**ANALISIS DAN PEMBAHASAN**

Pada bab ini akan dibahas hasil analisis dari pengolahan data yang telah dilakukan. Metode yang digunakan dalam analisis pada penelitian ini adalah klasifikasi dengan metode *Naïve Bayes Classifier* dan *Support Vector Machine* menggunakan data *tweet* yang telah didapatkan dengan Twitter API. Sebelum menganalisis, dilakukan praproses teks. Selain itu, dilakukan pula analisis dengan *Social Network Analysis* (SNA).

1. **Karakteristik Data**

Sebelum melakukan analisis lebih lanjut, terlebih dahulu dilakukan analisis untuk karakteristik data. Pada penelitian ini, data yang digunakan merupakan data *tweet* yang didapatkan dari hasil *crawling* dengan menggunakan Twitter API. Sebelum beranjak pada analisis karakteristik data yang digunakan, terlebih dahulu akan dilakukan perbandingan sosial media Twitter dari masing-masing penyedia jasa layanan transportasi umum, yaitu @Bluebirdgroup, @gojekindonesia, dan @GrabID.

**Gambar 1** Perbandingan Antar *Official Account* Twitter Tiap Taksi

Berdasarkan pada Gambar 1, dari data yang didapatkan per tanggal 5 Mei 2018, didapatkan bahwa *official account* dari @gojekindonesia merupakan *official account* dengan jumlah *followers* dan *tweet* tertinggi jika dibandingkan dengan *official account* @Bluebirdgroup dan @GrabID. Hal tersebut disebabkan karena pelayanan yang diberikan oleh @gojekindonesia bukan hanya sekedar pelayanan pada jasa transportasi umum taksi GoCar saja, namun adapula GoJek (pelayanan jasa transportasi umum motor ojek), GoFood (pelayanan jasa untuk pembelian makanan), GoSend (pelayanan antar barang), dan lain sebagainya. Pada @GrabID juga terdapat pelayanan yang sama, namun keaktifan @GrabID di media sosial Twitter masih jauh dibandingkan dengan @gojekindonesia. Sedangkan untuk @Bluebirdgroup, jumlah *tweet* pada *official account* tergolong paling sedikit jika dibandingkan dengan yang lainnya karena pelayanan yang diberikan oleh @Bluebirdgroup hanya sebatas pelayanan jasa transportasi umum taksi secara *offline* maupun *online*. Berikut merupakan data banyak *tweet* yang didapatkan dari periode 5 Februari 2018 hingga 5 Mei 2018 ­dan disajikan pada Gambar 2.

**Gambar 2** Perbandingan Banyak Data *Tweet* Dari Tiap Taksi

Dari Gambar 2 dapat diketahui bahwa banyak *tweet* yang didapatkan dari periode tanggal 5 Februari 2018 hingga 5 Mei 2018 dengan menggunakan *keyword* BlueBird Taxi yaitu sebanyak 17264 *tweet*, dengan *keyword* GoCar sebanyak 11671 *tweet*, dan dengan *keyword* GrabCar didapatkan sebanyak 6752 *tweet*. Selain itu dapat pula dilihat perbandingan banyaknya *tweet* yang dilakukan oleh masing-masing *official account*. *Tweet* terbanyak dilakukan oleh *official account* dari penyedia jasa layanan GoCar yaitu @gojekindonesia. Namun perlu diketahui, *tweet* yang diunggah oleh @gojekindonesia bukan saja tentang GoCar, akan tetapi bercampur dengan pelayanan yang lain. Sehingga wajar saja jika @gojekindonesia merupakan *official account* dengan jumlah *­­tweet* terbanyak. Jika dibandingkan dengan *official account* dari penyedia jasa layanan BlueBird Taxi memang sangat berbeda jauh, karena pelayanan yang diberikan oleh pihak @Bluebirdgroup hanya sebatas pelayanan jasa transportasi umum secara *offline* dan *online*. Dari Gambar 2 dapat dilakukan analisis lebih detail untuk mengetahui jenis *tweet* apa saja yang didapatkan seperti ditampilkan pada Gambar 3.

**Gambar 3** Kategori *Tweet* Pengguna

Gambar 3 menunjukkan beberapa macam *tweet* dari pengguna yang dibagi dalam beberapa kategori yang telah ditentukan, yaitu tentang pelayanan, aplikasi, tarif, promo atau voucher, berita, pertanyaan, dan lain-lain. Dapat diketahui bahwa sebagian besar *tweet* yang didapatkan dengan menggunakan *keyword* BlueBird Taxi adalah pada kategori berita yaitu sebanyak 11277 *tweet*. Sedangkan, untuk *tweet* dari pengguna tentang pelayanan BlueBird Taxi hanya sebanyak 2319 *tweet*. Kemudian pada *keyword* GoCar didapatkan sebagian besar dari pengguna merupakan *tweet* yang tidak termasuk dalam beberapa kategori yang diberikan atau dapat dikatakan termasuk dalam kategori *tweet* dengan kategori lain-lain yaitu sebanyak 4372 *tweet*. Hal tersebut dikarenakan sebagian besar pengguna Twitter menggunakan kata GoCar sebagai bahan bercanda atau hanya sekedar istilah pada *tweet* yang diunggahnya, sehingga tidak termasuk dalam beberapa kategori tentang pelayanan atau hal lain yang berhubungan dengan GoCar. Namun, dari keseluruhan *tweet* didapatkan sebanyak 3544 *tweet* dari *tweet* pengguna yang menggunggah tentang pelayanan dari jasa penyedia transportasi umum taksi GoCar.

Selanjutnya pada *keyword* GrabCar, pengguna Twitter sebagian besar menggunggah *tweet* yang tidak termasuk dalam beberapa kategori yang telah ditentukan, yaitu sebanyak 2020 *tweet*. Penyebab dari munculnya hal tersebut sama seperti pada jenis taksi GoCar, yaitu sebagian besar para pengguna Twitter menggunakan kata GrabCar hanya sebagai kata kiasan ataupun hanya sekedar sebagai bahan bercanda, sehingga *tweet* tersebut tidak dapat dikategorikan ke dalam jenis *tweet* tentang pelayanan, aplikasi, tarif, dan sebagainya. Akan tetapi, pada kategori pelayanan didapatkan jumlah *tweet* sebanyak 1662 *tweet*. Jumlah tersebut lebih besar jika dibandingkan dengan kategori-kategori lainnya Sehingga dapat dikatakan bahwa kategori pelayanan juga mendominasi pada *tweet* yang diunggah oleh pengguna GrabCar. Berikut akan diberikan beberapa contoh *tweet* untuk setiap kategori beserta contoh *tweet* dari *official account* dari masing-masing penyedia jasa transportasi umum taksi yang disajikan pada Tabel 1.

**Tabel 1** Contoh Isi *Tweet* Tiap Kategori

|  |  |  |
| --- | --- | --- |
| Kategori | Taksi | *Tweet* |
| *Official Account* | BlueBird | @W\_Thok Terima kasih Pak Wahyu atas kepercayaannya dan apresiasinya kepada pelayana taksi CCD7431. Semoga pengemudi… https://t.co/8alltuR8uJ |
| GoCar | @remahremahastor Hai Dessy, silakan informasikan nomor order yg Anda keluhkan dan detail kronologi Anda melalui DM. Tks ^yun |
| GrabCar | @AsmaraAra Hai Kak, saya Gita dari Grab Indonesia. Terima kasih sudah berpartisipasi dan mendukung program… https://t.co/HlPNRm8EYH |
| Pelayanan | BlueBird | Kami sangat apresiasi sekali dgn pengemudi @Bluebirdgroup ini. Bertanggung jawab &amp; ikhlas memutar ruter menjemput p… https://t.co/dh1UHdWrQO |
| GoCar | Gue berterimakasih bgt sama bapak driver gocar yg td plus mamas2 driver gojek yg anterin ke hotel trs ke bandara ju… https://t.co/xgyFd52jUJ |
| GrabCar | @GrabID Hai,sy mau komplen kejadian barusan jm 16.12. Psn grabcar utk papa,di cek ud in transit dan… https://t.co/cVJU9dsPrl |
| Aplikasi | BlueBird | Fitur Easy Ride di @Bluebirdgroup sangat menyenangkan, kalau saya lagi nggak pegang uang cash. |
| GoCar | Setelah update @gojekindonesia di appstore tampilan pemesanan berubah. Jujur tampilan baru sangat tdk mendukung use… https://t.co/UItEQWHC6r |

**Tabel 1** Contoh Isi *Tweet* Dari BlueBird Taxi (Lanjutan)

|  |  |  |
| --- | --- | --- |
| Kategori | Taksi | *Tweet* |
| Aplikasi | GrabCar | Secara user interface , grab car memiliki desain yang lebih elegan dibandingkan dengan taksi lainnya #katague |
| Tarif | BlueBird | Hai @bluebirdgroup! Tarifnya skrg semakin mahal. Ada kenaikan charge argo kah?? Biasanya estimasi dr Mampang Depok… https://t.co/AQnY894Dsr |
| GoCar | Go ride sekarang mahal yah... Biasa pake gopay cuma 6K jadi 10K. Malah murah naik gocar masaa.. @gojekindonesia https://t.co/I3xc7YONKL |
| GrabCar | lama tak naik grab tapi kenapa grabcar mahal gila ah skrg. |
| Promo atau Voucher | BlueBird | Coba e-voucher @Bluebirdgroup.. Mantap.. Lihat perjalanan saya menggunakan Blue Bird: https://t.co/gYvNdDzBsN |
| GoCar | Enakny pakai GO Car pakai voucher. Jarak jauh dekat cuma segitu :) #gocar #gojek #gocar @… https://t.co/v4KvFRwUIt |
| GrabCar | Grab ini makin lama makin pelit "Diskon 10% utk 10x naik GrabCar tp maks Rp3ribu" |
| Berita | BlueBird | Blue Bird Siap Hadapi Transportasi Online https://t.co/HY8qBgwvLm |
| GoCar | SOLOPOS HARI INI : Gocar Mogok, Tarif Melonjak 300% https://t.co/70WpNZVloM |
| GrabCar | Taksi Online GrabCar Resmi Beroperasi di Bandara Husein Sastranegara Bandung https://t.co/P9ArrLe07g |
| Pertanyaan | BlueBird | @Bluebirdgroup kalo pesan taxi di area bandara via aplikasi mybluebird apakah kena tambahan service charge? |
| GoCar | @gojekindonesia saya mau bergabung menjadi driver go car.tapi mobil saya masih plat putih atau profit.stnk nya belu… https://t.co/GeARu9vdu7 |

**Tabel 1** Contoh Isi *Tweet* Dari BlueBird Taxi (Lanjutan)

|  |  |  |
| --- | --- | --- |
| Kategori | Taksi | *Tweet* |
| Pertanyaan | GrabCar | @GrabID min sy tadi naik grabcar terus ada barang ketinggalan di mobil sedangkan gaada nomer drivernya. kode booking ADR99616093499 gmn min? |
| Lain-Lain | BlueBird | @otong\_leyong Nasib bluebird (dan taksi non online) semua bergantung permenhub baru mengenai layanan taksi online.… https://t.co/nk4PKFlnc1 |
| GoCar | RT @nadiarvh: Aku : Pak, saya minta turun di lobby south ya. Driver gocar: Mba saya turunin di west ya. Saya malu nemenin mba, mba cantik,… |
| GrabCar | Gue sih #TimDudukBelakang, supaya kalo gue inget mantan terus sedih, ga ketauan sama mitra GrabCar. Kan malu. Wkwk https://t.co/ZSmPNcy9HV |

Dari semua *tweet* pengguna jasa transportasi umum taksi yang didapatkan pada Gambar 2 tidak semua *tweet* mengandung sentimen, sehingga peneliti melakukan langkah awal dengan melakukan klasifikasi sentimen dengan dua kategori yaitu sentimen positif dan sentimen negatif pada masing-masing *tweet* pengguna yang didapatkan secara manual dan berdasarkan dari persepsi peneliti. Sehingga didapatkan banyak *tweet* yang mengandung sentimen positif atau sentimen negatif yang disajikan pada Gambar 4.

