

Sentimen Masyarakat Terkait Perpindahan Ibukota Via Model *Random Forest* dan *Logistic Regression*

Martaliana Putri Agustina¹, Hendry²

^{1,2}Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana; (0298) 321212
¹⁻²Jl. Dr. O. Notohamidjojo No.1 - 10, Blotongan, Kec. Sidorejo, Kota Salatiga 50715
e-mail: *¹martalianaputri.13@gmail.com

Received: 19-10-2021

Riwayat artikel:
Revised: 19-11-2021

Accepted: 23-11-2021

Abstract

This study aims to determine public opinion regarding the relocation of the capital of Indonesia. The pros and cons conveyed by the community are important because they can be constructive input for the government. The applications used to support the research are Orange and Twitter. The data obtained must go through several processes such as preprocess text, sentiment analysis, and testing algorithms to ensure data accuracy before making a decision tree. This research uses Random Forest and Logistic Regression as the research models. As for the results, Random Forest obtains higher accuracy value than Logistic Regression and it is considered that time did not affect the classification. The result obtained from the decision tree is that more people choose to have a neutral opinion on this program. People prefer to entrust everything to the government, because the government must have thought about the positive or negative impacts in the long term.

Keywords: *Public Opinion, Removal of the Capital, Random Forest, Logistic Regression.*

Abstrak

Penelitian ini bertujuan untuk mengetahui pendapat masyarakat mengenai pemindahan ibukota. Pro kontra yang disampaikan oleh masyarakat penting karena dapat menjadi masukan yang membangun bagi pemerintah. Aplikasi yang digunakan untuk mendukung proses penelitian tersebut adalah Orange dan Twitter. Data yang didapat harus melewati beberapa proses seperti *preprocess text*, *sentiment analysis*, dan pengujian algoritma untuk memastikan akurasi data sebelum pembuatan pohon keputusan. Penelitian ini menggunakan *Random Forest* dan *Logistic Regression* sebagai model penelitian. Hasil yang didapat adalah *Random Forest* memiliki nilai akurasi lebih tinggi dibanding *Logistic Regression* dan dianggap waktu tidak mempengaruhi klasifikasi. Hasil pohon keputusan lebih banyak masyarakat yang memilih berpendapat netral terhadap program ini. Masyarakat lebih memilih mempercayakan semua kepada pemerintah, karena pemerintah pasti sudah memikirkan dampak positif atau negatifnya dalam jangka panjang.

Kata kunci: *Pendapat masyarakat, Perpindahan Ibukota, Random Forest, Logistic Regression.*

Pendahuluan

Program pemindahan ibukota Indonesia sebenarnya sudah direncanakan beberapa tahun yang lalu, namun mulai dapat direalisasikan di tahun 2019. Media berita yang mengunggah tentang wacana pemindahan ibukota, selalu muncul dan tenggelam karena tidak ada putusan yang pasti mengenai program tersebut, serta perencanaan yang tidak matang. Pemerintah mengatakan bahwa tujuan dari pemindahan ibukota sendiri yaitu pemerataan pembangunan ke wilayah Indonesia lainnya, mengurangi beban ibukota di Jakarta, mengubah *mindset* masyarakat yang awalnya Jawa Sentris menjadi Indonesia Sentris, serta masih ada tujuan-tujuan lainnya. Hasil dari pencarian data di aplikasi Twitter, program pemindahan ibukota ini menimbulkan banyak pro kontra bagi penduduk Indonesia. Beberapa masyarakat yang mengatakan jika anggaran yang digunakan untuk pembangunan ibukota baru dapat direalisasikan untuk hal yang lebih penting seperti pandemi *Covid-19* yang sedang melanda Indonesia. Oleh sebab itu, penelitian ini dibuat dengan tujuan untuk mengetahui pendapat masyarakat mengenai salah satu program yang sedang berjalan ini. Pendapat masyarakat dapat menjadi masukan yang membangun untuk pemerintah, serta pemerintah juga dapat memberi penjelasan terhadap masyarakat mengenai hal-hal positif program tersebut.

