentimen untuk menilai kecenderungan saran terkait dengan berbagai faktor yang berkontribusi pada keberhasilan proses pembelajaran seperti metode pengajaran, suasana akademik, fasilitas dan ruang kelas, laboratorium, perpustakaan dan fasilitas kampus lainnya. Metode yang digunakan penelitian ini adalah algoritma Naive Bayes. Dalam penelitian ini mendapatkan akurasi 86,9%

Keywords: Klasifikasi, Naive Bayes, Kepuasan, Mahasiswa

1. Pendahuluan

Pendapat orientasi opini adalah bagian terpenting dalam pengambilan keputusan untuk suatu kebijakan. Keputusan yang tepat sangat dipengaruhi oleh analisis opini dari berbagai sumber yang terkait dengan pengambilan keputusan. Sebagai contoh pada dunia bisnis, penambahan produk oleh manajer produksi sangat memerlukan analisis dari review produk barang yang ada di pasaran. Contoh lain misalnya pada dunia manajemen pelayanan pendidikan di perguruan tinggi, pengukuran tentang tingkat kepuasan layanan pembelajaran dapat diukur dari opini mahasiswa tentang proses pembelajaran. Opini muncul pada berbagai situasi, misalnya yang dengan sengaja diminta oleh suatu alat peninjauan opini melalui permintaan saran dalam aktivitas kuesener, atau muncul secara alami dari suatu forum on line yang disediakan oleh situs resmi perguruan tinggi. Volume opini on line yang berupa teks bebas ini semakin hari semakin banyak dan umumnya tidak dimanfaatkan karena bentuknya yang tidak terstruktur. Meskipun mengandung informasi berharga, opini ini juga sering menggunakan bahasa informal, misalnya: “ pak Amril parah bngat ..”, atau “AC ruang B115 parah bngit..., tolong dong diperbaiki”. Di sini kata “ parah .” apakah memuat opini positif atau negative tentang dosen yang bersangkutan, sedangkan kata “parah” memuat opini negatif tentang AC. Tentu saja menangani opini yang diungkap dengan bahasa informal akan menjadi tantangan tersendiri.

Saat ini sumber opini teks yang tersedia melimpah di intranet atau internet belum sepenuhnya dapat dimanfaatkan karena belum adanya tool yang memadai. Di sisi lain, keberadaan internet dan sumber informasi on-line lainnya yang berkembang sangat pesat. Saat ini diperkirakan ada sekitar 30 trilyun web-pages terindeks di google dan 100 milyar ases per bulan (Koetsier, 2013). Data dan informasi online dari perusahaan dan organisasi pada umumnya berbentuk tidak terstruktur (Bridge,2011), terutama berbentuk teks yang mencapai 80% (Grimes,2013). Ditemukannya media social seperti Facebook (2004) dan Tweeter (2006) telah mendorong kegiatan seperti review, forum diskusi, blog, micro-blog, komentar, dan posting yang melipatgandakan keberadaan dokumen teks di internet.

Hal ini karena media sosial tersebut telah digunakan baik oleh individu maupun organisasi untuk berbagai kepentingan sharing informasi. Kondisi ledakan informasi ini semakin menyulitkan proses data mining sebagaimana jauh hari telah diprediksi

(Putten, et.al.,2002). Untuk itu pengembangan penelitian di bidang opinion mining menjadi topik sangat penting disamping topik-topik sebelumnya, yaitu data mining dan text mining.

Salah satu cabang riset yang kemudian berkembang dari situasi ledakan informasi di internet adalah sentiment analysis. Cabang ini menjadi riset yang menantang karena didalamnya terdapat akumulasi dari berbagai tantangan riset, yaitu antara lain information extraction, information summarization, document classification (Pang and Lee, 2002). Sentiment Analysis atau opinion mining adalah studi komputasional dari opini-opini orang, appraisal dan emosi melalui entitas, event dan atribut yang dimiliki (Liu, 2010). Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur, yaitu apakah pendapat yang dikemukakan dalam dokumen, kalimat atau fitur tersebut bersifat positif, negatif atau netral (Dehaff, M., 2010). Riset bidang ini akan menjadi jawaban keterbatasan mesin pencari (search engine) saat ini yang masih bersifat “fact-based” atau “information-based” untuk dikembangkan lebih kaya menjadi mesin pencari yang bersifat “opinion-based” (Pang and Lee,2008)

Aplikasi sentiment analysis untuk melakukan evaluasi kebijakan dan pengambilan keputusan menjanjikan cara yang lebih praktis dan ekonomis dibandingkan dengan metode klasik menggunakan pendekatan kuesioner. Kritik terhadap metode kuesioner sebagai metode yang lama dan mahal, disamping juga memberikan hasil yang kadang kurang dapat menangkap problem yang sebenarnya. Kuesioner dan interview dinilai lemah karena pada umumnya orang kurang suka menjawab pertanyaan survei yang kadang bertele-tele. Penggalan opini dengan cara mendengar (by listening) dinilai lebih baik dari pada dengan bertanya seperti kuesioner (by asking), karena lebih akurat mencerminkan realitas sebenarnya (Shelke, et.al.,2012). Bahkan lebih jauh sentiment analysis memungkinkan untuk menangkap emosi pemilik opini (Loia and Senatore, 2014). Contoh bagus dalam masalah ini adalah penelitian Greaves et.al. (2013) di English National Health Service website yang menangkap 6412 comment bebas dari pasien yang dirawat. Analisis tentang comment terkait dengan kebersihan, pelayanan rumah sakit dan berbagai aspek tanggung jawab rumah sakit memberikan hasil kesesuaian antara 81% sampai 89% dibandingkan dengan metode rating kuantitative yang diberikan melalui kuesioner.