Pada Gambar 4 dapat diketahui bahwa dari 17264 *tweet* pengguna yang mengunggah *tweet* menggunakan *keyword* BlueBird Taxi didapatkan 3111 *tweet* yang mengandung sentimen. Kemudian dapat diketahui pula dari *tweet* pengguna yang mengunggah *tweet* dengan *keyword* GoCar yaitu sebanyak 11671 *tweet*. Namun setelah dilakukan klasifikasi sentimen, ternyata hanya didapatkan sebanyak 4334 *tweet* pengguna yang mengandung sentimen. Sedangkan pada *tweet* pengguna yang menggunakan *keyword* GrabCar, yaitu sebanyak 6752 *tweet* pengguna yang didapatkan, hanya terdapat sebanyak 2286 *tweet* yang mengandung sentimen. Berdasarkan dari Gambar 4, maka didapatkan perbandingan antara ketiga jenis penyedia jasa transportasi umum taksi yang disajikan pada Gambar 5.

**Gambar 4** *Tweet* Sentimen

Berdasarkan Gambar 4 dapat diketahui bahwa banyak *tweet* yang mengandung sentimen yang didapatkan merupakan *tweet* tentang jasa transportasi umum taksi GoCar dibandingkan dengan BlueBird ataupun GrabCar. *Tweet* yang diunggah oleh para pengguna tentang jenis taksi GoCar jika berdasarkan dari Gambar 5 lebih banyak mengarah ke hal yang positif daripada hal negatif. Jika dibandingkan antara jenis taksi BlueBird Taxi dan GrabCar, berdasarkan Gambar 4 pengguna lebih sering mengunggah *tweet* yang mengandung sentimen tentang BlueBird Taxi daripada tentang GrabCar. Namun berdasarkan Gambar 5, sebagian besar *tweet* yang diunggah oleh pengguna tentang BlueBird Taxi lebih mengarah pada sentimen negatif daripada sentimen positif. Akan tetapi pada jenis taksi GrabCar, walaupun pengguna jarang mengunggah *tweet* tentang GrabCar namun terlihat pada Gambar 5 bahwa *tweet* yang mengandung sentimen yang diunggah oleh pengguna lebih banyak mengarah pada hal positif daripada negatifnya. Berdasarkan Gambar 5, dapat diketahui pula bahwa perbandingan *tweet* yang mengandung sentimen positif dan sentimen negatif dari masing-masing penyedia jasa transportasi umum taksi cenderung *balanced* atau seimbang antara *tweet* yang mengandung sentimen positif dan *tweet* yang mengandung sentimen negatif.

**Gambar 5** Perbandingan *Tweet* Sentimen Antar Jenis Taksi

Berdasarkan Gambar 4 dan Gambar 5 didapatkan bahwa dari *tweet* sentimen yang diperoleh dari topik tentang taksi konvensional dan taksi berbasis *online* didapatkan bahwa pengguna lebih banyak mengunggah *tweet* bersentimen positif pada taksi berbasis *online* baik itu GoCar maupun GrabCar jika dibandingkan dengan taksi konvensional BlueBird Taxi. Hal tersebut dapat menunjukkan bahwa pengguna sebagian besar telah merasa terpuaskan dengan jasa yang diberikan sehingga memberikan respon yang positif terhadap jasa yang telah diberikan dengan mengunggah *tweet* bersentimen positif pada taksi berbasis *online* GoCar dan GrabCar. *Tweet* yang diunggah dapat berupa jasa dari segi pelayanan secara langsung dari sopir atau *customer service*, jasa dari aplikasi yang telah disediakan, promo atau voucher yang diberikan, dan lain-lain. Sedangkan, jasa dari taksi konvensional BlueBird Taxi sebagian besar belum dapat memuaskan pelanggan dari segi pelayanan dari sopir atau *customer service*, jasa dari aplikasi yang disediakan, promo atau voucher yang diberikan, dan lain-lain sehingga pengguna taksi konvensional BlueBird Taxi lebih cenderung memberikan respon yang negatif dengan mengunggah *tweet* bersentimen negatif. Dalam hal ini, kepuasan pengguna terhadap jasa yang diberikan oleh taksi berbasis *online* lebih baik jika dibandingkan dengan taksi konvensional.

Setelah didapatkan *tweet* yang mengandung sentimen dari masing-masing jenis taksi, maka dapat dilakukan analisis selanjutnya yaitu klasifikasi dengan menggunakan metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) pada data *tweet* yang telah diklasifikasikan mengandung sentimen positif ataupun negatif. Namun sebelumnya akan dilakukan praproses teks.

1. **Praproses Teks**

Data *tweet* mengenai taksi konvensional BlueBird Taxi dan taksi berbasis *online* GoCar dan GrabCar yang telah terkumpul dilakukan praproses teks yang meliputi *cleansing*, *case folding*, *stemming, stopword*, dan *tokenizing*. Penjelasan mengenai hasil praproses teks dari setiap tahapan akan dijabarkan pada simulasi praproses teks pada sebuah data *tweet*. *Tweet* yang akan digunakan sebagai contoh adalah dua contoh *tweet* dari BlueBird Taxi. Tahapan dan hasil praproses menggunakan salah satu *tweet* tersebut ditunjukkan pada Tabel 2. Pada Tabel 2 menunjukkan contoh *tweet* yang melalui tahapan praproses teks hingga didapatkan hasil akhir pada proses *stopwords* dan *tokenizing*. Praproses teks yang telah dilakukan bertujuan untuk meningkatkan ketepatan klasifikasi dan mengurangi kesalahan klasifikasi data.

**Tabel 2** Simulasi Praproses Teks

|  |  |
| --- | --- |
| Tahapan | Hasil Praproses |
| Contoh *tweet* 1 | @Bluebirdgroup Itu yang saya males.. Ribet pake telp... Akun saya juga gak bisa login..... Nunggu yang lewat aja lah..... |
| Contoh *tweet* 2 | @Bluebirdgroup luar biasa inisiatif driver Bapak Rudyanto. Membuat penumpang nyaman dan Tissue, bolpen,… https://t.co/x4fo6xFyp |
| *Cleansing tweet* (menghapus simbol RT, *username*, dan *link* URL) | Itu yang saya males.. Ribet pake telp... Akun saya juga gak bisa login..... Nunggu yang lewat aja lah..... |
| luar biasa inisiatif driver Bapak Rudyanto. Membuat penumpang nyaman dan Tissue, bolpen, |
| *Case folding* | itu yang saya males ribet pake telp akun saya juga gak bisa login nunggu yang lewat aja lah |
| luar biasa inisiatif driver bdu2412 bapak rudyanto membuat penumpang nyaman dan tissue bolpen |
| *Stemming* | itu yang saya males ribet pake telp akun saya juga gak bisa login nunggu yang lewat aja lah |
| luar biasa inisiatif driver bapak rudyanto buat tumpang nyaman dan tissue bolpen |
| *Stopwords* dan *Tokenizing* | ['males', 'ribet', 'pake', 'telp', 'akun', 'login', 'nunggu'] |
| ['inisiatif', 'driver', 'rudyanto', 'tumpang', 'nyaman', 'tissue', 'bolpen'] |

Dari kedua contoh *tweet* dari *tweet* tentang BlueBird Taxi yang telah melalui praproses teks tersebut, maka didapatkan struktur data setelah praproses teks ditunjukkan pada Tabel 3. Tabel 3 menunjukkan pembentukan struktur data setelah dilakukan praproses teks dengan menjadikan setiap kata menjadi variabel prediktor dan meletakkannya pada satu baris. Jika terdapat tambahan kata (variabel prediktor) dari *tweet* baru, maka kata tersebut diletakkan pada baris yang sama dan di kolom berikutnya. Namun jika terdapat kata yang sama atau kata yang telah ada pada struktur data, maka kata tersebut tidak dimasukkan lagi pada struktur data. Sehingga tidak terdapat kata atau variabel prediktor yang sama dalam struktur data. Nilai dari setiap kata tersebut merupakan frekuensi kemunculan kata dalam *tweet* ke-*i* seperti yang terdapat pada Tabel 3.

**Tabel 3** Struktur Data Setelah Praproses Teks

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Tweet* ke- | Variabel Prediktor | | | | | | | | |
| males | ribet | pake | telp | akun | login | … | tissue | bolpen |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | … | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | … | 1 | 1 |
| . | . | . | . | . | . | . | … | . | . |
| . | . | . | . | . | . | . | … | . | . |
| 3111 | 0 | 0 | 0 | 0 | 0 | 0 | … | 0 | 0 |

Data yang telah terbentuk *document term matrix* seperti pada Tabel 3 dapat dilakukan perhitungan banyak kata yang selanjutnya akan menjadi banyak variabel dari setiap jenis taksi. Data *tweet* BlueBird Taximemiliki kata kunci sebanyak 3737 kata dan data *tweet* GoCar memiliki kata kunci sebanyak 5184 kata, sedangkan data *tweet* GrabCar memiliki kata kunci sebanyak 3147 kata. Setelah terbentuk struktur data yang diinginkan, maka dapat dilakukan perhitungan frekuensi kemunculan kata pada setiap taksi. Berikut merupakan 10 kata kunci dengan frekuensi tertinggi untuk tiap penyedia jasa layanan taksi yang ditampilkan pada Tabel 4.

Pada Tabel 4 dapat disimpulkan bahwa hal yang paling diperhatikan atau disampaikan pendapatnya oleh para pengguna taksi konvensional dan taksi berbasis *online* sama, yaitu tentang ‘driver’. Hal tersebut terbukti karena pada kata kunci BlueBird Taxi dan GoCar, kata kunci ‘driver’ merupakan kata kunci terbanyak pertama, dan pada GrabCar kata kunci ‘driver’ berada pada urutan kelima. Hal ini tentu saja dapat menjadi pertimbangan untuk perusahaan penyedia jasa taksi, baik itu taksi konvensional maupun taksi berbasis *online* untuk lebih memperhatikan drivernya agar lebih dapat memberikan pelayanan yang lebih baik kepada para penggunanya, karena driver merupakan salah satu hal yang penting yang sangat diperhatikan oleh para pengguna jasa transportasi umum taksi. Dari daftar 10 kata kunci dengan frekuensi kemunculan tertinggi pada setiap penyedia jasa layanan transportasi umum taksi pada Tabel 4 menunjukkan bahwa kata-kata tersebut merupakan kata-kata yang mempunyai pengaruh yang signifikan dalam pembangunan model klasifikasi.

**Tabel 4** Frekuensi Kata Kunci Untuk Masing-Masing Taksi

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BlueBird | | GoCar | | GrabCar | |
| Kata Kunci | **Frekuensi** | **Kata Kunci** | **Frekuensi** | **Kata**  **Kunci** | **Frekuensi** |
| Driver | 463 | Driver | 1392 | Mitra | 456 |
| App | 457 | Uber | 916 | Ngobrol | 450 |
| Tumpang | 327 | Selamat | 772 | Timdudukdepan | 438 |
| Supir | 319 | Rating | 757 | Inspirasi | 429 |
| Bandara | 263 | Bahas | 756 | Driver | 393 |
| Tarif | 249 | Aneh | 753 | Pake | 194 |
| Gojek | 248 | Simpel | 753 | Katague | 144 |
| Org | 210 | Gojek | 389 | Grabbike | 142 |
| Nyegat | 202 | Order | 346 | Mobil | 140 |
| Gada | 200 | Pake | 330 | Order | 139 |

1. **Klasifikasi Menggunakan *Naïve Bayes Classifier* (NBC)**

Metode *Naïve Bayes Classifier* merupakan klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Kelebihan dalam NBC adalah algoritmanya sederhana tetapi memiliki akurasi yang tinggi. Dalam analisis dengan menggunakan NBC, data akan dibagi menjadi data *training* dan data *testing*. Proses ini akan dilakukan pada masing-masing jenis taksi dengan menggunakan seluruh data *tweet* yang telah dilakukan pengkategorian sentimen sebelumnya. Klasifikasi dengan menggunakan metode *Naïve Bayes Classifier* menghasilkan probabilitas yang digunakan untuk menentukan apakah *tweet* masuk ke dalam kategori sentimen positif atau negatif. Model klasifikasi yang didapatkan dari data *training* untuk menghitung nilai probabilitas untuk tiap kategori sentimen ditampilkan dalam Tabel 5.