Untuk mendukung proses penelitian ini, diperlukan aplikasi *Orange* dan Twitter. *Orange* adalah *software open-source* yang digunakan untuk proses *data mining* melalui konsep visual *programming*. API Twitter diinputkan pada *tools* yang sudah membentuk alur *data mining*, setelah diinputkan kemudian menentukan kata kunci pencarian, maka akan menghasilkan data yaitu *tweet* masyarakat dengan berbagai pendapat mengenai program yang sedang berjalan ini yaitu pemindahan ibukota [1].

Setiap proses *Data Training* yang dilakukan akan menghasilkan jumlah *tweet* yang berbeda karena memang bersifat *random*. Proses *Data Training* dilakukan sebanyak 13 kali untuk mendapat hasil *tweet* terbanyak, setelah proses tersebut selesai, *tweet-tweet* tersebut dilakukan *Sentiment Analysis* untuk mendeteksi pro, kontra, serta netral. Pohon keputusan ini kemudian dibuat untuk menyimpulkan hasil penelitian tersebut.

Pada penelitian sebelumnya mendapatkan referensi yang sebagian besar menggunakan metode *Naïve Bayes*, dengan hasil nilai akurasi yang bermacam-macam namun relatif baik/tinggi jumlah persentasenya. Program pemindahan ibukota ini cukup membuat rakyat Indonesia terkejut. Banyak penulis lain yang penasaran dengan tanggapan-tanggapan masyarakat mengenai program tersebut. Rata-rata penulis lain juga mengelompokkan pendapat masyarakat menjadi kelas positif, negatif, dan netral.

Kajian Pustaka

Penelitian ini dilakukan berdasarkan penelitian terdahulu yang memiliki relevansi pada pembahasannya. Dalam jurnal penelitian berjudul “Analisa Sentimen Persepsi Masyarakat Terhadap Pemindahan Ibukota Baru di Kalimantan Timur Pada Media Sosial Twitter” membahas tentang persepsi masyarakat Indonesia di Twitter mengenai pemindahan ibukota. Data yang didapat disimpan dalam *database* MySQL kemudian dilakukan proses *Text Processing*. Penelitian ini menggunakan metode *Naïve Bayes Classifier* dan proses klasifikasi dibagi menjadi sentimen positif dan sentimen negatif yang menggunakan 200 data *tweet* yaitu terdiri dari 41 *data testing* dan 159 *data training* kemudian menghasilkan akurasi sebesar 78% [2].

Dalam jurnal penelitian berjudul “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)” membahas tentang topik perbincangan di aplikasi Twitter mengenai perpindahan ibukota Indonesia. Terdapat kontroversi dari pihak pro dan kontra yang memiliki sudut pandang berbeda. *Sentiment Analysis* adalah proses mengolah data secara otomatis untuk mendapatkan informasi sentimen pada sebuah opini. Metode yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM) untuk diterapkan pada topik pemindahan ibukota bertujuan untuk mengklasifikasi data sentimen. Teknik yang digunakan untuk klasifikasi adalah mengklasifikasi menjadi dua kelas yaitu positif dan negatif [3].

Dalam jurnal penelitian berjudul “Analisis Sentimen Terhadap Pemindahan Ibu Kota Pada Media Sosial Twitter Menggunakan Algoritma *Naive Bayes* Berbasis Particle Swarm Optimization” membahas tentang analisis sentimen untuk mengidentifikasi opini masyarakat yang terdapat pada aplikasi Twitter. Algoritma *Naive Bayes* adalah metode pengklasifikasian sederhana untuk teks, karena dianggap kurang optimal sehingga perlu menggunakan fitur tambahan untuk meningkatkan akurasi. Tujuan dari penelitian tersebut adalah untuk mengoptimalkan algoritma *Naive Bayes* dengan fitur *Particle Swarm Optimization*. Hasil akurasi dari menggunakan fitur tersebut adalah 91.50% dan jika hanya menggunakan *Naive Bayes* tunggal hanya mendapat hasil akurasi 78,88% [4].

