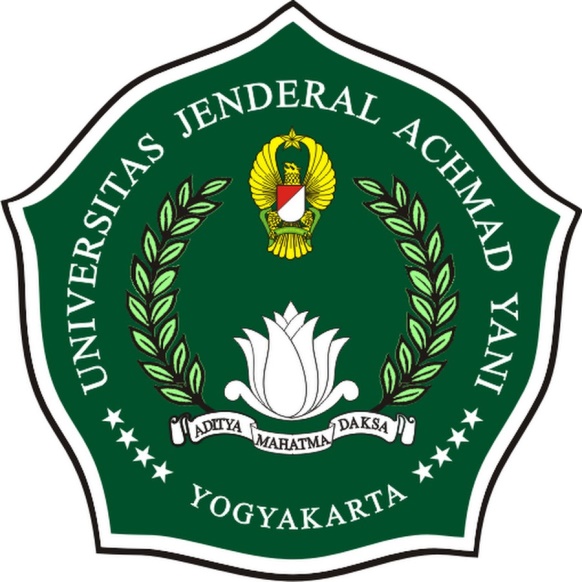
TUGAS REVIEW PAPER ATAU JURNAL

DARI PORTAL ONLINE ATAU MEDIA SOSIAL

Tugas Web Data Extraction  
Program Studi S-1 Informatika



diajukan oleh

Ridho Marhaban

182102031

kepada

UNIVERSITAS JENDERAL ACHMAD YANI

FAKULTAS TEKNIK DAN TEKNOLOGI INFORMASI

YOGYAKARTA

September, 2021

1. **Judul paper :**

ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE INDONESIA DENGAN KOMENTAR FOLLOWERS MENGGUNAKAN TOOLS ORANGE DATA MINING

1. **Tahun terbit**: 2020-01-14
2. **Masalah yang diangkat:**

malakukan analisis text clustering akun fanpage shopee dengan komentar follower yang berhubungan dengan kata kuis berhadiah, retweet yang memberikan handphone, dan tebakan dalam bentuk vote dimana topik pembicaraan dalam sebuah tweet untuk diteliti

1. **Metode yang digunakan :**

* Metode yang digunakan: metode pengumpulan data yaitu scraping dengan menggunakan web scraper untuk pengumpulan data, orange data mining untuk mengolah data analytics dan visualisasi data
* Metode pengolahan data menggunakan preprocess text, untuk membuat unit yang lebih kecil seperti transformation, tokenization, normalization, filtering
* Metode analisis data menggunakan penerapan data mining. Data selection, pre-processing/cleaning, transformation, data mining, interpretation

1. **Data yang digunakan serta cara pengambilan datanya seperti apa:**

* Data yang digunakan orange Data Mining pengambilan data komentar follower shopee dengan teknik scraping dalam widget Preprocess Text. **cara pengambilan data :**
* mengubah semua teks menjadi huruf kecil (Where the Vocher = where the vocher)
* Remove Accents akan menghapus semua diakritik/aksen dalam teks. naïve = naïve
* Parse html akan mendeteksi tag html dan mem-parsing teks saja. a sample of text = a sample of text
* Remove url akan menghapus url dari teks. This url http://orange.biolab.si/, = this url
* Setelah melakukan transformation, langkah kedua dari preprocess text adalah tokenization yaitu metode memecah teks menjadi komponen yang lebih kecil (kata, kalimat, bigrams)
* membagi teks dengan kata per kata dan membiarkan juga simbol tanda baca. This example = (This), (example), (.)
* Text yang telah dipisahkan kata per kata akan menjadi sebuah text yang berdiri sendiri dalam sebuah kalimat
* Proses terakhir pada preprocess text adalah filtering yaitu menghapus atau menyimpan pilihan kata

1. **Hasil:**

Dari hasil pengelompokkan data konten status Shopee diatas, maka peneliti melakukan analisis dari konten status Shopee Indonesia dengan komentar followers-nya untuk mendapatkan hasil kata yang dominan muncul menjadi sebuah kategori tema pembahasan yang menjadi topik yang banyak diperbincangan oleh penggunanya