**Tabel 5** Model Klasifikasi NBC

|  |  |  |
| --- | --- | --- |
| Taksi | Sentimen | Model |
| BlueBird | Positif |  |
| Negatif |  |
| GoCar | Positif |  |
| Negatif |  |
| GrabCar | Positif |  |
| Negatif |  |

Pada Tabel 5, menunjukkan *fb1* merupakan frekuensi kata pertama pada BlueBird Taxi, *fb2* merupakan frekuensi kata kedua pada BlueBird Taxi, hingga *fb3734* yaitu banyaknya kata kunci yang didapatkan pada BlueBird Taxi. Kemudian *fgo1* merupakan frekuensi kata pertama pada GoCar, *fgo2* merupakan frekuensi kata kedua pada GoCar, hingga *fgo5181* yaitu banyaknya kata kunci yang didapatkan pada GoCar. Sedangkan, *fgr1* merupakan frekuensi kata pertama pada GrabCar, *fgr2* merupakan frekuensi kata kedua pada GrabCar, hingga *fgr3144* yaitu banyaknya kata kunci yang didapatkan pada GrabCar. Frekuensi setiap kata kunci yang didapatkan untuk masing-masing sentimen tidak sama, tergantung berapa banyak kata kunci tersebut muncul pada sentimen positif atau sentimen negatif. Berdasarkan model yang telah didapatkan, maka untuk menentukan kategori sentimen pada data *tweet* dapat dicari probabilitas terbesar dari hasil yang telah didapatkan dari masing-masing kategori sentimen. Berikut adalah hasil *accuracy* pada data *testing* dengan menggunakan metode *k-fold cross validation* dengan menggunakan 10-*fold* untuk tiap jenis taksi yang ditunjukkan pada Tabel 6.

**Tabel 6** Nilai *Accuracy* Metode NBC dengan 10*-fold CV*

|  |  |  |  |
| --- | --- | --- | --- |
| *fold ke-* | BlueBird | GoCar | GrabCar |
| 1 | 0.750 | 0.647 | 0.778 |
| 2 | 0.701 | 0.802 | 0.896 |
| 3 | 0.791 | 0.906 | 0.848 |
| 4 | 0.871 | 0.880 | 0.873 |
| 5 | 0.775 | 0.832 | 0.732 |
| 6 | 0.711 | 0.694 | 0.662 |
| 7 | 0.932 | 0.617 | 0.680 |
| 8 | 0.920 | 0.624 | 0.899 |
| 9 | 0.897 | 0.729 | 0.759 |
| 10 | 0.746 | 0.685 | 0.697 |
| Rata-Rata | 0.809 | 0.741 | 0.782 |

Berdasarkan Tabel 6 didapatkan rata-rata nilai *accuracy* pada masing-masing penyedia jasa transportasi umum taksi dengan menggunakan 10-*fold cross validation*. Pada BlueBird Taxi, didapatkan rata-rata nilai *accuracy* sebesar 80.9% dengan nilai *accuracy* tertinggi pada *fold* ke 7. Kemudian, untuk GoCar didapatkan rata-rata nilai *accuracy* sebesar 74.1% dengan nilai *accuracy* tertinggi pada *fold* ke 3. Sedangkan, pada GrabCar didapatkan rata-rata nilai *accuracy* sebesar 78.2% dengan nilai *accuracy* tertinggi pada *fold* ke 8. Selanjutnya akan dilakukan pengukuran ketepatan klasifikasi dengan membentuk *confusion matrix* berdasarkan *fold* dengan nilai *accuracy* tertinggi dari masing-masing jenis taksi. Berikut merupakan hasil *confusion matrix* untuk tiap jenis taksi yang disajikan pada Tabel 7.

**Tabel 7** *Confusion Matrix* dengan Metode NBC

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Jenis Taksi | Jumlah *Tweet* | *True Positive* | *False Positive* | *True Negative* | *False Negative* |
| *Training* | BlueBird | 2800 | 939 | 33 | 1677 | 151 |
| GoCar | 3900 | 2012 | 147 | 1525 | 216 |
| GrabCar | 2058 | 1165 | 83 | 766 | 44 |

**Tabel 7** *Confusion Matrix* dengan Metode NBC (Lanjutan)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Jenis Taksi | Jumlah *Tweet* | *True Positive* | *False Positive* | *True Negative* | *False Negative* |
| *Testing* | BlueBird | 311 | 120 | 20 | 170 | 1 |
| GoCar | 434 | 247 | 40 | 146 | 1 |
| GrabCar | 228 | 129 | 18 | 76 | 5 |

Setelah terbentuk *confusion matrix* pada Tabel 7, langkah selanjutnya yaitu melakukan perhitungan ketepatan klasifikasi. Hasil perhitungan ketepatan klasifikasi untuk setiap jenis taksi dirangkum menjadi satu tabel berdasarkan data *training* dan data *testing*. Berikut merupakan hasil pengukuran ketepatan klasifikasi dari setiap penyedia jasa transportasi umum taksi dengan menggunakan pengukuran nilai *accuracy*, *sensitivity*, *specificity*, *G-Mean*, dan AUCyang ditunjukkan pada Tabel 8.

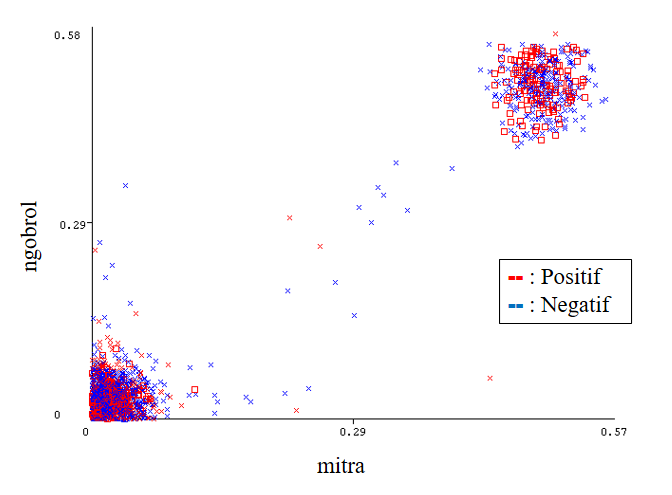
**Tabel 8** Ketepatan Klasifikasi Metode NBC

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Taksi | *Accuracy* | *Sensitivity* | *Specificity* | *G-Mean* | AUC |
| *Training* | BlueBird | 0.934 | 0.861 | 0.981 | 0.919 | 0.921 |
| GoCar | 0.907 | 0.903 | 0.912 | 0.908 | 0.908 |
| GrabCar | 0.938 | 0.964 | 0.902 | 0.932 | 0.933 |
| *Testing* | BlueBird | 0.932 | 0.992 | 0.895 | 0.942 | 0.943 |
| GoCar | 0.906 | 0.996 | 0.785 | 0.884 | 0.890 |
| GrabCar | 0.899 | 0.963 | 0.809 | 0.882 | 0.886 |

Berdasarkan Tabel 8 dapat diketahui bahwa hasil klasifikasi pada data *training* untuk ketiga jenis taksi didapatkan nilai *accuracy* yang cukup tinggi yaitu diatas 90%. Kemudian pada nilai AUC, didapatkan hasil nilai AUC untuk ketiga jenis taksi berada diatas 0.9. Jika mengacu pada Tabel 2.3, maka dapat diartikan ketepatan klasifikasi pada data *training* untuk ketiga jenis taksi adalah baik sekali. Karena data yang digunakan cenderung *balanced*, maka selanjutnya perbandingan dapat dilakukan dengan menggunakan nilai *accuracy*. Pada data *testing*, didapatkan nilai *accuracy* untuk penyedia jasa transportasi umum BlueBird Taxi, yaitu sebesar 93.2%. Sedangkan untuk penyedia jasa transportasi umum GoCar dan GrabCar, didapatkan nilai *accuracy* yaitu sebesar 90.6% untuk GoCar dan 89.9% untuk GrabCar.

**4 Klasifikasi Menggunakan *Support Vector Machine* (SVM)**

*Support Vector Machine* (SVM) merupakan sistem pembelajaran yang menggunakan hipotesis fungsi linear dalam ruang berdimensi tinggi dan dilatih dengan algoritma berdasarkan teori optimasi dengan menerapkan *learning bias* yang berasal dari teori statistik. Tujuan utama dari metode ini adalah membangun pemisah optimum yang disebut OSH (*Optimal Separating Hyperplane*) sehingga dapat digunakan untuk klasifikasi. Pembahasan pada SVM ini akan sama dengan pembahasan pada NBC. Perbedaannya adalah ditambahkan parameter untuk SVM dengan Kernel Linear yaitu parameter *C*, dan parameter untuk SVM dengan Kernel RBF (*Radial Basis Function*) yaitu parameter *C* dan parameter *gamma* (*γ*). Data yang akan dianalisis menggunakan metode SVM merupakan data yang telah dilakukan pembobotan dengan menggunakan *term frequency-inverse document frequency* (TF-IDF) terlebih dahulu pada setiap kata. Pembobotan tersebut menyebabkan adanya sebaran data. Berikut ini merupakan salah satu contoh sebaran data pada variabel kata ‘driver’ dan ‘app’ yang merupakan kata kunci yang didapatkan dari BlueBird Taxi pada kategori sentimen positif dan sentimen negative yang disajikan pada Gambar 6.



**Gambar 6** *Scatterplot* Variabel Ngobrol dan Variabel Mitra

Pada Gambar 6 menunjukkan bahwa *scatterplot* antara variabel ‘mitra’ dan variabel ‘ngobrol’ terdapat kemungkinan data dapat terpisah secara linear, sehingga pada pembahasan klasifikasi menggunakan metode SVM akan dilakukan dengan menggunakan Kernel Linear dan Kernel RBF. Pembahasan pertama akan dilakukan dengan menggunakan Kernel Linear dan kemudian dilanjutkan dengan menggunakan Kernel RBF.

**4.1 SVM Menggunakan Kernel Linear**

Pada mulanya data *tweet* dibagi menjadi dua klasifikasi sentimen sama seperti pembahasan pada NBC. Parameter *C* yang digunakan pada Kernel Linear akan dicoba dari 10-2 hingga 104 (Huang, Lee, Lin, & Huang, 2007). Dalam analisis dengan menggunakan SVM Kernel Linear, data akan dibagi menjadi data *training* dan data *testing*. Berikut merupakan hasil ketepatan klasifikasi dari metode SVM dengan menggunakan Kernel Linear untuk menentukan parameter *C* yang memberikan hasil ketepatan klasifikasi optimum untuk masing-masing penyedia jasa transportasi umum taksi yang disajikan pada Tabel 9.

**Tabel 9** Penentuan Parameter SVM Kernel Linear

|  |  |  |  |
| --- | --- | --- | --- |
| *C* | Ketepatan Klasifikasi | | |
| **BlueBird Taxi** | **GoCar** | **GrabCar** |
| 0.01 | 0.606 | 0.577 | 0.590 |
| 0.1 | 0.738 | 0.698 | 0.706 |
| 1 | 0.851 | 0.808 | 0.806 |
| 10 | 0.839 | 0.791 | 0.796 |
| 100 | 0.832 | 0.767 | 0.792 |
| 1000 | 0.824 | 0.766 | 0.792 |
| 10000 | 0.824 | 0.766 | 0.792 |

Dari Tabel 9 didapatkan ketepatan klasifikasi optimum yang didapat menggunakan SVM Kernel Linear pada masing-masing jenis taksi adalah dengan menggunakan parameter *C* sebesar 1. Selanjutnya adalah dilakukan pengukuran ketepatan klasifikasi dengan membentuk *confusion matrix* dengan menggunakan parameter *C* sebesar 1. Selanjutnya, dilakukan pemecahan data *tweet* menjadi data *training* dan data *testing* berdasarkan metode *k-fold cross validation* dengan menggunakan 10-*fold*. Berikut adalah nilai *accuracy* pada data *testing* dengan menggunakan 10-*fold cross validation* pada setiap penyedia jasa transportasi umum taksi yang disajikan pada Tabel 10.

Dari Tabel 10 menunjukkan rata-rata nilai *accuracy* pada BlueBird Taxi sebesar 81.3% dengan nilai *accuracy* tertinggi pada *fold* ke 7. Kemudian pada GoCar, didapatkan rata-rata nilai *accuracy* dengan menggunakan 10-*fold cross validation* adalah sebesar 86.6% dengan nilai *accuracy* tertinggi pada *fold* ke 3. Sedangkan untuk GrabCar, didapatkan rata-rata nilai *accuracy* adalah sebesar 75.8% dengan nilai *accuracy* tertinggi didapatkan pada *fold* ke 8.