Dalam jurnal penelitian berjudul “Klasifikasi Argument Pada Teks dengan Menggunakan Metode *Multinomial Logistic Regression* Terhadap Kasus Pemindahan Ibu Kota Indonesia di Twitter” membahas tentang keterbatasan pengguna dalam menyaring informasi sehingga mempersulit masyarakat atau pemerintah dalam menganalisis banyak informasi yang terdapat pada *tweet*. Oleh karena itu, dibuatlah sebuah sistem yang dapat melakukan klasifikasi terhadap data yang berupa *tweet* secara otomatis ke dalam tiga kelas, yaitu non argumen, *unknown*, dan argumen. Metode yang digunakan dalam penelitian tersebut adalah *Multinomial Logistic Regression* (MLR) yang merupakan metode generalisasi dari

Logistic Regression. Sebelum dilakukan proses klasifikasi, data/*tweet* harus dilakukan *preprocessing* terlebih dahulu agar bersih dari segala *noise* yang ada [5].

Dalam jurnal penelitian berjudul “Analisis Sentimen: Pemindahan Ibu Kota Indonesia pada Twitter” membahas tentang aktivitas-aktivitas bersosial media yang dilakukan masyarakat. Topik yang sedang banyak diperbincangkan di aplikasi Twitter adalah perpindahan ibukota negara Indonesia. Banyaknya opini masyarakat menimbulkan permasalahan seperti perbedaan pendapat. Hal tersebut dapat dimanfaatkan untuk dilakukan *Sentiment Analysis* yaitu sentimen negatif, positif, dan netral. Langkah-langkah yang perlu dilakukan adalah pengumpulan data kemudian diolah dengan data mining. Pada penelitian ini menggunakan metode *Naïve Bayes* yang menggunakan tiga macam dengan tujuan untuk mendapat nilai akurasi terbaik. Penelitian ini menghasilkan akurasi terbaik yang didapatkan oleh *Multinomial Naïve Bayes* [6].

Dalam jurnal penelitian berjudul “Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan Algoritma Term-Based Random Sampling dan Metode Klasifikasi Naïve Bayes” membahas tentang analisis sentimen terhadap kebijakan pemindahan ibukota dengan menggunakan aplikasi Twitter sebagai alat untuk mencari data tersebut. Proses pengembangan sistem menggunakan perhitungan *Raw Term Frequency*, *Preprocessing* dan *Data Scraping*. Pada *Preprocessing* terdapat proses *filtering* yang menggunakan algoritma *Term-Based Random Sampling* untuk *stoplist*. Proses pengujian menggunakan dua metode yaitu *Confusion Matrix* berjenis *multi class* dan pengujian parameter [7].

Dalam jurnal penelitian berjudul “Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial” membahas tentang penambahan teks yang melibatkan pendapat, komentar dan tanggapan. Kelas sentimen pada penelitian ini menggunakan *scoring*, metode yang digunakan adalah *Naïve Bayes* dengan model *Bernoulli* dan *Multinomial*. *Bernoulli* menghasilkan tingkat *Recall* 93,45% sedangkan *Multinomial* menghasilkan tingkat *Recall* 90,19%, yang berarti *Bernoulli* maupun *Multinomial* memiliki hasil yang baik [8].

Dalam jurnal penelitian berjudul “Perbandingan Metode Regresi Logistik dan *Random Forest* untuk Klasifikasi Data Imbalanced” yang mengangkat studi kasus Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017. Hasil dari studi kasus ini adalah hasil terbaik berada pada model *Logistic Regression* yaitu dengan nilai akurasi 78,13% sedangkan *Random Forest* hanya 63,91%, hal itu membuktikan bahwa waktu berpengaruh pada penelitian ini, karena perekonomian masyarakat dapat berubah-ubah setiap harinya [9].

Tabel 1 merupakan rangkuman kajian penelitian terdahulu.