Universitas Buana Perjuangan Karawang (UBP Karawang) sebagai lembaga pendidikan tinggi senantiasa ingin meningkatkan layanan dalam manajemen

pembelajaran. Untuk maksud tersebut pada setiap akhir semester bagian administrasi akademik mengadakan evaluasi layanan pembelajaran menggunakan instrumen kuesioner dengan butir-butir jawaban yang telah disediakan. Selama ini ada data kuesioner yang tidak dapat dimanfaatkan dan dianalisis yaitu data saran mahasiswa. Data ini jumlahnya mencapai ribuan saran atau lebih tepatnya opini yang berasal dari seluruh seluruh peserta dari seluruh mata kuliah. Saran/opini dapat mengenai suasana akademik, dosen, ruang kuliah, AC, OHP, atau fasilitas kampus lainnya. Dalam beberapa tahun data ini semakin menumpuk. Data ini sangat mungkin dapat mengungkap hal-hal yang belum terungkap melalui butir kuesioner yang jumlah pertanyaannya terbatas, dan diberikan berulang setiap semester.

Informasi dalam bentuk teks adalah informasi yang penting dan banyak didapatkan dari berbagai sumber seperti buku, surat kabar, situs web, ataupun pesan e-mail. Teks merupakan sebuah hamparan bahasa, baik dalam pembicaraan ataupun dalam tulisan, yang memiliki makna, bersifat praktis dan berguna untuk umum serta berhubungan dengan dunia nyata (Bolshakov & Gelbukh). Sebuah teks dapat terdiri dari hanya satu kata ataupun susunan kalimat (Carter & McCarthy). Pengambilan informasi dari teks (text mining) antara lain dapat meliputi kategorisasi teks atau dokumen, analisis sentimen (sentiment analysis), pencarian topik yang lebih spesifik (search engine), serta spam filtering. Gagasan umum text mining adalah untuk mengetahui cakupan atau topik dari permasalahan dalam teks (Maning, et al.)

Analisis sentimen adalah studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan dalam teks (Liu). Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas mempunyai arti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif. Salah satu teknik pembelajaran mesin untuk analisis sentimen adalah Naïve Bayes classifier (NBC). NBC merupakan teknik pembelajaran mesin yang berbasis probabilistik. NBC adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks (Routray, et)

Berdasarkan Latarbelakang yang ada maka dapat ditarik suatu latarbelakang yang akan di bahas adalah sebagai berikut. Bagaimanakah mengklasifikasi tingkat kepuasan mahasiswa terhadap pembelajaran dengan dosen teknik informatikan semester 4 tahun ajaran 2019/2020. Mengukur tingkat kepuasan mahasiswa dengan algoritma naïve bayes yang gunakan.

2. Tinjauan Pustaka

Kepuasan Mahasiswa

Kepuasan menurut Kotler adalah tingkat perasaan seseorang setelah membandingkan kinerja atau hasil dengan harapannya (Tjiptono, 2000: 90). Suatu pelayanan memiliki potensi untuk memenuhi atau tidak memenuhi harapan pelanggan. Suatu jasa dianggap memuaskan jika memiliki kualitas. Kualitas merupakan suatu kondisi dinamis yang berhubungan dengan jasa, proses, dan lingkungan yang memenuhi atau melebihi harapan. Penilaian kualitas pelayanan atau jasa dapat diukur dengan dua variable, yaitu jasa yang dirasakan (*perceived service*) dan jasa yang diharapkan (*expected service*). Kepuasan dapat dilihat dari kesesuaian antara harapan dengan apa yang didapat dari suatu pelayanan (Tjiptono, 2000: 52).

Kepuasan merupakan persepsi seseorang terhadap sesuatu yang telah memenuhi harapannya. Kepuasan mahasiswa terhadap pembelajaran dapat dilihat dari 5 dimensi kepuasan yaitu: *tangible*, *reliability*, *responsiveness*, *assurance*, dan *empathy*. Dimensi pertama dari kualitas pelayanan adalah *tangible*. *Tangible* merupakan dimensi fisik. Suatu jasa tidak dapat dicium, dan tidak dapat diraba, sehingga bukti fisik menjadi penting sebagai ukuran terhadap pelayanan. *Tangible* merupakan kemampuan untuk memberi fasilitas fisik kampus dan perlengkapan perkuliahan yang memadai menyangkut penampilan dosen serta sarana umum, misalnya: ketersediaan sarana prasarana. Mahasiswa akan menilai suatu kualitas pembelajaran dari segala sarana dan fasilitas yang ada, mengukur kehandalan dari pendidikan tinggi dalam memberikan pelayanan kepada mahasiswanya. Ada dua aspek dari dimensi ini, yaitu (1) kemampuan dosen untuk memberikan metode pembelajaran seperti yang dijanjikan dan (2) seberapa jauh dosen memberikan pembelajaran secara akurat. *Reliability*. Merupakan kemampuan dosen dalam memberikan pembelajaran sesuai dengan yang dijanjikan (tepat waktu), dengan segera, relevan dan akurat sehingga memuaskan mahasiswa.