**Tabel 10** Nilai *Accuracy* Metode SVM Kernel Linear dengan 10*-fold CV*

|  |  |  |  |
| --- | --- | --- | --- |
| *fold ke-* | BlueBird | GoCar | GrabCar |
| 1 | 0.731 | 0.733 | 0.791 |
| 2 | 0.768 | 0.802 | 0.887 |
| 3 | 0.817 | 0.866 | 0.830 |
| 4 | 0.900 | 0.843 | 0.855 |
| 5 | 0.711 | 0.841 | 0.724 |
| 6 | 0.727 | 0.765 | 0.671 |
| 7 | 0.932 | 0.663 | 0.583 |
| 8 | 0.920 | 0.762 | 0.904 |
| 9 | 0.875 | 0.769 | 0.684 |
| 10 | 0.752 | 0.731 | 0.649 |
| Rata-Rata | 0.813 | 0.778 | 0.758 |

Langkah selanjutnya adalah melakukan pengukuran ketepatan klasifikasi dengan membentuk *confusion matrix* berdasarkan *fold* dengan nilai *accuracy* tertinggi yang terlah didapatkan pada Tabel 10. Berikut adalah hasil *confusion matrix* pada data *training* dan *testing* dari setiap jenis taksi yang disajikan pada Tabel 11.

**Tabel 11** *Confusion Matrix* dengan Metode SVM Kernel Linear

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Jenis Taksi | Jumlah *Tweet* | *True Positive* | *False Positive* | *True Negative* | *False Negative* |
| *Training* | BlueBird | 2800 | 1010 | 28 | 1682 | 80 |
| GoCar | 3900 | 2129 | 88 | 1584 | 99 |
| GrabCar | 2058 | 1177 | 18 | 831 | 32 |
| *Testing* | BlueBird | 311 | 118 | 18 | 172 | 3 |
| GoCar | 434 | 247 | 57 | 129 | 1 |
| GrabCar | 228 | 129 | 17 | 77 | 5 |

Tabel 11 menunjukkan hasil *confusion matrix* yang didapatkan dengan menggunakan SVM Kernel Linear dengan parameter *C* sebesar 1. Setelah didapatkan *confusion matrix* tersebut, maka langkah selanjutnya adalah melakukan perhitungan ketepatan klasifikasi untuk setiap penyedia jasa transportasi umum taksi yang akan dirangkum menjadi satu tabel berdasarkan data *training* dan data *testing*. Berikut ini merupakan hasil pengukuran ketepatan klasifikasi untuk setiap jenis taksi yang disajikan pada Tabel 12.

**Tabel 12** Ketepatan Klasifikasi Metode SVM Kernel Linear

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Taksi | *Accuracy* | *Sensitivity* | *Specificity* | *G-Mean* | *AUC* |
| *Training* | BlueBird | 0.961 | 0.927 | 0.984 | 0.955 | 0.955 |
| GoCar | 0.952 | 0.956 | 0.947 | 0.951 | 0.951 |
| GrabCar | 0.976 | 0.974 | 0.979 | 0.976 | 0.976 |
| *Testing* | BlueBird | 0.932 | 0.975 | 0.905 | 0.940 | 0.940 |
| GoCar | 0.866 | 0.996 | 0.694 | 0.831 | 0.845 |
| GrabCar | 0.904 | 0.963 | 0.819 | 0.888 | 0.891 |

Dari Tabel 12 menunjukkan hasil ketepatan klasifikasi untuk ketiga jenis taksi yang mendapatkan nilai *accuracy* cukup tinggi untuk data *training* yaitu berada diatas 90%. Kemudian untuk nilai AUC, pada ketiga jenis taksi didapatkan nilai AUC yang berada dikisaran angka 0.9 hingga 1. Jika mengacu pada Tabel 2.3, maka dapat disimpulkan bahwa hasil ketepatan klasifikasi yang didapatkan dapat dikatakan baik sekali jika ditinjau dari nilai AUC. Karena data yang digunakan cenderung *balanced*, maka selanjutnya perbandingan dapat dilakukan dengan menggunakan nilai *accuracy*. Kemudian pada data *testing*, nilai *accuracy* yang didapatkan oleh BlueBird Taxi merupakan nilai *accuracy* yang tertinggi jika dibandingkan dengan taksi yang lainnya, yaitu sebesar 93.2%. Sedangkan untuk jenis taksi GoCar dan GrabCar mendapatkan nilai *accuracy* sebesar 86.6% dan 90.4%.

**4.2 SVM Menggunakan Kernel *Radial Basis Function* (RBF)**

Pada SVM dengan menggunakan Kernel RBF, parameter yang digunakan ada dua, yaitu *C* dan *γ*. Parameter *C* yang digunakan pada Kernel Linear akan dicoba dari 10-2 hingga 104 (Huang, Lee, Lin, & Huang, 2007). Sedangkan untuk *gamma* (*γ*) akan didapatkan melalui hasil percobaan untuk mendapatkan hasil yang paling baik. Dalam analisis dengan menggunakan SVM Kernel RBF, data akan dibagi menjadi data *training* dan data *testing*. Berikut merupakan hasil ketepatan klasifikasi dari metode SVM dengan Kernel RBF yang disajikan pada Tabel 13.

**Tabel 13** Penentuan Parameter SVM Kernel RBF

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jenis Taksi | *C* | *Gamma* (*γ*) | | | | | | |
| 1000 | 100 | 10 | 1 | 0.1 | 0.01 | 0.001 |
| BlueBird | 0.01 | 0.606 | 0.606 | 0.606 | 0.606 | 0.606 | 0.606 | 0.606 |
| 0.1 | 0.699 | 0.699 | 0.702 | 0.719 | 0.640 | 0.606 | 0.606 |
| 1 | 0.777 | 0.777 | 0.779 | 0.845 | 0.769 | 0.646 | 0.606 |
| 10 | 0.777 | 0.777 | 0.781 | 0.859 | 0.857 | 0.779 | 0.646 |
| 100 | 0.777 | 0.777 | 0.781 | 0.859 | 0.841 | 0.855 | 0.781 |
| 1000 | 0.777 | 0.777 | 0.781 | 0.859 | 0.841 | 0.833 | 0.855 |
| 10000 | 0.777 | 0.777 | 0.781 | 0.859 | 0.841 | 0.833 | 0.834 |
| GoCar | 0.01 | 0.571 | 0.571 | 0.571 | 0.570 | 0.570 | 0.570 | 0.570 |
| 0.1 | 0.669 | 0.669 | 0.669 | 0.675 | 0.617 | 0.570 | 0.570 |
| 1 | 0.701 | 0.701 | 0.703 | 0.796 | 0.741 | 0.608 | 0.570 |
| 10 | 0.701 | 0.701 | 0.704 | 0.808 | 0.802 | 0.742 | 0.608 |
| 100 | 0.701 | 0.701 | 0.704 | 0.808 | 0.792 | 0.798 | 0.742 |
| 1000 | 0.701 | 0.701 | 0.704 | 0.808 | 0.789 | 0.787 | 0.798 |
| 10000 | 0.701 | 0.701 | 0.704 | 0.808 | 0.789 | 0.774 | 0.784 |

**Tabel 13** Penentuan Parameter SVM Kernel RBF (Lanjutan)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jenis Taksi | *C* | *Gamma* (*γ*) | | | | | | |
| 1000 | 100 | 10 | 1 | 0.1 | 0.01 | 0.001 |
| GrabCar | 0.01 | 0.590 | 0.590 | 0.590 | 0.590 | 0.590 | 0.590 | 0.590 |
| 0.1 | 0.684 | 0.684 | 0.684 | 0.694 | 0.617 | 0.590 | 0.590 |
| 1 | 0.717 | 0.717 | 0.724 | 0.814 | 0.737 | 0.622 | 0.590 |
| 10 | 0.717 | 0.717 | 0.726 | 0.817 | 0.814 | 0.750 | 0.627 |
| 100 | 0.717 | 0.717 | 0.726 | 0.817 | 0.807 | 0.812 | 0.750 |
| 1000 | 0.717 | 0.717 | 0.726 | 0.817 | 0.808 | 0.796 | 0.813 |
| 10000 | 0.717 | 0.717 | 0.726 | 0.817 | 0.808 | 0.795 | 0.794 |

Dari Tabel 13 didapatkan ketepatan klasifikasi terbaik yang didapat menggunakan SVM Kernel RBF pada masing-masing penyedia jasa transportasi umum taksi dengan menggunakan parameter *C* sebesar 10 dan parameter *gamma* (*γ*) sebesar 1. Setelah didapatkan parameter yang sesuai digunakan dengan untuk metode SVM Kernel RBF, akan dilakukan pemecahan data *tweet* menjadi data *training* dan data *testing* dengan menggunakan metode *k-fold cross validation*. Pada proses ini akan dilakukan dengan menggunakan 10-*fold*. Tabel 14 menunjukkan hasil *accuracy* pada data *testing* dengan menggunakan 10-*fold cross validation* pada setiap penyedia jasa transportasi umum taksi.

**Tabel 14** Nilai *Accuracy* Metode SVM Kernel RBF dengan 10*-fold CV*

|  |  |  |  |
| --- | --- | --- | --- |
| *fold ke-* | BlueBird | GoCar | GrabCar |
| 1 | 0.718 | 0.724 | 0.800 |
| 2 | 0.717 | 0.809 | 0.896 |
| 3 | 0.814 | 0.876 | 0.852 |
| 4 | 0.875 | 0.882 | 0.886 |
| 5 | 0.730 | 0.848 | 0.715 |
| 6 | 0.788 | 0.698 | 0.675 |
| 7 | 0.961 | 0.649 | 0.649 |
| 8 | 0.920 | 0.776 | 0.895 |
| 9 | 0.894 | 0.782 | 0.789 |
| 10 | 0.746 | 0.713 | 0.693 |
| Rata-Rata | 0.816 | 0.776 | 0.785 |

Pada Tabel 14 menunjukkan hasil nilai *accuracy* pada masing-masing jenis taksi dengan menggunakan 10-*fold cross validation*. Nilai *accuracy* rata-rata yang didapatkan untuk BlueBird Taxi adalah sebesar 96.1% dengan nilai *accuracy* tertinggi pada *fold* ke 7. Kemudian untuk GoCar, didapatkan rata-rata nilai *accuracy* sebesar 88.2% dengan nilai *accuracy* tertinggi pada *fold* ke Sedangkan, nilai rata-rata *accuracy* yang didapatkan pada GrabCar adalah sebesar 78.5% dengan nilai *accuracy* yang tertinggi diperoleh pada *fold* ke 2. Setelah didapatkan *fold* dengan nilai *accuracy* tertinggi pada masing-masing jenis taksi, maka selanjutnya akan dilakukan pengukuran ketepatan klasifikasi dengan membentuk *confusion matrix*. Berikut adalah hasil *confusion matrix* pada data *training* dan data *testing* dari setiap taksi yang ditunjukkan pada Tabel 15.

**Tabel 15** *Confusion Matrix* dengan Metode SVM Kernel RBF

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Jenis Taksi | Jumlah *Tweet* | *True Positive* | *False Positive* | *True Negative* | *False Negative* |
| *Training* | BlueBird | 2800 | 1080 | 4 | 1706 | 10 |
| GoCar | 3900 | 2225 | 7 | 1665 | 3 |
| GrabCar | 2056 | 1207 | 4 | 844 | 1 |
| *Testing* | BlueBird | 311 | 119 | 10 | 180 | 2 |
| GoCar | 434 | 241 | 44 | 142 | 7 |
| GrabCar | 230 | 133 | 22 | 73 | 2 |

Pada Tabel 15 didapatkan hasil *confusion matrix* dengan menggunakan SVM Kernel RBF dengan parameter *C* sebesar 10 dan parameter *gamma* (*γ*) sebesar 1. Kemudian akan dilakukan perhitungan ketepatan klasifikasi untuk setiap penyedia jasa transportasi umum taksi yang akan dirangkum dalam satu tabel berdasarkan data *training* dan data *testing*. Berikut adalah hasil pengukuran ketepatan klasifikasi yang disajikan pada Tabel 16.