Tabel 1 Penelitian Terdahulu

No.	Tahun	Penulis	Metode
1.	2020	1. Icha Adellia Safra 2. Eri Zuliarso	<i>Naïve Bayes Classifier</i>
2.	2021	1. Primandani Arsi 2. Retno Waluyo	<i>Support Vector Machine (SVM)</i>
3.	2020	1. Ahmad Taufiq	<i>Naïve Bayes</i>
4.	2020	1. Mochammad Naufal Rizaldi 2. Adiwijaya 3. Said Al Faraby	<i>Multinomial Logistic Regression (MLR)</i>
5.	2018	1. Erica Mas'udah 2. Eka Dyar W., S.Kom., M.Kom. 3. Amalia Anjani, S.Kom., M.Kom.	<i>Naïve Bayes</i>
6.	2019	1. Akhmad Sa'rony 2. Putra Pandu Adikara 3. Randy Cahya Wihandika	<i>Naïve Bayes</i>
7.	2020	1. Nabila Surya Wardani 2. Alan Prahutam 3. Puspita Kartikasari	<i>Naïve Bayes Classifier</i>
8.	2019	1. Taly Purwa	<i>Decision Tree</i>

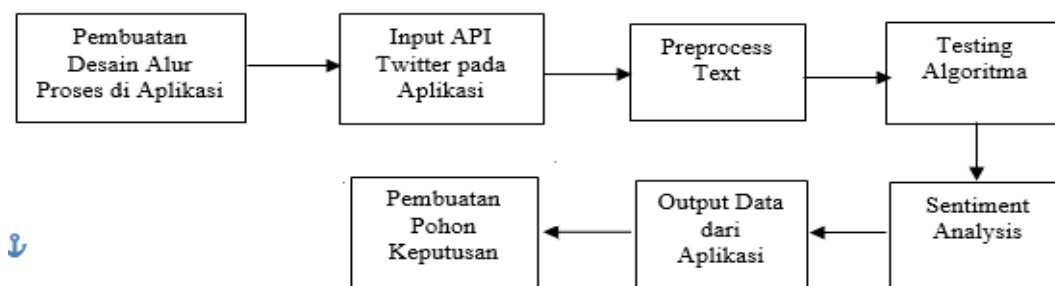
Metode yang digunakan pada penelitian ini yaitu *Decision Tree* atau yang biasa disebut pohon keputusan, metode berupa diagram alir yang menyerupai pohon berfungsi membantu dalam mempertimbangkan suatu pilihan. Tujuan dari penelitian ini yaitu untuk menentukan sentimen negatif, sentimen positif dan netral. Model yang digunakan pada penelitian ini adalah *Random Forest* yang berarti kombinasi dari masing-masing *tree*, kemudian dikombinasikan dalam satu model, dimana waktu tidak akan mempengaruhi hasil perhitungan yang berarti kapanpun penelitian ini dilakukan maka hasilnya akan sama. *Logistic Regression* adalah teknik regresi yang berfungsi untuk memisahkan dataset menjadi dua kelompok, namun penelitian ini menggunakan *Multinomial Logistic Regression* yang menjadikan dataset dalam tiga kelompok yaitu negatif, positif, dan netral.

Pada riset sebelumnya, *Logistic Regression* menghasilkan akurasi lebih tinggi daripada *Random Forest*, karena penelitian menyangkut perekonomian yang bersifat dapat berubah-ubah hasilnya atau dapat disebut waktu sangat berpengaruh dalam penelitian. Berbeda dengan penelitian ini dimana hasil yang didapat adalah hasil *Random Forest* lebih tinggi atau dapat disebut waktu tidak mempengaruhi hasil.