Hasil Pengelompokan data komentar follower shopee

|  |  |  |  |
| --- | --- | --- | --- |
| **No cluster** | **Akun** | **Nama** | **Komentar** |
| 1 | @du2kpalingdepan | EL | Bulan Februari lalu ikut 7 lomba (Foto,nulis,kuis). 5 diantaranya kalah. 1 belum diumumkan. 1 lagi, Alhamdulillah menang. Yang penting banyak usaha. Pasti ada hasilnya. Btw, thankyou @ShopeeID pic.twitter.com/m0ROgdrTaT |
| 2 | @benzbara\_ | bernard batubara | ada yg lagi butuh hp baru? saya kasih tahu nih cara gampangnya. gapake duit. retweet tweet @ShopeeID yang ini + comment yg kreatif pake hestek #PilihShopee. kamu bisa dapet iPhone XR! belanja juga di Shopee pake kode PILIHSHOPEE dapet cashback 50%, sampai 18 Feb besok lho.https://twitter.com/ShopeeID /status/1096824564573556736 |
| 3 | @OfficialGTVID | GTV Official | Ayo dukung kontestan favoritmu untuk jadi Pemenang The Voice Indonesia. Download Shopee dan vote kontestan favoritmu di Shopee! Ada hadiah uang tunai untuk 10 voter beruntung setiap minggunya & 2 paket liburan ke Bali! @ShopeeID #VoteDiShopee #TheVoiceGTV #MakeItLouder pic.twitter.com/fFp5RGXvQ7 |

1. **Judul paper:**

MENGGUNAKAN DATA MINING UNTUK SEGMENTASI CUSTOMER PADA BANK UNTUK MENINGKATKAN CUSTOMER RELATIONSHIP MANAGEMENT (CRM) DENGAN METODE KLASIFIKASI (AGORITMA J-48, ZERO-R DAN NAIVE BAYES)

1. **Tahun terbit:** 2015
2. **Masalah yang diangkat:**

mengenai salah satu model strategi pemasaran yaitu Customer Segmentation yang membantu pihak bank untuk membagi pasar menjadi kelompok nasabah yang terbedakan dengan kebutuhan, karakteristik atau tingkah laku yang berbada yang mungkin membutuhkan produk atau bauran pemasaran yang terpisah. Customer Segmentation dapat dilakukan dengan bantuan teknik Data Mining

1. **Metode yang digunakan:**

Bank Direct Marketing Data dan Classification Method. Metode ini banyak digunakan dalam klasifikasi pelanggan dengan menganalisis data dalam sampel database pelanggan dan membuat deskripsi yang tepat untuk masing-masing kategori untuk mendirikan model pelanggan yang berbeda. menganalisis pelanggan dengan kegiatan yang sama dalam kondisi yang berbeda dan menyaring potensi masa depan pelanggan. Dengan memanfaatkan metode klasifikasi, perusahaan dapat mengklasifikasikan pelanggan potensial dalam kategori yang ada.

1. **Data yang digunakan serta cara pengambilan datanya seperti apa:**

Dataset yang digunakan dalam paper ini merupakan data mengenai nasabah bank yang berbentuk file microsoft excel. File tersebut perlu diubah dulu menjadi ekstensi .csv baru kemudian menjadi .arff untuk dapat diolah menggunakan Weka. Dataset memiliki 45211 instances dan 17 attributes. Pada masing masing attribute tidak terdeteksi adanya nilai yang kosong (missing value) namun jelas terdapat beberapa informasi yang tidak diketahui.

1. **Hasil:**

Dalam paper ini dilakukan analisis bank dataset dengan menggunakan beberapa algoritma dari metode classification, dan diperoleh kesimpulan hasil seperti yang ditunjukkan pada tabel berikut ini :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metode | Class Yes | Class No | Error Rate | Waktu (Detik) |
| Naive Bayes | 2908 | 37014 | 11.944% | 0.21 |
| Rules-- Zero -R | 0 | 39922 | 11.6985% | 0.03 |
| Trees-J48 | 783 | 39139 | 5.8769 % | 4.87 |

Untuk dataset yang digunakan yaitu bank dataset, metode classification dengan algoritma J-48 memberikan hasil terbaik dengan error rate terkecil, yaitu 5.8769%. Sedangkan jika dilihat dari segi efiesiensi waktu dan hasil klasifikasi, algoritma Zero-R memberikan hasil terbaik (0,03 detik), walaupun tidak terlalu signifikan dibandingkan dengan Naive Bayes.