**Tabel 16** Ketepatan Klasifikasi Metode SVM Kernel RBF

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Taksi | *Accuracy* | *Sensitivity* | *Specificity* | *G-Mean* | *AUC* |
| *Training* | BlueBird | 0.995 | 0.991 | 0.998 | 0.994 | 0.994 |
| GoCar | 0.997 | 0.999 | 0.996 | 0.997 | 0.997 |
| GrabCar | 0.998 | 0.999 | 0.995 | 0.997 | 0.997 |
| *Testing* | BlueBird | 0.961 | 0.983 | 0.947 | 0.965 | 0.965 |
| GoCar | 0.882 | 0.972 | 0.763 | 0.861 | 0.868 |
| GrabCar | 0.896 | 0.985 | 0.768 | 0.870 | 0.877 |

Dari Tabel 16 didapatkan hasil ketepatan klasifikasi untuk ketiga jenis taksi menunjukkan nilai *accuracy* untuk data *training* cukup tinggi yaitu berada diatas 90%. Kemudian untuk nilai AUC, didapatkan hasil nilai AUC untuk ketiga jenis taksi berada diatas 0.9. Jika mengacu pada Tabel 2.3, maka dapat diartikan ketepatan klasifikasi pada data *training* untuk ketiga jenis taksi adalah baik sekali. Karena data yang digunakan cenderung *balanced*, maka selanjutnya perbandingan dapat dilakukan dengan menggunakan nilai *accuracy*. Selanjutnya untuk data *testing*, didapatkan nilai *accuracy* untuk penyedia jasa transportasi umum BlueBird Taxi merupakan nilai *accuracy* tertinggi jika dibandingkan dengan jenis taksi yang lain, yaitu sebesar 96.1%. Sedangkan untuk penyedia jasa transportasi umum taksi GoCar dan GrabCar, didapatkan nilai *accuracy* sebesar 88.2% dan 89.6%.

**4.3 Pengukuran Performa SVM**

Dari hasil analisis SVM dengan menggunakan Kernel Linear dan Kernel RBF didapatkan perbedaan parameter dan hasil ketepatan klasifikasi pada masing-masing Kernel. Sehingga akan dilakukan perbandingan ketepatan klasifikasi antara metode SVM dengan menggunakan Kernel Linear dan Kernel RBF dengan menggunakan rata-rata nilai *accuracy* pada data *testing* dengan 10-*fold cross validation* yang akan disajikan pada Tabel 17.

**Tabel 17** Perbandingan Ketepatan Klasifikasi Antara SVM Kernel Linear dan Kernel RBF

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Jenis Taksi | Kernel Linear | | Kernel RBF | |
| **Parameter** | ***Accuracy*** | **Parameter** | ***Accuracy*** |
| BlueBird | *C* = 1 | 0.813 | *C* = 10  *γ* = 1 | 0.816 |
| GoCar | 0.778 | 0.776 |
| GrabCar | 0.758 | 0.785 |

Berdasarkan Tabel 17 dapat disimpulkan bahwa dari perbandingan metode SVM dengan menggunakan Kernel Linear dan Kernel RBF, didapatkan metode SVM yang memiliki ketepatan klasifikasi terbaik atau yang tinggi adalah dengan menggunakan metode SVM Kernel RBF dengan parameter yang digunakan yaitu parameter *C* sebesar 10 dan parameter *gamma* (*γ*) sebesar 1 pada jenis taksi BlueBird Taxi dan GrabCar. Namun, pada jenis taksi GoCar ketepatan klasifikasi yang tinggi didapatkan dengan menggunakan metode SVM Kernel Linear dengan menggunakan parameter C sebesar 1. Akan tetapi, nilai yang didapatkan dengan menggunakan metode SVM Kernel Linear dan SVM Kernel RBF tidak berbeda jauh, sehingga metode SVM yang terpilih adalah metode SVM Kernel RBF dengan parameter yang digunakan yaitu parameter *C* sebesar 10 dan parameter *gamma* (*γ*) sebesar 1.

Hasil ketepatan klasifikasi terbaik dari ketiga penyedia jasa transportasi umum taksi pada metode SVM Kernel RBF menggunakan nilai parameter yang sama yaitu menggunakan parameter *C* sebesar 10 dan parameter *gamma* (*γ*) sebesar 1. Kemudian parameter *gamma* tersebut disubstitusikan pada persamaan Kernel RBF yang terdapat pada Tabel 2.1, sehingga fungsi Kernel RBF pada ketiga taksi yang terbentuk sama yaitu sebagai berikut.



Fungsi *hyperplane* didapatkan dengan mensubstitusikan fungsi Kernel pada persamaan (2.25). Sehingga didapatkan fungsi *hyperplane* dari data *training* untuk masing-masing penyedia jasa transportasi umum taksi yang disajikan pada Tabel 18.

**Tabel 18** Persamaan *Hyperplane* pada Setiap Taksi

|  |  |
| --- | --- |
| Jenis Taksi | Persamaan *Hyperplane* |
| BlueBird |  |
| GoCar |  |
| GrabCar |  |

Pada persamaan *hyperplane* pada Tabel 18, *αi*merupakan vektor koefisien atau *lagrange multiplier* dari *support vector* yang berukuran (2766,1) untuk BlueBird Taxi (nilainya terlampir pada Lampiran 18), dan berukuran (3883,1) untuk GoCar (nilainya terlampir pada Lampiran 19), serta berukuran (2044,1) untuk GrabCar (nilainya terlampir pada Lampiran 20). Kemudian *y*i merupakan label kelas yang memiliki dua nilai yaitu +1 dan -1, serta *x* merupakan nilai *input* yang akan diklasifikasikan. Nilai 0.2119, -0.1029, dan -0.1097 merupakan nilai bias untuk masing-masing persamaan *hyperplane* jenis taksi.  merupakan persamaan *hyperplane* untuk BlueBird Taxi,  merupakan persamaan *hyperplane* untuk GoCar, dan  merupakan persamaan *hyperplane* untuk GrabCar.

**4.5 Perbandingan Hasil Klasifikasi Antara Metode NBC dan SVM**

Setelah mengetahui hasil dari masing-masing ketepatan klasifikasi pada kedua metode, maka langkah selanjutnya adalah membandingkan. Berikut merupakan perbandingan antara kedua metode berdasarkan rata-rata nilai *accuracy* dari data *testing* dengan 10-*fold cross validation* yang ditunjukkan pada Tabel 19.

Berdasarkan perbandingan metode NBC dan SVM untuk masing-masing jenis taksi pada Tabel 19, jika dilihat dari rata-rata nilai *accuracy* pada data *testing*, maka dapat disimpulkan bahwa metode SVM dengan Kernel RBF lebih baik jika dibandingkan dengan metode NBC. Sehingga metode yang cocok digunakan untuk pengklasifikasian pada penelitian ini adalah metode SVM dengan Kernel RBF dengan menggunakan parameter *C* sebesar 10 dan *gamma* (*γ*) sebesar 1. Ketepatan klasifikasi yang didapatkan untuk masing-masing jenis taksi dengan metode SVM Kernel RBF didapatkan rata-rata nilai *accuracy* pada data *testing* sebesar 81.6% untuk BlueBird Taxi, kemudian untuk GoCar sebesar 77.6%, sedangkan untuk GrabCar sebesar 78.5%.

**Tabel 19** Perbandingan Metode NBC dan SVM

|  |  |  |
| --- | --- | --- |
| Metode | Taksi | *Accuracy* |
| NBC | BlueBird | 0.809 |
| GoCar | 0.741 |
| GrabCar | 0.782 |
| SVM Kernel RBF | BlueBird | 0.816 |
| GoCar | 0.776 |
| GrabCar | 0.785 |

**4.6 Visualisasi *Word Cloud***

Visualisasi data teks dengan menggunakan *word cloud* bertujuan untuk mengetahui kata-kata yang paling sering muncul pada data. Pada penelitian ini, visualisasi *word cloud* digunakan untuk visualisasi *tweet* berdasarkan kategori sentimennya, sehingga dapat diketahui kata-kata yang sering muncul pada setiap sentimen tentang masing-masing jenis taksi. Ukuran *font* pada *word cloud* menunjukkan frekuensi kemunculan kata. Jadi, semakin besar ukuran *font* berarti semakin besar frekuensi kemunculan kata tersebut.

Visualisasi *word cloud* akan dilakukan dengan membandingkan antara data *tweet* yang memiliki sentimen positif dan data *tweet* yang memiliki sentimen negatif. Perbandingan tersebut dilakukan dengan tujuan mengetahui penyebab mayoritas pengguna menilai masing-masing pelayanan dari penyedia jasa taksi berdasarkan kategori sentimen. Hasil visualisasi dengan *word cloud* untuk masing-masing kategori sentimen pada BlueBird Taxi ditunjukkan pada Gambar 7.

|  |  |
| --- | --- |
| (a) | (b) |

**Gambar 7** *Word Cloud* BlueBird Taxi Sentimen Positif (a) dan Sentimen Negatif (b)

Pada Gambar 7 menunjukkan kata kunci yang mengarah pada BlueBird Taxi, terlihat bahwa kata kunci terbanyak yang digunakan untuk respon pengguna pada sentimen positif terbesar yaitu kata ‘tarif’. Hal tersebut disebabkan karena para pengguna sering membandingkan tarif antara taksi konvensional dengan tarif taksi berbasis *online* yang kenyataannya tarif dari BlueBird Taxi memang lebih murah dibandingkan taksi berbasis *online* menurut pengguna. Sedangkan, pada sentimen negatif kata kunci yang terbesar adalah ‘app’, dikarenakan memang aplikasi dari BlueBird Taxi masih kalah jika dibandingkan dengan aplikasi dari taksi berbasis *online*, sehingga banyak dikeluhkan oleh para pengguna. Kemudian, didapatkan pula kata kunci ‘driver’ dan ‘supir’ yang mendapat banyak perhatian dari para pengguna, sehingga banyak pengguna yang berpendapat positif atau negatif terhadap driver atau supir dari BlueBird Taxi. Kata-kata lain pada sentimen positif dan sentimen negatif mempunyai frekuensi yang jauh kecil yang ditandai dengan ukuran *font* yang berbeda. Selanjutnya, dilakukan visualisasi *word cloud* untuk masing-masing kategori sentimen pada GoCar yang disajikan pada Gambar 8.

Berdasarkan Gambar 8 didapatkan bahwa kata kunci terbesar untuk GoCar dari sentimen positif dan negatif sama, yaitu kata ‘driver’. Hal tersebut memang sangat dirasakan oleh para pengguna GoCar, karena terkadang pengguna GoCar mendapatkan driver yang memang baik dan memberikan pelayanan yang memuaskan, sehingga pengguna memberikan pendapatnya tentang pelayanan yang diperoleh dari driver. Terkadang pula pengguna mendapatkan driver yang foto dan plat nomor kendaraannya tidak sesuai dengan aplikasi, sehingga hal tersebut juga sering dikeluhkan oleh pengguna. Terlebih lagi, banyak juga beredar pemberitaan tentang kasus-kasus yang dilakukan oleh driver GoCar.

|  |  |
| --- | --- |
| (a) | (b) |

**Gambar 8** *Word Cloud* GoCar Sentimen Positif (a) dan Sentimen Negatif (b)

Selain itu, pada kategori sentimen positif didapatkan pula bahwa pengguna jasa layanan taksi GoCar juga sering menggunakan kata ‘uber’ untuk menyampaikan pendapatnya tentang pelayanan dari GoCar, yang berarti banyak pengguna yang membandingkan pelayanan dari penyedia jasa taksi GoCar dengan penyedia jasa taksi lain yaitu Uber. Selanjutnya, dilakukan visualisasi *word cloud* pada masing-masing kategori sentimen dari penyedia jasa transportasi umum GrabCar yang disajikan pada Gambar 9.

Dari Gambar 9 didapatkan kata kunci dari sentimen positif untuk GrabCar adalah kata ‘ngobrol’. Hal tersebut karena banyak pengguna yang suka dengan driver GrabCar yang suka mengajak berinteraksi selama diperjalanan, sehingga hal tersebut menjadi nilai tambah untuk citra dari GrabCar. Namun pada sentimen negatif, didapatkan kata kunci yang paling banyak digunakan adalah ‘driver’. Permasalahan driver ini memang banyak dikeluhkan oleh pengguna GrabCar, lantaran terkadang pengguna GrabCar sulit mendapatkan driver jika sedang melakukan pemesanan, atau terkadang driver tidak datang untuk memenuhi pesanan dan tiba-tiba membatalkan pesanan yang dilakukan pelanggan. Sehingga hal tersebut menjadi sebuah permasalahan yang seharusnya diperhatikan oleh pihak GrabCar. Selain itu, sama seperti halnya GoCar, pengguna jasa transportasi umum taksi juga sering membandingkan GrabCar dengan jasa transportasi umum taksi berbasis *online* yang lain yaitu Uber. Hal tersebut ditunjukkan dari adanya kata ‘uber’ yang cukup dominan pada *word cloud*. Namun berbeda dengan GoCar, sebagian besar pengguna jasa transportasi umum taksi GrabCar lebih banyak mengunggah *tweet* dengan berisi sentimen negatif pada saat membandingkan antara penyedia jasa transportasi umum taksi berbasis *online* Uber dengan GrabCar.

|  |  |
| --- | --- |
| (a) | (b) |

**Gambar 9** *Word Cloud* GrabCar Sentimen Positif (a) dan Sentimen Negatif (b)

**4.7 *Social Network Analysis* (SNA)**

Setelah dilakukan analisis klasifikasi dengan menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine*, maka selanjutnya dilakukan anailisis dengan menggunakan *Social Network Analysis* (SNA) untuk mengetahui pengguna Twitter yang berpengaruh berdasarkan topik tentang taksi konvensional dan taksi berbasis *online* dengan menganalisa hasil representasi *graph* yang terbentuk. *Graph type* dari *network* yang digunakan adalah *directed*, yaitu jenis *graph* yang mempertimbangkan arah interaksi. Berikut adalah hasil analisis *Social Network Analysis* untuk masing-masing jenis taksi.