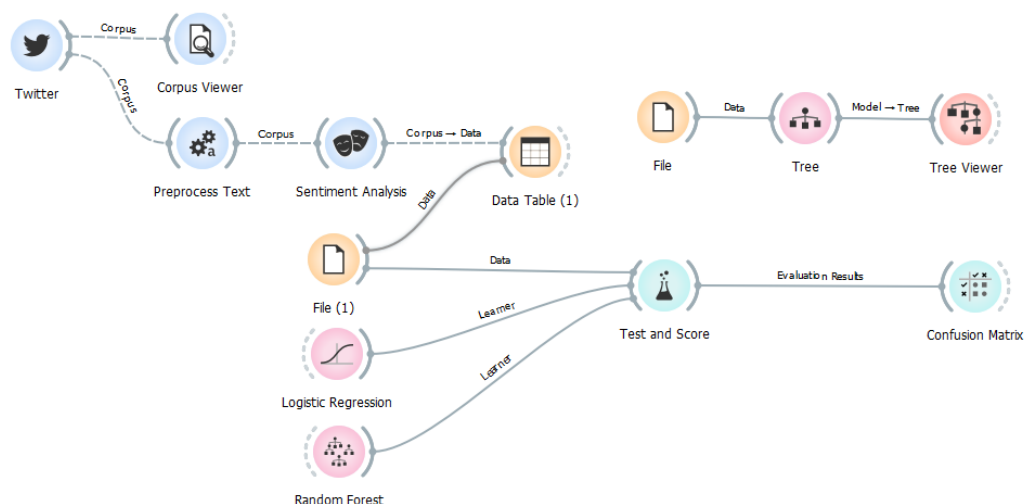
Metode Penelitian

Kerangka kerja pada Gambar 1 merupakan langkah-langkah yang akan dilakukan selama proses penelitian ini berjalan dalam penyelesaian masalah. Proses

yang dilewati harus berhasil dilakukan, karena setiap proses terhubung untuk melakukan proses selanjutnya. Adapun detail kerangka kerja pada penelitian ini ada tahapan demi tahapannya dapat dilihat seperti pada Gambar 1 [10][11].



Gambar 1. Kerangka Kerja



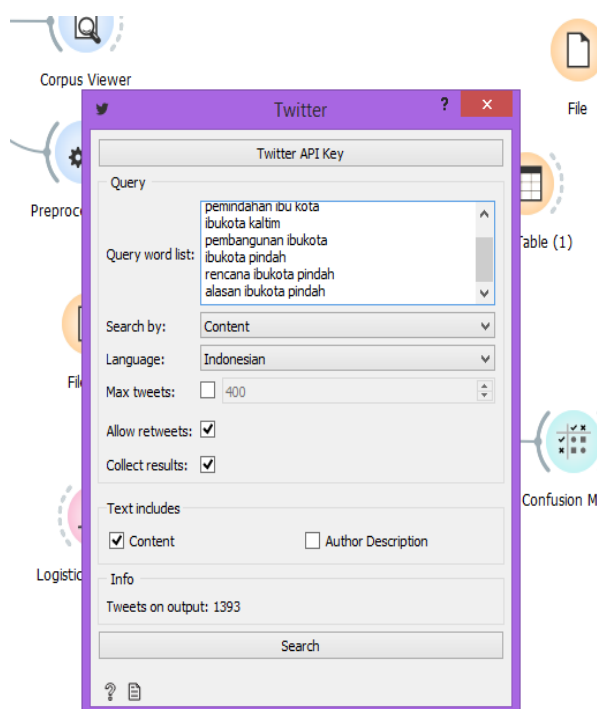
Gambar 2. Desain Alur Proses pada Aplikasi *Orange* [12]

Penelitian ini mendapat referensi dalam pembuatan desain alur proses tersebut. Data diproses dalam beberapa tahapan hingga pada proses terakhir yaitu pembuatan pohon keputusan yang dibuat dalam satu halaman yang sama. Beberapa tahapan atau proses dapat dilihat pada Gambar 2, dimana diperlukan banyak proses untuk mendapatkan data yang akurat. Proses ini akan menghasilkan banyak hal seperti topik opini politik jaringan, tren *hashtag*, *brand sentiment*, dan sebagainya. Hasil pencarian yang diperoleh ini akan diproses terlebih dahulu dan akan menghasilkan *corpus* atau data yang menarik untuk diulas atau layak digunakan sebagai bahan penelitian [12].