Ketiga, *responsiveness* adalah dimensi kualitas pelayanan yang dinamis. *Responsiveness* merupakan kesediaan dan daya tanggap para dosen untuk membantu dan memberikan pembelajaran sesuai kebutuhan mahasiswa. Dimensi ini tampak pada situasi dimana dosen mudah ditemui untuk keperluan konsultasi. Harapan mahasiswa terhadap keakuratan pelayanan akan selalu berubah dari waktu ke waktu. Dimensi

keempat dari dimensi kualitas pelayanan yang menentukan kepuasan pelanggan adalah assurance, yaitu dimensi jaminan kualitas yang berhubungan dengan perilaku staf pengajar atau dosen dalam menanamkan rasa percaya dan keyakinan kepada para mahasiswa. Assurance mencakup kompetensi, pengetahuan, keterampilan, kesopanan. Contoh dimensi ini ditunjukkan seperti sikap dosen yang menyampaikan kuliah sesuai dengan bidang keahlian masing-masing. Dosen berusaha menambah wawasan dengan membaca, menghadiri seminar, mengikuti pelatihan, atau melakukan penelitian. Terdapat empat aspek dari dimensi assurance, yaitu keramahan, kompetensi, kredibilitas, dan keamanan.

Dimensi kepuasan yang terakhir adalah empathy. Empathy adalah sikap dosen dalam memberikan pelayanan sepenuh hati, seperti perhatian secara pribadi serta pemahaman bahwa setiap mahasiswa memiliki kemampuan dan kebutuhan yang berbeda. Sikap ini dapat ditunjukkan dengan pemahaman

Metode Naïve bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Kelebihan Naive Bayes Mudah untuk dipahami dan Hanya memerlukan pengkodean yang sederhana dan kekurangan dari metoda Naïve Bayes classifier ini adalah banyaknya celah untuk mengurangi keefektifan metoda ini dan akibatnya meloloskan dokumen ke dalam kelas tertentu padahal jelas -jelas dokumen tersebut tidak layak berada di kelas tersebut.

Persamaan Metode Naive Bayes secara umum:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad 1$$

Di mana:

H = Prediksi

X = Aktual

$P(H)$ = Peluang Prediksi

$P(X)$ = Peluang Aktual

$P(X|H)$ = Peluang kelas Aktual untuk kelas Prediksi

Klasifikasi

Klasifikasi merupakan salah satu topik utama dalam data mining atau machine learning. Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target. Sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam supervised learning atau pembelajaran yang diawasi. Maksud dari pembelajaran yang diawasi adalah data label atau target ikut berperan sebagai ‘supervisor’ atau ‘guru’ yang mengawasi proses pembelajaran dalam mencapai tingkat akurasi atau presisi tertentu.

Contoh klasifikasi hewan dengan label atau target adalah reptile (reptile), serangga (insect), ikan (fish), burung (bird) dan mamalia (mammal), dan menggunakan fitur-fitur tertentu, seperti mempunyai telinga, dapat berenang, beranak, bertelur dan sebagainya, sebagai indikator-indikator suatu hewan masuk dalam kelompok hewan tertentu.

Beberapa metode standar dapat digunakan untuk menyelesaikan masalah klasifikasi yaitu: Backpropagation neural network, support vector classification (SVC), extreme learning machine (ELM), K-NN, Naïve Bayes dan masih banyak lagi.

Confusion matrik

Pengukuran terhadap kinerja suatu sistem klasifikasi merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya [1]

Berdasarkan jumlah keluaran kelasnya, sistem klasifikasi dapat dibagi menjadi 4 (empat) jenis yaitu klasifikasi binary, multi-class, multi-label dan hierarchical [2]. Pada klasifikasi binary, data masukan dikelompokkan ke dalam salah satu dari dua kelas. Jenis klasifikasi ini merupakan bentuk klasifikasi yang paling sederhana dan banyak digunakan. Contoh penggunaannya antara lain dalam sistem yang melakukan

deteksi orang atau bukan, sistem deteksi kendaraan atau bukan, dan sistem deteksi pergerakan atau bukan.

Sementara itu, pada bentuk klasifikasi multi-class, data masukan diklasifikasikan menjadi beberapa kelas. Sebagai contoh sistem yang dapat mengklasifikasikan jenis kendaraan seperti sepeda, sepeda motor, mobil, bus, truk, dan sebagainya. Bentuk klasifikasi multi-label pada dasarnya sama dengan multi-class dimana data dikelompokkan menjadi beberapa kelas, namun pada klasifikasi multi-label, data dapat dimasukkan dalam beberapa kelas sekaligus. Bentuk klasifikasi yang terakhir adalah hierarchical. Data masukan dikelompokkan menjadi beberapa kelas, namun kelas tersebut dapat dikelompokkan kembali menjadi kelas-kelas yang lebih sederhana secara hirarkis. Contohnya dalam penelitian ini, arah pergerakan dikelompokkan menjadi 12 arah pergerakan yang tentunya dapat disederhanakan menjadi 4 arah.

Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Nilai True Negative (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan False Positive (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, True Positive (TP) merupakan data positif yang terdeteksi benar. False Negative (FN) merupakan kebalikan dari True Positive, sehingga data positif, namun terdeteksi sebagai data negatif. Pada jenis klasifikasi binary yang hanya memiliki 2 keluaran kelas, confusion matrix dapat disajikan seperti pada Tabel 1 [2]

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (True Positive)	FN (False Negative)
Negatif	FP (False Positive)	TN (True Negative)

Kelas Terklasifikasi Positif Terklasifikasi Negatif

Positif TP (True Positive) FN (False Negative)

Negatif FP (False Positive) TN (True Negative)

Berdasarkan nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 1. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 2. Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 3.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad 2$$

$$Presisi = \frac{TP}{FP+TP} * 100\% \quad 3$$

$$Recall = \frac{TP}{FN+TP} * 100\% \quad 4$$

dimana:

TP adalah True Positive, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.