**4.7.1 *Social Network Analysis* BlueBird Taxi**

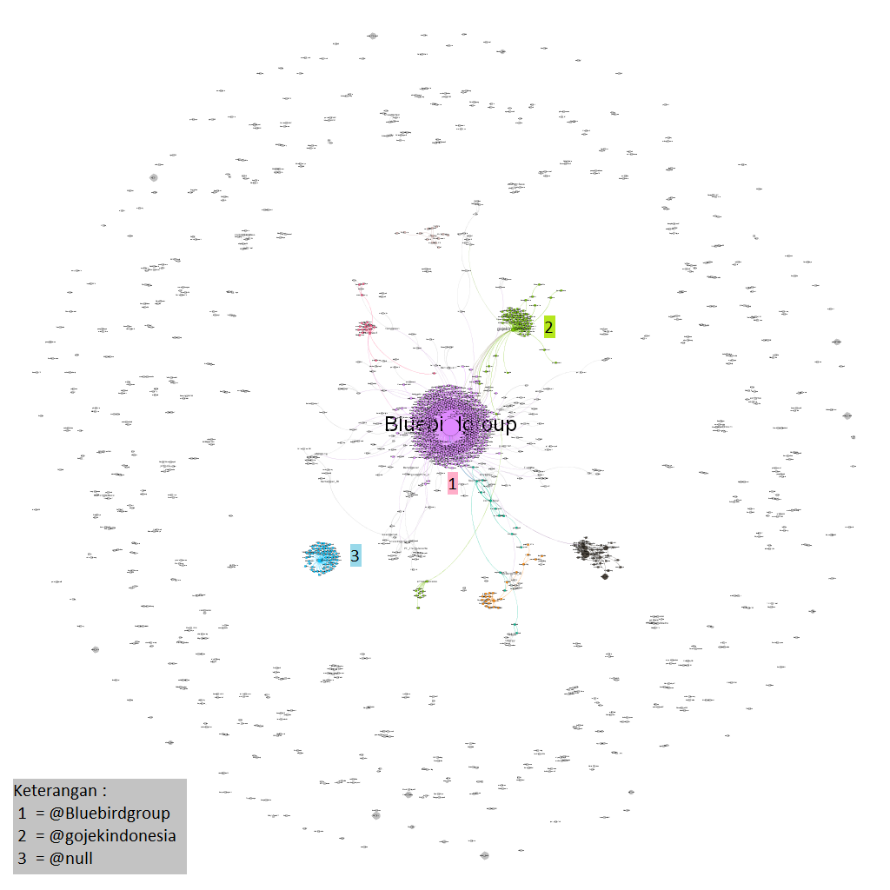
Berikut ini adalah hasil observasi SNA pada *keyword* BlueBird Taxi dalam media sosial Twitter. Sebelumnya, dilakukan analisis dari data perbandingan jaringan dari BlueBird Taxi yang ditunjukkan pada Tabel 20.

**Tabel 20** Data Perbandingan Jaringan BlueBird Taxi

|  |  |  |
| --- | --- | --- |
| *Degree Range* | *Nodes* | *Edges* |
| 0 | 4623 | 1651 |
| 1 | 1721 | 1651 |
| 2 | 417 | 523 |
| 3 | 150 | 213 |

Dari Tabel 20 didapatkan jumlah *node* dan *edge* dari jaringan BlueBird Taxi. Dapat diketahui bahwa jika tanpa menggunakan *filter degree range*, jumlah *nodes* dan *edges* yang didapatkan adalah sebanyak 4623 *nodes* dan 1651 *edges*. Sedangkan, jika menggunakan *filter degree range*, jumlah *nodes* yang didapatkan menjadi 1721 dan jumlah *edges* tetap. Hal tersebut karena jika menggunakan *filter degree range* minimal satu, maka *nodes* yang tidak memiliki interaksi *edges* minimal satu akan dihilangkan, sehingga dapat mengurangi *noise* pada *graph*. Berikut adalah hasil visualisasi *network* yang terbentuk untuk jaringan BlueBird Taxi dengan menggunakan *filter degree range* satu yang ditunjukkan pada Gambar 10.

Gambar 10 merupakan visualisasi *network* untuk BlueBird Taxi dengan menggunakan tipe jaringan *Frunchterman Reingold* dan *ForceAtlas 2*. Tipe tersebut merupakan tipe yang umum digunakan dalam pembuatan *graph* untuk SNA. Dari Gambar 10 didapatkan nilai *density* sebesar 0.001. *Density* adalah kepadatan *graph* suatu *network* yang menunjukkan jumlah hubungan yang hadir dalam suatu kelompok. Ketika menghitung *density*, SNA melihat bagaimana erat hubungan seseorang yang satu dengan yang lain. Nilai *density* sebesar 0.001 menunjukkan bahwa *network* BlueBird Taxi memiliki kepadatan yang kecil atau renggang. *Density* yang kecil atau renggang ini terjadi karena rendahnya interaksi antar *account*, baik berupa *mention*, *quote retweet*, atau *reply* yang dilakukan antar *node* dalam jaringan. Berikut adalah hasil analisis *centrality* dari Gambar 10 yang ditunjukkan pada Tabel 21.



**Gambar 10** Visualisasi *Network* BlueBird Taxi

Analisis *centrality* pada Tabel 21 bertujuan untuk menemukan *account* yang paling berperan dalam sebuah *network*. *Metric* yang digunakan dalam penentuan *centrality* ini adalah *degree centrality*, *closeness centrality*, *betweenness centrality*, dan *eigenvector centrality*. Analisis *degree centrality* menentukan *account* yang paling berperan berdasarkan banyaknya *edge* atau hubungan yang terjadi antara sebuah *node* dengan *node* yang lainnya. *Account* @Bluebirdgroup memiliki nilai *degree centrality* tertinggi yaitu 635 (pada Gambar 10 ditunjukkan dengan angka 1). Hal ini dapat diartikan bahwa @Bluebirdgroup memiliki hubungan atau interaksi yang berupa *mention*, *quote retweet*, atau *reply* antar *node* lain sebanyak 635 kali. Sedangkan, yang kedua adalah *account* @gojekindonesia dengan nilai *degree centrality* sebesar 82 (pada Gambar 10 ditunjukkan dengan angka 2), dan yang ketiga adalah *account* @null dengan nilai *degree centrality* sebesar 67 (pada Gambar 10 ditunjukkan dengan angka 3).

**Tabel 21** Analisis *Centrality* dari Jaringan BlueBird Taxi

|  |  |  |
| --- | --- | --- |
| *Centrality* | *Username* | *Score* |
| *Degree Centrality* | @Bluebirdgroup | 635 |
| @gojekindonesia | 82 |
| @null | 67 |
| *Closeness Centrality* | @PT\_TransJakarta | 0.888889 |
| @ancoltmnimpian | 0.857143 |
| @ekyology | 0.833333 |
| *Betweenness Centrality* | @gojekindonesia | 990.0 |
| @WayToLombok | 189.0 |
| @Lostpacker | 146.0 |
| *Eigenvector Centrality* | @Bluebirdgroup | 1.0 |
| @gojekindonesia | 0.116251 |
| @null | 0.102723 |

Analisis *closeness centrality* digunakan untuk melihat *node-node* yang dapat menjangkau *node* lainnya dengan jalur yang lebih pendek. Semakin mendekati 1 atau sama dengan 1, maka semakin dekat *node* tersebut dengan *node* lain. Agar lebih mudah dipahami dan mengurangi bias, maka semua *node* yang memiliki nilai *closeness centrality* sebesar 1 tidak diperhatikan. Terdapat tiga *account* dengan nilai *closeness centrality* tertinggi, yaitu @PT\_TransJakarta, @ancoltmnimpian, dan @ekyology. Nilai *closeness centrality* untuk peringkat pertama yaitu 0.888889 dan peringkat kedua sebesar 0.857143. selisih nilai antara peringkat pertama dan kedua yaitu sebesar 0.031746. Hal ini mengakibatkan *node-node* dengan nilai *closeness centrality* 0.888889 lebih cepat dan lebih mudah dalam berkomunikasi dengan *node* lain ketika melakukan *mention*, *quote retweet*, atau *reply* tanpa melalui banyak perantara yang dilalui. Akan tetapi, *node-node* dengan nilai *closeness centrality* 0.857143 sama baiknya dalam hal kecepatan dan kemudahan berkomunikasi dengan *node* lain tanpa melalui banyak perantara yang dilalui.

Analisis *betweenness centrality* bertujuan untuk mengetahui posisi *node* dalam *network*, dimana *node* tersebut tidak boleh hilang. Jika *node* tersebut hilang maka akan terjadi gangguan komunikasi dalam *network*. *Node* dengan *username* @gojekindonesia memiliki nilai *betweenness centrality* tertinggi dalam *network*. Artinya, @gojekindonesia adalah *account* penghubung atau jembatan dari seluruh aliran informasi dalam percakapan mengenai BlueBird Taxi. Kemudian analisis *eigenvector centrality* bertujuan untuk menemukan *account* yang memiliki performa paling baik dalam *network*. Nilai *eigenvector centrality* paling besar dimiliki oleh *username* @Bluebirdgroup sebesar 1. Selanjutnya diikuti oleh *username* @gojekindonesia dan @null, yang artinya *username* @Bluebirdgroup memiliki interaksi baik berupa *mention*, *quote retweet*, dan *reply* dengan *node-node* tertentu seperti @gojekindonesia dan @null yang mempunyai interaksi tinggi atau menjadi sumber informasi.

**4.7.2 *Social Network Analysis* GoCar**

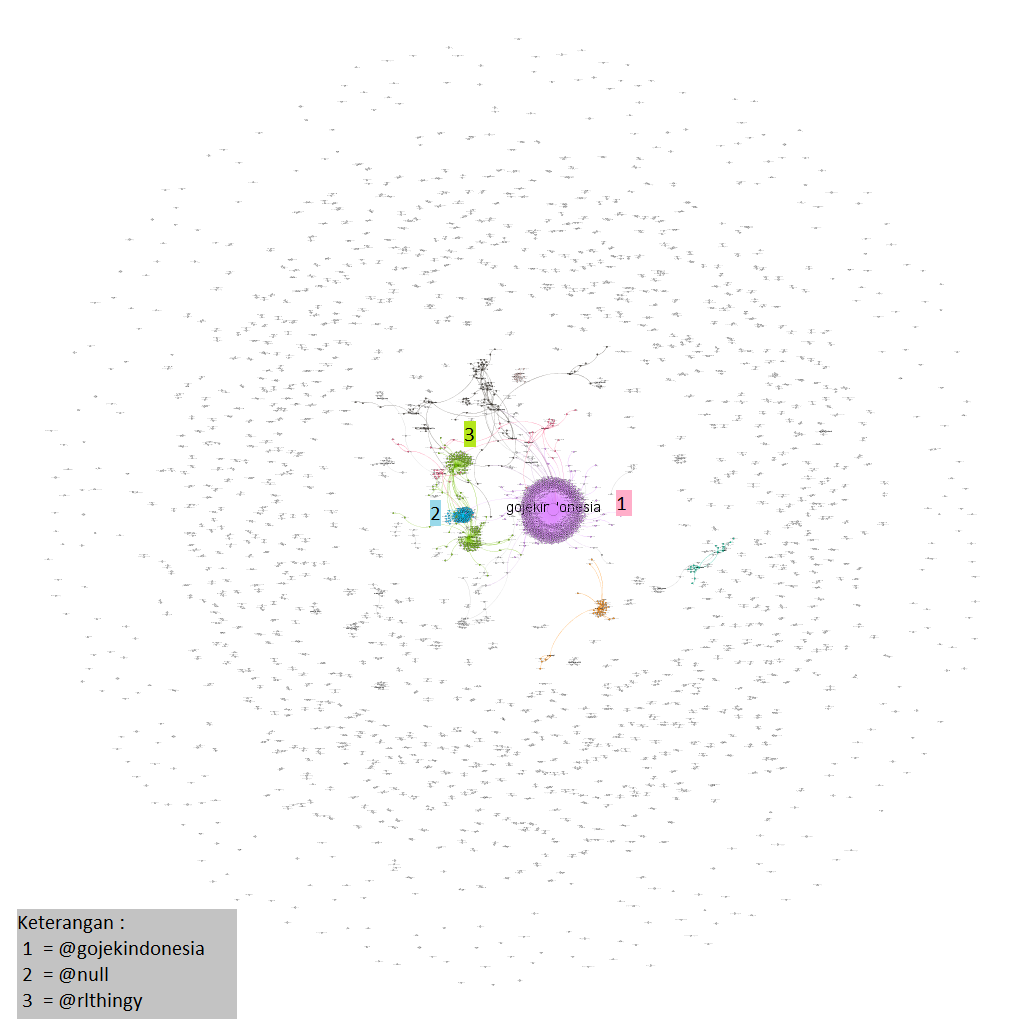
Berikut ini adalah hasil obsevasi SNA pada *keyword* GoCar dalam media sosial Twitter. Sebelumnya, dilakukan analisis dari data perbandingan jaringan yang ditunjukkan pada Tabel 22.