Pada *tool File*, membutuhkan *Excel* sebagai perantara antara hasil data yang didapat dengan proses pengujian algoritma. Data yang didapat dipindahkan ke *excel*

kemudian diinputkan pada *tool file* agar dapat terdeteksi, setelah data terdeteksi pada *tool file* kemudian proses selanjutnya adalah pengujian algoritma yang akan mengoutputkan dua hasil dari *Random Forest* dan *Logistic Regression* untuk selanjutnya dibandingkan hasil yang terbaik dari kedua model. Proses terakhir yaitu pembuatan pohon keputusan, karena data pada awal sangat kompleks jadi diperlukan penghapusan beberapa kolom pada data, hanya tersisa kolom nama responden, isi *tweet* responden, dan sentimennya untuk pembuatan pohon keputusan tersebut.

a. *Input API Twitter pada Aplikasi*



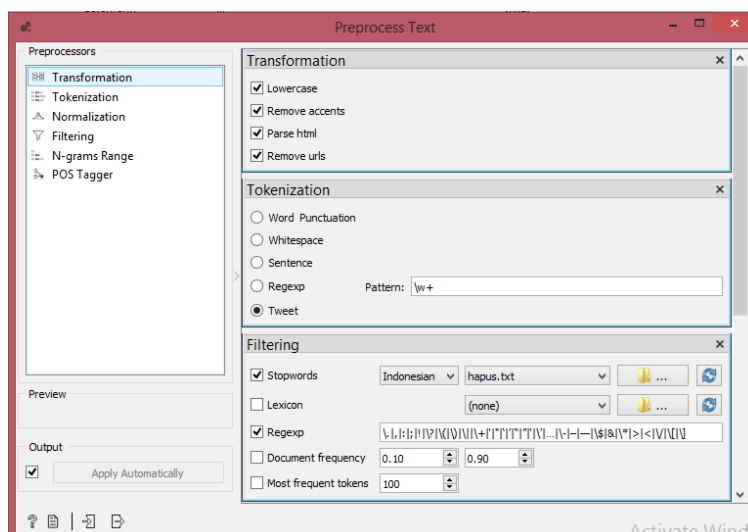
Gambar 3. Hasil Pencarian *Tweet*

Gambar 3 adalah proses pencarian data berupa *tweet*, dimana pada gambar tersebut terlihat mendapatkan 1393 data yang akan digunakan. Sebelum tahap pencarian data harus menghubungkan aplikasi *Orange* yang digunakan dengan aplikasi Twitter menggunakan API. Cara mendapatkan API Twitter adalah dengan masuk pada alamat *developer.Twitter.com* dengan mengisi beberapa pertanyaan terkait kegunaan API tersebut, jika disetujui maka akan mendapat *keys* dan *tokens*. *Keys* tersebut dapat digunakan untuk penelitian ini.

Tahap selanjutnya adalah API yang sudah didapat diinputkan pada *tools* Twitter, kemudian penentuan kata kunci pencarian, semakin banyak kata kunci maka akan semakin banyak data yang didapat karena pencarian data semakin luas. Klik *tools* pada *search* dan proses pencarian data berjalan, jika proses *search* selesai, maka data akan masuk pada *tools Corpus Viewer*. Diperlukan

proses *search* atau *Data Training* sebanyak 13 kali untuk mendapat hasil data atau *corpus* yang maksimal jumlahnya.

b. *Preprocess Text*

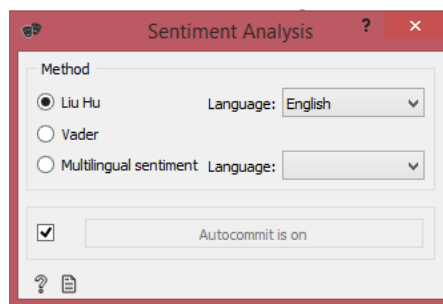


Gambar 4. *Preprocess Text*

Tahap *Preprocess Text* adalah tahap normalisasi data dan menerangkan berbagai tipe proses dari data mentah untuk mempersiapkan ke proses selanjutnya, dimana akan diinputkan file *.txt* yang berisi kata-kata untuk *memfilter* seperti yang digambar bertuliskan *hapus.txt*. Kata-kata yang tidak penting seperti hanya sekedar kata imbuhan, simbol-simbol akan dihapus dan sumber data yang diinginkan hanya data yang menggunakan Bahasa Indonesia. Tahap ini harus dilakukan, karena apabila tidak dilakukan maka akan muncul data dengan berbagai bahasa serta membuat nilai data yang sebelumnya besar. Hasil *training* data pada aplikasi Twitter menjadi lebih kecil karena adanya proses ini tanpa merubah informasi yang ada, untuk hal tersebut dapat dilihat pada Gambar 4.