TN adalah True Negative, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.

FN adalah False Negative, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.

FP adalah False Positive, yaitu jumlah data positif namun terklasifikasi salah oleh sistem

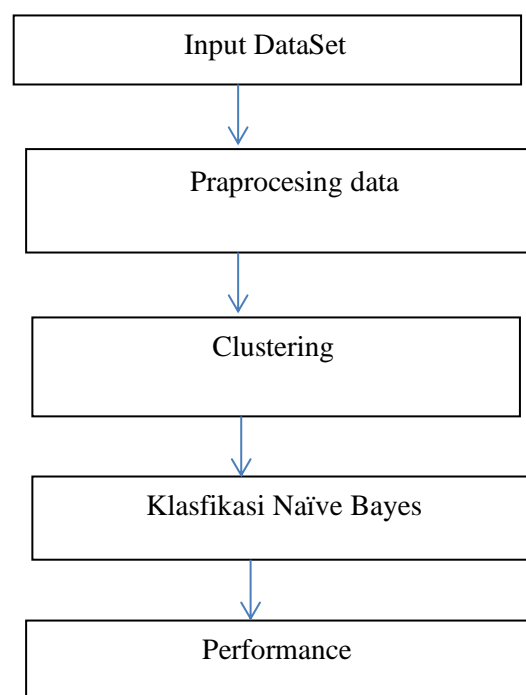
Penelitian terkait

Untuk itu penelitian ini bertujuan melakukan kajian penerapan teknik sentiment analysis untuk menganalisa data-data saran/opini mahasiswa, sehingga saran-saran dan opini tersebut dapat dimanfaatkan untuk mendukung hasil evaluasi menggunakan skor kuesioner. Pang and Lee (2008), menyatakan bahwa salah satu core dalam sentiment analysis adalah problem klasifikasi opini. Untuk konteks klasifikasi

dokumen teks secara umum (bukan teks opini), misalnya teks berita, metode NBC telah diterapkan oleh beberapa peneliti. Wibisono (2005) meneliti metode NBC untuk kategorisasi berita menghasilkan akurasi 86.9% sampai 90,23%. Penelitian Wulandini dan Nugroho (2009) membandingkan method klasifikasi teks NBC dengan method Support Vector machine (SVM), C4.5 dan K-Nearest Neighbour (K-NN). Hasil penelitian menunjukkan akurasi masing-masing method urut dari yang terbaik adalah SVM akurasi 92%, NBC akurasi 90% C4.5 akurasi 77.5% dan yang terendah K-NN akurasi 50%. SVM memiliki tingkat kompleksitas tinggi, sedangkan keunggulan NBC adalah dalam kesederhanaan komputasinya (Hamzah, 2012a;Hamzah 2012b). Akan tetapi keberhasilan metode-metode ini dalam klasifikasi opini masih harus dikaji lebih jauh.

3. Metode Penelitian

Metode penelitian berisi langkah-langkah yang digunakan dalam penelitian ini, Agar lebih mudah dipahami. Penelitian yang akan dilakukan adalah untuk merancang sistem yang diperoleh dari pengamatan data yang ada. Pada penelitian ini mengambil di universitas buana perjuangan karawang jurusan teknik informatika Metode pengumpulan data yang digunakan dalam penelitian ini adalah data kuisisioner terhadap pembelajaran semester 4 tahun ajaran 2019/2020 pada mahasiswa teknik informatika angkatan 18. Langkah yang dilakukan pada penelitian sebelumnya adalah dapat digambarkan pada kerangka pikir penelitian yang dapat dilihat pada gambar 3.1.



Gambar 3.1 Metode yang digunakan

Penjelasan dari diagram alir diatas adalah sebagai berikut:

Input

Data diinput sebagai dataset, adapun data yang diambil pada proses ini adalah hasil dari tanggapan. Mahasiswa UBP Karawang

Praprocessing data

Kategorikal data adalah data non-numerical, seperti jenis kelamin, status pernikahan, setuju atau tidak setuju. Contoh: pelayanan dosen terhadap bimbingan terlayani dengan baik atau tidak dan bisa lebih banyak, rata rata dan lebih sedikit.

Karena model hanya mengenal angka, kita harus mengubah kategori tersebut menjadi angka. Contoh: lebih kecil = 0, rata rata = 1 dan lebih banyak = 2.

Namun hal diatas bermasalah, karena model akan melihat adanya order atau urutan.

Clustering

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma clustering. Oleh karena itu, clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data. Clustering banyak digunakan dalam berbagai aplikasi seperti misalnya pada business intelligence, pengenalan pola citra, web search, bidang ilmu biologi, dan untuk keamanan (security). Di dalam business intelligence, clustering bisa mengatur banyak customer ke dalam banyaknya kelompok. Contohnya mengelompokan customer ke dalam beberapa cluster dengan kesamaan karakteristik yang kuat. Clustering juga dikenal sebagai data segmentasi karena clustering mempartisi banyak data set ke dalam banyak group berdasarkan kesamaannya. Selain itu clustering juga bisa sebagai outlier detection.