**Tabel 22** Data Perbandingan Jaringan GoCar

|  |  |  |
| --- | --- | --- |
| *Degree Range* | *Nodes* | *Edges* |
| 0 | 13831 | 3142 |
| 1 | 4027 | 3142 |
| 2 | 823 | 819 |
| 3 | 144 | 150 |

Pada Tabel 22 didapatkan jumlah *node* dan *edge* dari jaringan GoCar yaitu sebanyak 13831 *nodes* dan 3142 *edges*. Kemudian dilakukan proses untuk mengurangi *noise* dari *node* yang tidak memiliki interaksi dengan menggunakan *filter degree range*. Nilai minimal yang digunakan dalam *filter degree range* adalah satu, sehingga *node* yang muncul dalam *graph* adalah *node* yang memiliki minimal satu interaksi atau satu *edge*. Setelah dilakukan *filter degree range*, maka didapatkan jumlah *nodes* sebanyak 4027 *nodes* dan jumlah *edge* tetap. Berikut adalah hasil visualisasi *network* yang terbentuk untuk jaringan GoCar dengan *filter degree range* yang disajikan pada Gambar 11.

Gambar 11 merupakan visualisasi *network* untuk jaringan GoCar dengan menggunakan tipe jaringan *Frunchterman Reingold* dan *ForceAtlas 2*. Kedua tipe tersebut dipilih karena merupakan tipe yang umum digunakan dalam pembuatan *graph* untuk SNA. Nilai *density* yang didapatkan untuk jaringan GoCar sebesar 0. Nilai *density* sebesar 0 menandakan bahwa jaringan GoCar memiliki kepadatan yang kecil atau renggang. Nilai *density* yang kecil atau renggang ini terjadi karena rendahnya interaksi antar *account*, baik berupa *mention*, *quote retweet*, atau *reply* yang dilakukan antar *node* dalam jaringan. Berikut ini adalah hasil analisis *centrality* yang bertujuan untuk menemukan *account* yang paling berperan dalam sebuah *network* yang ditunjukkan pada Tabel 23.



**Gambar 11** Visualisasi *Network* GoCar

Pada hasil analisis *degree centrality* pada Tabel 23 didapatkan bahwa *account* @gojekindonesia memiliki nilai *degree centrality* tertinggi yaitu 735 (pada Gambar 11 ditunjukkan dengan angka 1), yang artinya *account* @gojekindonesia memiliki hubungan atau interaksi yang berupa *mention*, *quote retweet*, atau *reply* antar *node* lain sebanyak 735 kali. Kemudian yang kedua adalah *account* @null dengan nilai *degree centrality* sebesar 96 (pada Gambar 11 ditunjukkan dengan angka 2), dan yang ketiga adalah *account* @rlthingy dengan nilai *degree centrality* sebesar 87 (pada Gambar 11 ditunjukkan dengan angka 3). Selanjutnya dari nilai *closeness centrality* didapatkan tiga *account* dengan nilai tertinggi yaitu @gojekindonesia, @ffarliani, dan @qaqqah. Nilai *closeness centrality* pada peringkat pertama sebesar 0.983871 dan peringkat kedua sebesar 0.75 dengan selisih antara peringkat pertama dan kedua sebesar 0.233871 mengakibatkan *node-node* dengan nilai *closeness centrality* sebesar 0.984871 lebih cepat dan lebih mudah dalam berkomunikasi dengan *node* lain ketika melakukan *mention*, *quote retweet*, atau *reply* tanpa melalui banyak perantara. Sedangkan, *node-node* dengan nilai *closeness centrality* 0.75 perlu melalui banyak perantara untuk melakukan *mention*, *quote retweet*, dan *reply*.

**Tabel 23** Analisis *Centrality* dari Jaringan GoCar

|  |  |  |
| --- | --- | --- |
| *Centrality* | *Username* | *Score* |
| *Degree Centrality* | @gojekindonesia | 735 |
| @null | 96 |
| @rlthingy | 87 |
| *Closeness Centrality* | @gojekindonesia | 0.983871 |
| @ffarliani | 0.75 |
| @qaqqah | 0.75 |
| *Betweenness Centrality* | @gojekindonesia | 45095.0 |
| @JennyJusuf | 1815.0 |
| @kejO\_Online | 1778.666667 |
| *Eigenvector Centrality* | @gojekindonesia | 1.0 |
| @Ismailm62599808 | 0.110057 |
| @ari\_tengol | 0.109411 |

Analisis *betweenness centrality* digunakan untuk mengetahui posisi *node* dalam *network*. Pada Tabel 23 didapatkan nilai *betweenness centrality* tertinggi pada *account* @gojekindonesia. Artinya, *account* @gojekindonesia merupakan *account* penghubung atau jembatan dari seluruh aliran informasi dalam percakapan mengenai GoCar. Hal ini mengakibatkan banyak *node* lain yang bergantung pada *tweet* yang di-*posting* oleh @gojekindonesia, karena *node* tersebut adalah sebuah jembatan bagi *node* lain sebagai sumber informasil terkait percakapan mengenai GoCar. Selanjutnya, dari analisis *eigenvector centrality*, nilai paling besar dimiliki oleh *account* @gojekindonesia sebesar 1. Hal tersebut menunjukkan bahwa *account* @gojekindonesia memiliki interaksi baik berupa *mention*, *quote retweet*, dan *reply* dengan *node-node* tertentu.

**4.7.3 *Social Network Analysis* GrabCar**

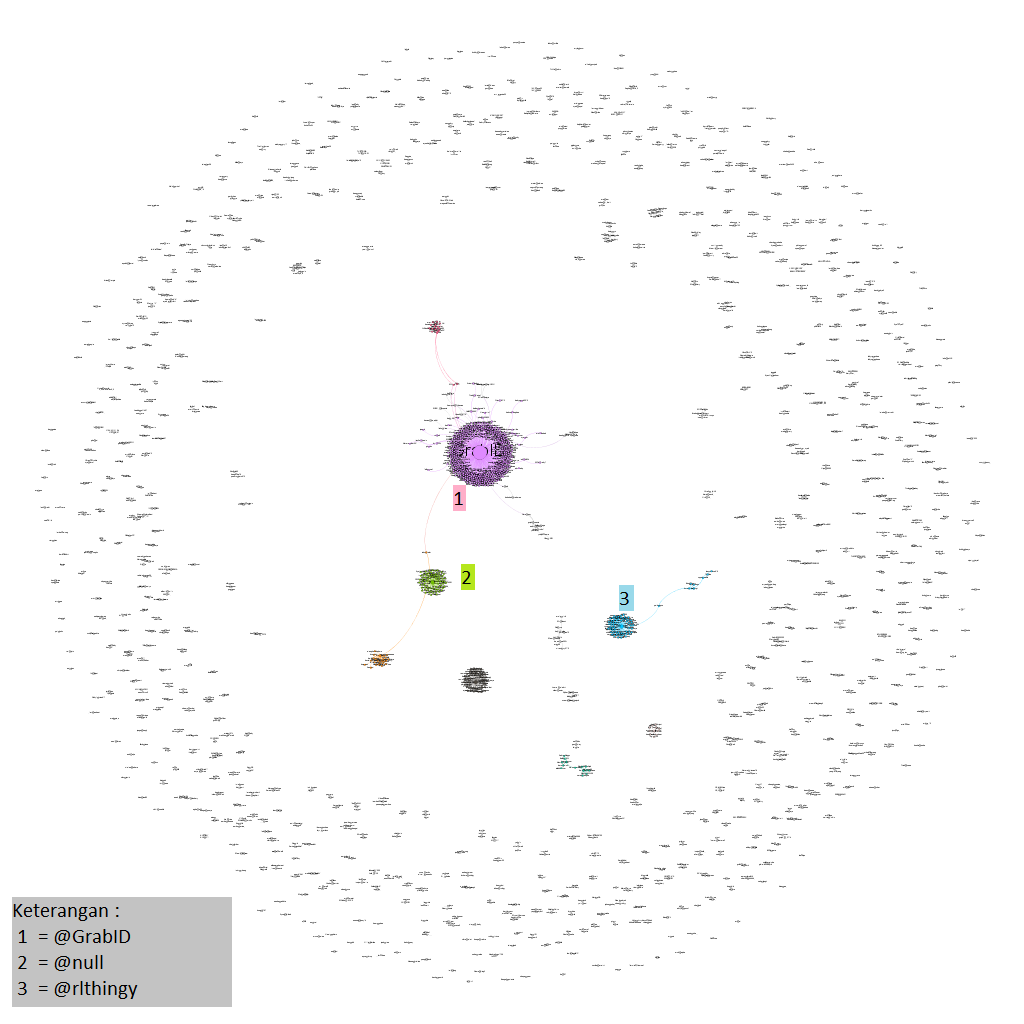
Berikut ini adalah hasil observasi SNA pada *keyword* GrabCar dalam media sosial Twitter. Sebelum dilakukan analisis lebih lanjut, maka terlebih dahulu dilakukan perbandingan jaringan dari GrabCar yang ditunjukkan pada Tabel 24.

**Tabel 24** Data Perbandingan Jaringan GrabCar

|  |  |  |
| --- | --- | --- |
| *Degree Range* | *Nodes* | *Edges* |
| 0 | 13674 | 1765 |
| 1 | 2301 | 1765 |
| 2 | 406 | 371 |
| 3 | 67 | 47 |

Berdasarkan Tabel 24 didapatkan jumlah *node* dan *edge* dalam jaringan GrabCar sebanyak 13674 *nodes* dan 1765 *edges*. Jumlah *node* sebanyak 13674 *nodes* tersebut masih mengandung *node* yang tidak memiliki interaksi atau *edge*, sehingga dilakukan proses *filtering* untuk mengurangi *noise* pada *graph* dengan menggunakan *filter degree range*. Nilai f*ilter degree range* yang digunakan adalah minimal satu, yaitu hanya memunculkan *node* yang memiliki interaksi sebanyak satu. Sehingga setelah dilakukan proses *filtering*, didapatkan jumlah *node* sebanyak 2301 *nodes* dengan jumlah *edge* tetap. Berikut adalah hasil visualisasi *network* yang terbentuk untuk jaringan GrabCar dengan menggunakan *filter degree range* minimal satu yang ditunjukkan pada Gambar 12.

Visualisasi *network* untuk GrabCar pada Gambar 12 menggunakan tipe jaringan *Frunchterman Reingold* dan *ForceAtlas 2*. Kedua tipe tersebut digunakan karena merupakan tipe yang umum untuk pembuatan *graph* SNA. Dari *graph* jaringan pada Gambar 12 didapatkan nilai *density* sebesar 0, yang artinya jaringan GrabCar memiliki kepadatan yang kecil atau renggang. Nilai *density* yang kecil atau renggang terjadi diakibatkan oleh rendahnya interaksi antar *account*, baik berupa *mention*, *quote retweet*, atau *reply* yang dilakukan antar *node* dalam jaringan tersebut. Berikut adalah hasil analisis *centrality* dari Gambar 12 yang ditunjukkan pada Tabel 25.