Tahap pada *Preprocess Text* ini menjadi salah satu tahap yang penting dan memerlukan ketelitian yang baik, sebab tahap ini akan menentukan kualitas data yang dihasilkan, apabila tahap ini terjadi kesalahan maka angka akurasi data yang dihasilkan pada tahap selanjutnya akan menurun dan akan dianggap kualitas data yang digunakan rendah. Penelitian ini dianggap baik atau berhasil jika keakuratan suatu data memiliki nilai yang tinggi, pada kolom *Filtering* pertama diubah “Indonesia” dari yang semula “none” yang artinya data yang didapat akan berbagai bahasa.

c. *Sentiment Analysis*



Gambar 5. *Sentiment Analysis*

Pada proses ini menggunakan *Liu Hu* yaitu menghitung skor normal tunggal dari sentimen dalam teks (skor yang memiliki nilai negatif untuk sentimen negatif, skor yang memiliki nilai positif untuk sentimen positif, sedangkan 0 adalah netral) Pada tahap *Sentiment Analysis*, akan dilakukan proses penentuan sentimen negatif, sentimen positif dan netral. Data yang akan dihasilkan nanti akan memunculkan keterangan bahwa *tweet* tersebut adalah *tweet* negatif atau positif dan dapat langsung dikelompokkan sesuai keterangan. *Sentiment Analysis* terdiri dari dua kategori yaitu *Coarse – grained sentiment analysis* dan *Fined – grained sentiment analysis*. *Coarse – grained sentiment analysis* dimana proses analisa serta klasifikasi data secara keseluruhan dan membaginya menjadi tiga yaitu positif, negatif dan netral, namun ada beberapa yang menjadikan nilai tersebut bersifat kontinu atau tidak diskrit. *Fined – grained sentiment analysis* yaitu objek yang dikelompokkan tidak pada level data namun terletak pada level kalimat di setiap data [13].

Sentiment Analysis terdiri dari tiga sub proses yaitu *subjectivity Classification*, *Orientation Detection*, dan *Opinion Holder and Target Detection*. *Subjectivity Classification* adalah proses penentuan kalimat opini pada data. *Orientation Detection* adalah pengelompokkan opini ke dalam *Coarse – grained sentiment analysis* dan *Opinion Holder and Target Detection* adalah penentuan bagian yang merupakan pemberi opini [13]. Proses ini dapat dilihat pada Gambar 5, dimana *Liu Hu* dipilih untuk penelitian kali ini.

d. *Output Data dari Aplikasi*

Setelah seluruh proses pencarian dan normalisasi data dilakukan, data yang dihasilkan akan sesuai dengan yang diinginkan, termasuk akan muncul keterangan *Sentiment Analysis* dan data akan langsung mengelompokkan sesuai sentimennya. Hal yang dilakukan setelah data didapat adalah memindahkan data pada *excel* kemudian menginputkan data tersebut pada *tool* “File” yang tersambung pada *tool* “*Test and Score*” untuk dilakukan proses selanjutnya yaitu *Testing Algoritma*. Pemindahan data ke dalam *excel* dan kemudian diinputkan kembali pada *tool* “File” dikarenakan data tidak dapat terbaca untuk *Testing*

Algoritma jika dari tool “*Data Table*” langsung disambungkan pada tool “*Test and Score*”.