Klasifikasi Naïve Bayes

Dihitung dari probability yang dilakukan dengan prediksi tanggapan mahasiswa dengan

mengelompokkan positif, negative dan netral lalu diprediksi melakukan algoritma Naïve Bayes

Performance

Hasil adalah hasil dari pengolahan data yang sudah dilakukan akan ditentukan rule yang terbentuk untuk memberikan suatu keputusan dari algoritma. Data yang dihasilkan adalah data yang telah diolah secara keseluruhan memakai metode yang telah dipilih. Dengan hasil output ini kita sudah mendapatkan akurasi dari setiap proses yang kita uji serta mendapatkan nilai recall dosetiap tanggapan dalam layanan akademik

4. Hasil dan Pembahasan

Hasil Penelitian

Pengumpulan data dalam penelitian ini dilakukan pada tanggal 27 juni 2020 melalui google formulir. Data yang terkumpul terdiri dari 159 responden dan 31 atribut. Berikut atribut atribut yang digunakan dalam pembuatan questioner.

Singkatan	Nama Atribut
Q1	Dosen menyediakan silabus mata kuliah *
Q2	Dosen mendiskusikan silabus dengan mahasiswa *
Q3	Dosen menyediakan bahan bacaan yang sesuai dengan materi silabus *
Q4	Dosen memperlihatkan penguasaan materi matakuliah *
Q5	Dosen mengajarkan materi dengan metode yang efektif *
Q6	Dosen selalu memberi contoh konkrit setiap menjelaskan suatu hal *
Q7	Dosen sangat komunikatif *
Q8	Dosen menciptakan suasana kelas yang kondusif/ membuat mahasiswa termotivasi *
Q9	Dosen mengajar tidak terlalu cepat/lambat, sehingga mudah dimengerti mahasiswa *
Q10	Dosen selalu memberi kesempatan mahasiswa untuk bertanya *
Q11	Materi dari matakuliah telah menambah / memperluas pengetahuan dan wawasan anda *
Q12	Mahasiswa puas setelah mengikuti perkuliahan matakuliah tersebut *

Q13	Matakuliah tersebut sangat mudah dipahami mahasiswa *
Q14	Dosen menciptakan suasana kelas yang menyenangkan *
Q15	Dosen memperlihatkan sikap menghormati mahasiswa dan mendorong / memotivasi mahasiswa *
Q16	Dosen terampil menggunakan sarana teknologi modern dalam memberi kuliah *
Q17	Dosen menyelesaikan seluruh materi sesuai isi silabus mata kuliah *
Q18	Dosen tidak banyak bercerita tentang hal di luar materi matakuliah yang bersangkutan *
Q19	Materi matakuliah selalu diperbaharui dengan contoh atau perkembangan terakhir *
Q20	Dosen selalu hadir memberi kuliah setiap pertemuan *
Q21	Dosen hadir di kelas tepat waktu *
Q22	Dosen tidak pernah meniadakan kuliah tanpa alasan *
Q23	Dosen meninggalkan kelas tepat waktu *
Q24	Dosen memberi penilaian yang obyektif *
Q25	Dosen selalu memberi penjelasan tentang cara menilai *
Q26	Dosen selalu mengembalikan hasil tes / tugas dengan catatan/komentar *
Q27	Materi tugas, tes, dan ujian sesuai dengan materi mata kuliah dan selaras dengan isi silabus *
Q28	Dosen selalu mengembalikan hasil tes / tugas kepada mahasiswa dalam waktu yang wajar *
Q29	Dosen mudah ditemui di luar kelas *
Q30	Dosen berwibawa di mata mahasiswa *
Q31	Dosen memberi pendidikan tentang nilai (values), moral, etika selain tentang materi matakuliah *

Penelitian ini menggunakan tools Rapidminer studio 9.2 untuk menganalisa data data kategorikal dengan menggunakan algoritme naïve bayes, berikut hasil akurasi menggunakan algoritme naïve bayes adalah

Confusion matrix

Akurasi 86,9% ± 9,4%

	True Tinggi	True rendah	True sedang	Class precision
Pred tinggi	18	3	1	81.82%
Pred rendah	0	0	0	100%
Pred sedang	2	0	2	85.74%
Class recall	90%	76,92%	92,3%	

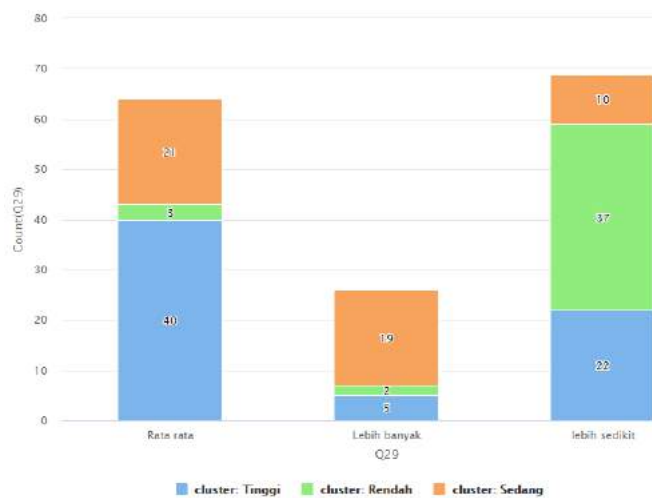
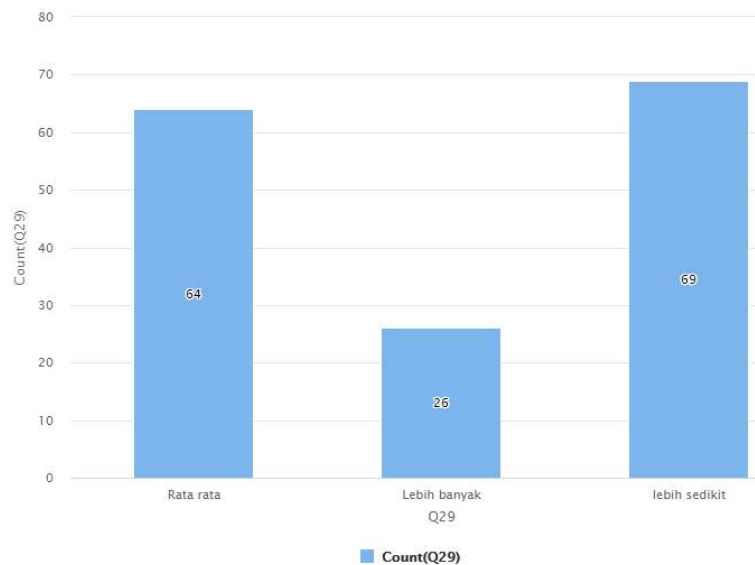
Pembahasan Penelitian