**Gambar 12** Visualisasi *Network* GrabCar

Analisis *centrality* pada Tabel 25, *metric* yang digunakan dalam penentuan *centrality* adalah *degree centrality*, *closeness centrality*, *betweenness centrality*, dan *eigenvector centrality*. Dari Tabel 21 didapatkan nilai *degree centrality* yang terbesar adalah pada *account* @GrabID yaitu sebesar 482 (pada Gambar 12 ditunjukkan dengan angka 1). Nilai *degree centrality* sebesar 482 menunjukkan bahwa *account* @GrabID memiliki hubungan atau interaksi yang berupa *mention*, *quote retweet*, atau *reply* antar *node* lain sebanyak 482 kali. Kemudian pada urutan kedua adalah *account* @null yaitu sebesar 73 (pada Gambar 12 ditunjukkan dengan angka 2), dan yang ketiga adalah *account* @rlthingy dengan nilai *degree centrality* sebesar 61 (pada Gambar 12 ditunjukkan dengan angka 3). Kemudian untuk analisis *closeness centrality* didapatkan tiga *account* dengan nilai tertinggi, yaitu @GrabID, @be\_em\_we323i, dan @nengbiker. Pada *account* @GrabID didapatkan nilai *closeness centrality* sebesar 0.948052 dan memiliki selisih sebesar 0.114719 dengan *account* @be\_em\_we323i. Hal ini mengakibatkan *node-node* dengan nilai *closeness centrality* sebesar 0.948052 lebih cepat dan lebih mudah dalam berkomunikasi dengan *node* lain ketika melakukan *mention*, *quote retweet*, atau *reply* tanpa melalui banyak perantara yang dilalui. Sedangkan, *node-node* dengan nilai *closeness centrality* sebesar 0.833333 perlu melalui banyak perantara.

**Tabel 25** Analisis *Centrality* dari Jaringan GrabCar

|  |  |  |
| --- | --- | --- |
| *Centrality* | *Username* | *Score* |
| *Degree Centrality* | @GrabID | 482 |
| @null | 73 |
| @rlthingy | 61 |
| *Closeness Centrality* | @GrabID | 0.948052 |
| @be\_em\_we323i | 0.833333 |
| @nengbiker | 0.8 |
| *Betweenness Centrality* | @GrabID | 30279.0 |
| @Heruu96 | 830 |
| @yonpurba | 415.0 |
| *Eigenvector Centrality* | @GrabID | 1.0 |
| @rik\_agus | 0.159444 |
| @Heruu96 | 0.157865 |

Selanjutnya dilakukan analisis *betweenness centrality* yang bertujuan untuk mengetahui posisi *node* dalam *network*. Pada Tabel 25 didapatkan *node* dengan *username* @GrabID memiliki nilai yang paling tinggi yaitu sebesar 30279. Hal tersebut menunjukkan bahwa @GrabID merupakan *account* penghubung atau jembatan dari seluruh aliran informasi di dalam percakapan mengenai GrabCar. Begitu pula pada analisis *eigenvector centrality*, didapatkan bahwa *account* @GrabID merupakan *acoount* yang memiliki nilai *eigenvector centrality* tertinggi, kemudian diikuti oleh *account* @arik\_agus dan @Heruu96. Sehingga dapat disimpulkan bahwa dari analisis *eigenvector centrality* diketahui bahwa *account* @GrabID memiliki interaksi baik berupa *mention*, *quote retweet*, dan *reply* dengan *node-node* tertentu, seperti @arik\_agus dan @Heruu96 yang mempunyai interaksi tinggi atau menjadi sumber informasi.

**4.7.4 *Social Network Analysis* Taksi Konvensional dan Taksi Berbasis *Online***

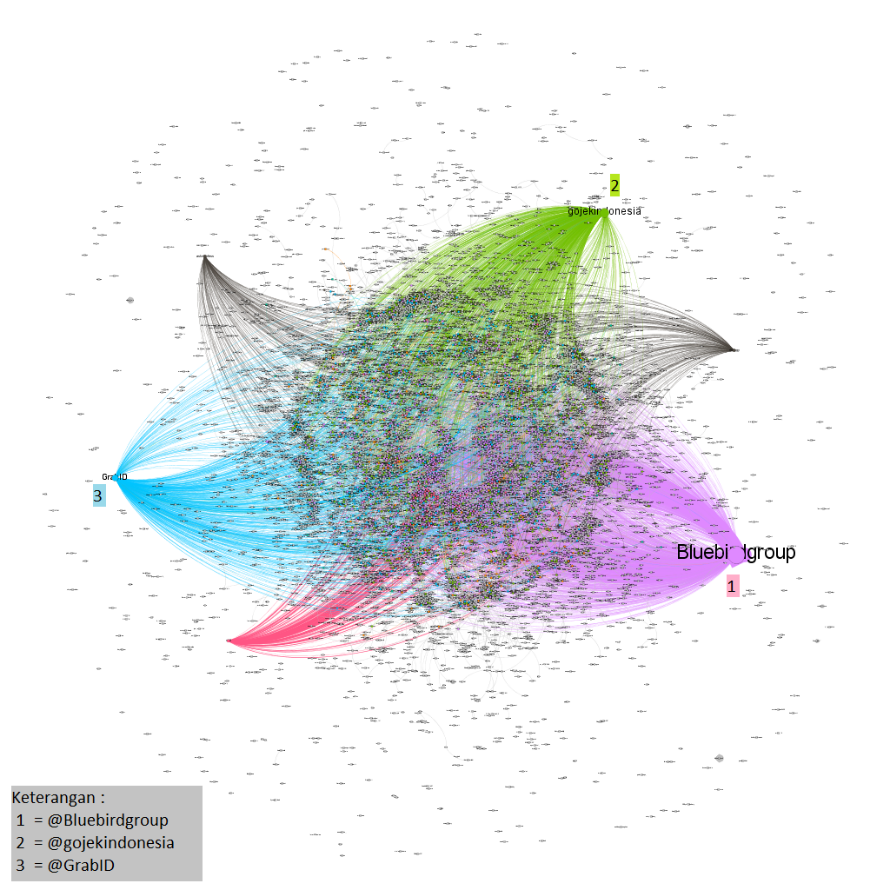
Setelah dilakukan *Social Network Analysis* pada masing-masing penyedia jasa transportasi umum taksi, selanjutnya akan dilakukan *Social Network Analysis* pada topik tentang taksi konvensional BlueBird Taxi dan taksi berbasis *online* GoCar dan GrabCar. Berikut ini adalah hasil observasi SNA pada *keyword* tentang taksi konvensional dan taksi berbasis *online* dalam media sosial Twitter. Sebelum dilakukan analisis lebih lanjut, maka terlebih dahulu dilakukan perbandingan jaringan yang ditunjukkan pada Tabel 26.

**Tabel 26** Data Perbandingan Jaringan

|  |  |  |
| --- | --- | --- |
| *Degree Range* | *Nodes* | *Edges* |
| 0 | 18201 | 7167 |
| 1 | 7525 | 7167 |
| 2 | 2186 | 3077 |
| 3 | 487 | 779 |

Dari Tabel 26 didapatkan jumlah *node* dan *edge* dari jaringan tentang taksi konvensional dan taksi berbasis *online*. Dapat diketahui bahwa jika tanpa menggunakan *filter degree range*, jumlah *nodes* dan *edges* yang didapatkan adalah sebanyak 18201 *nodes* dan 7167 *edges*. Sedangkan, jika menggunakan *filter degree range* dengan minimal adanya interaksi pada *edge* adalah satu, jumlah *nodes* yang didapatkan menjadi 7525 dan jumlah *edges* tetap. Hal tersebut karena jika menggunakan *filter degree range* minimal satu, maka *nodes* yang tidak memiliki interaksi *edges* minimal satu akan dihilangkan, sehingga dapat mengurangi *noise* pada *graph*. Berikut adalah hasil visualisasi *network* yang terbentuk untuk jaringan tentang taksi konvensional dan taksi berbasis *online* dengan menggunakan *filter degree range* satu yang ditunjukkan pada Gambar 13.

Gambar 13 merupakan visualisasi *network* untuk topik tentang taksi konvensional dan taksi berbasis *online* dengan menggunakan tipe jaringan *Frunchterman Reingold* dan *ForceAtlas 2*. Tipe tersebut merupakan tipe yang umum digunakan dalam pembuatan *graph* untuk SNA. Dari Gambar 13 didapatkan nilai *density* sebesar 0.000. Nilai *density* sebesar 0.000 menunjukkan bahwa *network* dari topik tentang taksi konvensional dan taksi berbasis *online* memiliki kepadatan yang kecil atau renggang. *Density* yang kecil atau renggang ini terjadi karena rendahnya interaksi antar *account*, baik berupa *mention*, *quote retweet*, atau *reply* yang dilakukan antar *node* dalam jaringan. Berikut adalah hasil analisis *centrality* dari Gambar 13 yang ditunjukkan pada Tabel 27.



**Gambar 13** Visualisasi *Network*

Pada hasil analisis *degree centrality* pada Tabel 27 didapatkan bahwa *account* @Bluebirgroup memiliki nilai *degree centrality* tertinggi yaitu sebanyak 1652 (pada Gambar 13 ditunjukkan dengan angka 1), yang artinya *account* @Bluebirdgroup memiliki hubungan atau interaksi yang berupa *mention*, *quote retweet*, atau *reply* antar *node* lain sebanyak 1652 kali. Kemudian yang kedua adalah *account* @gojekindonesia dengan nilai *degree centrality* sebesar 798 (pada Gambar 13 ditunjukkan dengan angka 2), dan yang ketiga adalah *account* @GrabID dengan nilai *degree centrality* sebesar 492 (pada Gambar 13 ditunjukkan dengan angka 3). Dari analisis *degree centrality* dapat disimpulkan bahwa *official account* penyedia jasa transportasi umum BlueBird Taxi merupakan *official account* yang paling banyak memiliki interaksi dibandingkan dengan *official account* dari taksi berbasis *online* GoCar dan GrabCar. Hal tersebut dikarenakan adanya pembatasan topik pada taksi berbasis *online* yang diambil yaitu hanya tentang pelayanan taksi, sedangkan untuk jenis pelayanan yang lain tidak diikutkan.

**Tabel 27** Analisis *Centrality*

|  |  |  |
| --- | --- | --- |
| *Centrality* | *Username* | *Score* |
| *Degree Centrality* | @Bluebirdgroup | 1652 |
| @gojekindonesia | 798 |
| @GrabID | 492 |
| *Closeness Centrality* | @wintermeshach | 0.8 |
| @nengbiker | 0.8 |
| @bestyoland | 0.8 |
| *Betweenness Centrality* | @Bluebirdgroup | 1023661.605813 |
| @gojekindonesia | 208158.941392 |
| @GrabID | 173641.886128 |
| *Eigenvector Centrality* | @Bluebirdgroup | 1.0 |
| @gojekindonesia | 0.355657 |
| @GrabID | 0.201661 |

Selanjutnya dari nilai *closeness centrality* didapatkan tiga *account* dengan nilai tertinggi yaitu @wintermeshach, @nengbiker, dan @bestyoland. Nilai *closeness centrality* yang didapatkan dari ketiga *account* tersebut sama, yaitu sebesar 0.8. Sehingga *node-node* dengan nilai *closeness centrality* sebesar 0.8 lebih cepat dan lebih mudah dalam berkomunikasi dengan *node* lain ketika melakukan *mention*, *quote retweet*, atau *reply* tanpa melalui banyak perantara. Sedangkan, *node-node* dengan nilai *closeness centrality* dibawah 0.8 perlu melalui banyak perantara untuk melakukan *mention*, *quote retweet*, dan *reply*.

Analisis *betweenness centrality* digunakan untuk mengetahui posisi *node* dalam *network*. Pada Tabel 27 didapatkan nilai *betweenness centrality* tertinggi pada *account* @Bluebirdgroup. Artinya, *account* @Bluebirdgroup merupakan *account* penghubung atau jembatan dari seluruh aliran informasi dalam percakapan mengenai taksi konvensional maupun taksi berbasis *online*. Hal ini mengakibatkan banyak *node* lain yang bergantung pada *tweet* yang di-*posting* oleh @Bluebirdgroup, karena *node* tersebut adalah sebuah jembatan bagi *node* lain sebagai sumber informasi terkait percakapan mengenai taksi konvensional dan taksi berbasis *online*. Selanjutnya, dari analisis *eigenvector centrality*, nilai paling besar dimiliki oleh *account* @Bluebirdgroup sebesar 1. Hal tersebut menunjukkan bahwa *account* @Bluebirdgroup memiliki interaksi baik berupa *mention*, *quote retweet*, dan *reply* dengan *node-node* tertentu.