Dalam penelitian ini, beberapa point yang dapat diambil informasi penting dari 31 atribut yang digunakan adalah kategori populer yang kurang baik dan yang baik.

Kategori yang **kurang baik** dalam pertanyaan kuisioner Q29 yaitu : ***Dosen mudah ditemui di luar kelas***

Pertanyaan ini mahasiswa merasakan 69 mahasiswa dari 159 mahasiswa dengan persentasi 44%, jadi persepsi mahasiswa 44 % yaitu: ***Dosen mudah ditemui di luar kelas***. Dari 69 mahasiswa 37 cluster rendah 10 cluster sedang dan 22 cluster tinggi

Berikut grafiknya

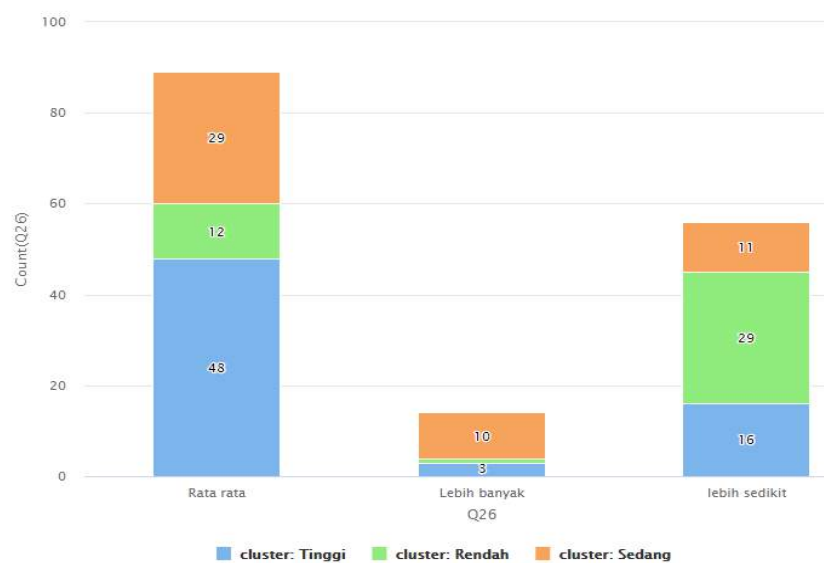
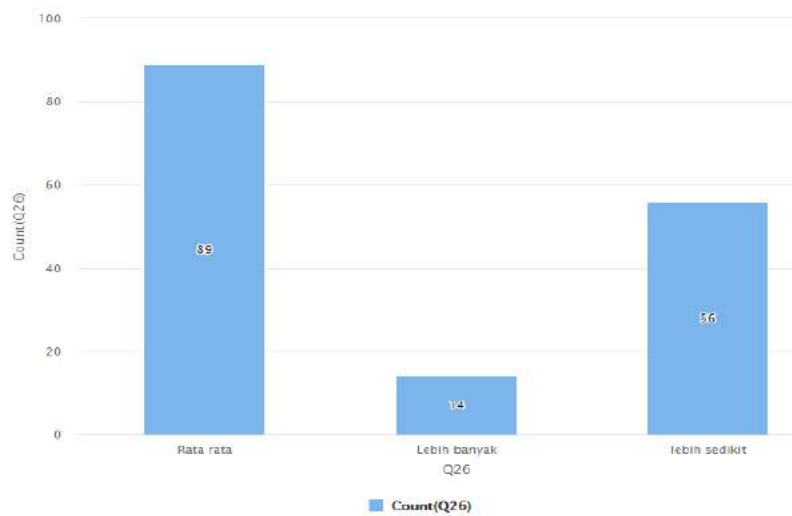


Kategori yang **kurang baik** dalam pertanyaan kuisioner Q26 yaitu: ***Dosen selalu mengembalikan hasil tes / tugas dengan catatan/komentar***

Pertanyaan ini mahasiswa merasakan 14 mahasiswa dari 159 mahasiswa dengan persentasi 9%, jadi persepsi mahasiswa 9 % yaitu: ***Dosen selalu mengembalikan hasil tes / tugas dengan catatan/komentar***.

Dari 14 mahasiswa 11 cluster rendah 0 cluster sedang dan 3 cluster tinggi

Berikut grafiknya



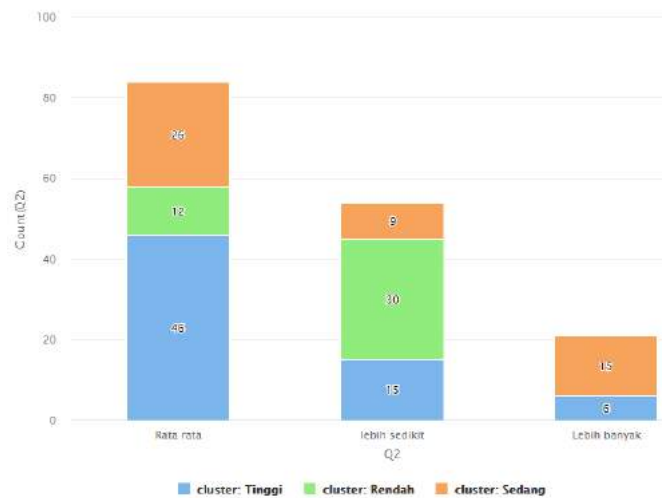
Kategori yang **kurang baik** dalam pertanyaan kuisioner Q2 yaitu : ***Dosen mendiskusikan silabus dengan mahasiswa***

Pertanyaan ini mahasiswa merasakan 21 mahasiswa dari 159 mahasiswa dengan persentasi 13%, jadi persepsi mahasiswa 13 % yaitu: ***Dosen mendiskusikan silabus dengan mahasiswa.***

Dari 21 mahasiswa 0 cluster rendah 15 cluster sedang dan 16 cluster tinggi

Berikut grafiknya



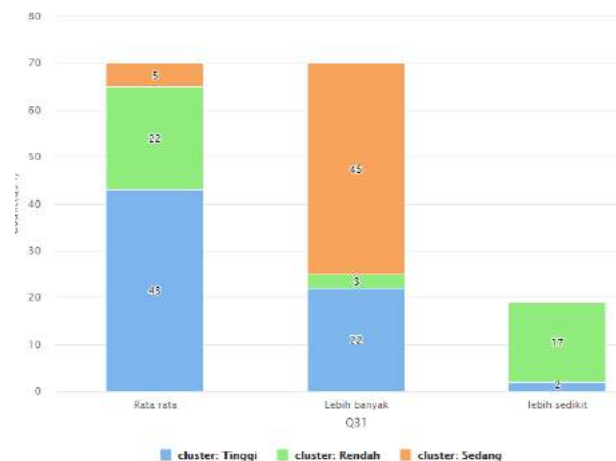
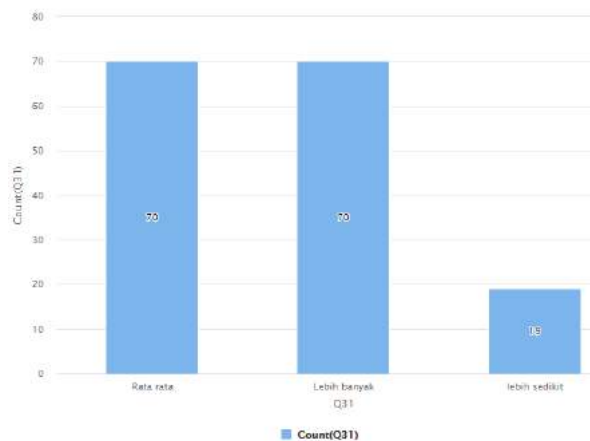


Kategori yang *baik* dalam pertanyaan kuisisioner Q31 yaitu *Dosen memberi pendidikan tentang nilai (values), moral, etika selain tentang materi matakuliah*

Pertanyaan ini mahasiswa merasakan 70 mahasiswa dari 159 mahasiswa dengan persentasi 44%, jadi persepsi mahasiswa 44 % yaitu: *Dosen memberi pendidikan tentang nilai (values), moral, etika selain tentang materi matakuliah.*

Dengan 70 mahasiswa 3 cluster rendah 45 cluster sedang dan 22 cluster tinggi

Berikut grafiknya

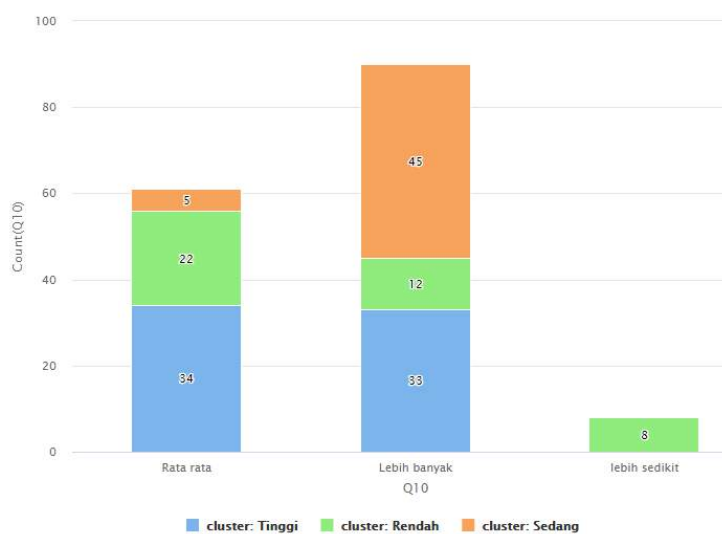
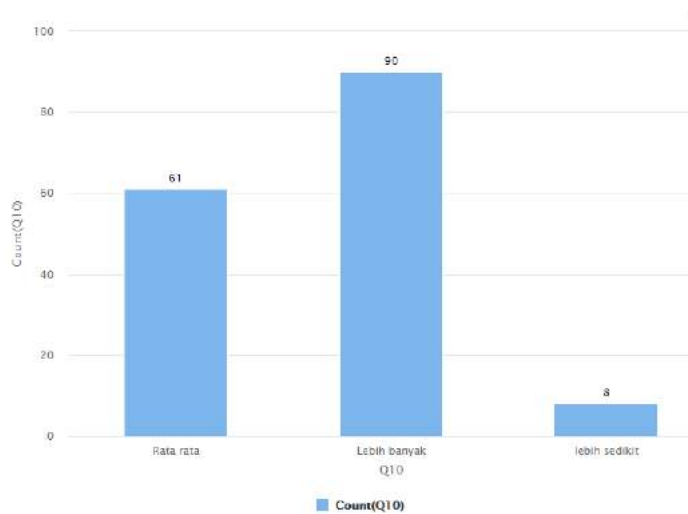


Kategori yang baik dalam pertanyaan kuisisioner Q10 yaitu : ***Dosen selalu memberi kesempatan mahasiswa untuk bertanya***

Pertanyaan ini mahasiswa merasakan 90 mahasiswa dari 159 mahasiswa dengan persentasi 57%, jadi persepsi mahasiswa 57 % yaitu: Dosen selalu memberi kesempatan mahasiswa untuk bertanya.

Dengan 90 mahasiswa 12 cluster rendah 45 cluster sedang dan 33 cluster tinggi

Berikut grafiknya



5. Kesimpulan

Setelah melakukan penelitian dengan menggunakan beberapa langkah seperti melakukan mengelompokkan dataset untuk mendapatkan kelas tinggi, sedang dan rendah. setelah dibuatkan database dengan memiliki kelas atribut menjadi dataset untuk supervised learning. Penelitian ini menggunakan metode naïve bayes untuk mengklasifikasikan terhadap dataset training dan dan uji, beberapa hasil yang didapatkan, bisa ambil beberapa kesimpulan yaitu:

1. Penelitian dengan menggunakan algoritme naïve bayes dengan akurasi 86,9 %

2. Dosen 57% lebih banyak melakukan “*Dosen selalu memberi kesempatan mahasiswa untuk bertanya*”
3. Dosen 44% lebih banyak melakukan “*Dosen memberi pendidikan tentang nilai (values), moral, etika selain tentang materi matakuliah*”
4. Dosen 9 % lebih banyak melakukan “*Dosen selalu mengembalikan hasil tes / tugas dengan catatan/komentar*”
5. Dosen 13% lebih banyak melakukan “*Dosen mendiskusikan silabus dengan mahasiswa*”
6. Dosen 44% lebih sedikit melakukan “*Dosen mudah ditemui di luar kelas*”

Saran

Berdasarkan kesimpulan yang didapat ada beberapa saran yaitu dengan menggunakan beberapa algoritme klasifikasi untuk mencari performace yang paling baik dengan kasus yang sama.

DAFTAR PUSTAKA

- Bridge, C., 2011, *Unstructured Data and the 80 Percent Rule*.(Online di: <http://clarabridge.com/default.aspx?tabid=137&ModuleID=635&ArticleID=551> di akses 29 September 2012)
- Dehaff, M. 2010. *Sentiment Analysis, Hard But Worth It!*. [Online]. Tersedia di: <http://www.customerthink.com/blog/sentiment-analysis-hard-but-worth-it> (diunduh 12 April 2013)
- Greaves,F., Cano, D.R., Millet,C., Darzi,A., and Donaldson, L., 2013, Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments,Journal Of Medical Internet Research 15:11, e239. Online publication date: 1-Jan-2013
- Grimes, S., 2013, *Unstructured Data and the 80 Precent Rule*,[online: <http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>][diases 12 April 2014]
- Hamzah, A., 2012b, *Meningkatkan Kinerja Naïve Bayes Classifier (NBC) Untuk Klasifikasi Teks dengan Menggunakan Clustering untuk Pemilihan Feature Kata*, Prosiding Seminar Nasional TEKNOIN 2012, Universitas Islam Indonesia, Yogyakarta, 10 November 2012.
- Liu,B., 2010,Sentiment Analysis, Muti Facet Problem, to Appear in IEEE intelligentSystem.[<http://www.cs.uic.edu/~liub/FBS/IEEE-Intell-Sentiment-Analysis.pdf>]
- Loia,L. and Senatore,S. , 2014, *A fuzzy-oriented sentic analysis to capture the human emotion in Webbased content*, Knowledge-Based Systems **58**, 75-85 Online publication date: 1-Mar-2014.
- Pang,B. , Lee, L. and Vaithyanathan, S.,2002, *Thumbs up? Sentiment Classification Using Machine Learning Techniques*, in Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP’02), USA, 2002, pp.
- Konferensi Nasional Penelitian dan Pengabdian (KNPP) Ke-1
Karawang, 25 Februari 2021
Universitas Buana Perjuangan Karawang

79 – 86.

Pang, B. and Lee, L., 2008, *Opinion mining and sentiment analysis* in Foundations and Trends In Information Retrieval 2(1-2), pp. 1–135, 2008

Putten, P.V.D., Kok, J. and Gupta, A., 2002, *Why the Information Explosion can be Bad for Data Mining, and How Data Fusion Provides a Way Out*, Proc. of the 2nd SIAM International Conference on Data Mining, pp:11-13

Shelke, N.M, Deshpande, S. and Thakre, 2012, *Survey of Techniques for Opinion Mining*, International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012

Tjiptono, Fandy. 2000. *Perspektif Manajemen & Pemasaran kontemporer*. Yogyakarta : ANDI.

Wulandini, F. & Nugroho, A. N. 2009. *Text Classification Using Support Vector Machine for Webmining Based Spation Temporal Analysis of the Spread of Tropical Diseases*. International Conference on Rural Information and Communication Technology 2009. (Online di: http://asnugroho.net/papers/rict2009_textclassification.pdf ; diases 28 September 2012).