Data yang didapat sebanyak 1393 *tweet* sesuai dengan yang ada pada tahap pencarian data. Data tersebut hanya dapat digunakan untuk tahap *Testing Algoritma*, jika data tersebut akan digunakan untuk pembuatan pohon keputusan, data perlu diubah lebih minimalis yaitu menyisakan kolom nama responden, isi *tweet* responden dan hasil sentimen.

e. Pengujian Algoritma

Proses selanjutnya apabila data sudah didapat adalah tahap pengujian algoritma, dimana untuk mengetahui apakah algoritma yang digunakan baik atau tidak. Hasil dari pengujian algoritma ini adalah *Random Forest* yaitu algoritma untuk klasifikasi dalam jumlah besar serta waktu tidak akan mempengaruhi hasil dan *Logistic Regression* yaitu statistika yang digunakan untuk memprediksi probabilitas kejadian suatu peristiwa, waktu juga sangat berpengaruh model ini atau biasa disebut *Time Series* dimana suatu hari hasil dapat berubah [10]. Hasil yang menjadi acuan penelitian yang menggunakan model *Random Forest* dan *Logistic Regression* ini adalah *Recall* dan *Precision*, jika hasil *Recall* dan *Precision* ini semakin tinggi maka data dianggap akurat dan layak digunakan sebagai penelitian [11]. Proses *Data Training* diperlukan beberapa kali untuk menghasilkan data yang maksimal dan hal tersebut dapat dilihat dari perolehan *Recall* dan *Precision* pada kedua model yang digunakan.

Tujuan dari digunakannya dua model pada penelitian ini adalah untuk membandingkan hasilnya. Hasil yang didapat pada model *Random Forest* dan *Logistic Regression* berupa nominal angka yang terlihat jelas hasilnya. Hasil yang kemudian dibandingkan untuk mendapatkan hasil yang paling tinggi/baik, model yang memiliki nilai tertinggi dianggap tepat untuk penelitian yang sedang dilakukan.

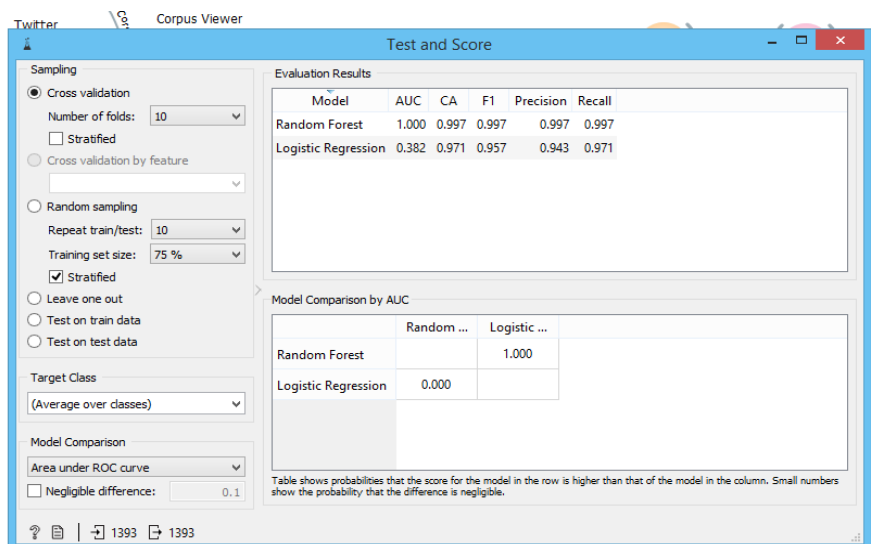
f. Pembuatan Pohon Keputusan

Tahap akhir dari penelitian ini yaitu pembuatan pohon keputusan. Data yang telah melalui beberapa proses seperti *Sentiment Analysis* dan *Testing Algoritma* kemudian diubah bentuk menjadi pohon keputusan agar mempermudah pembaca dalam memahami hasil penelitian. Pohon keputusan adalah sebuah struktur yang membagi kumpulan data besar menjadi himpunan yang lebih kecil dengan menerapkan beberapa aturan keputusan. Aturan keputusan tersebut biasanya disebut dengan atribut, atribut yang dinyatakan sebagai parameter dibuat untuk memisahkan hasil berdasarkan kriteria yang ditentukan. Atribut yang digunakan untuk penelitian pemindahan ibukota ini adalah pro, kontra, dan netral. Setiap atribut memiliki target atribut masing-masing. Contoh target *attribute Pro* adalah adanya *hashtag* dukung pemindahan

ibukota [14]. Pada tahap output data dari aplikasi dijelaskan bahwa data yang digunakan untuk membuat pohon keputusan adalah kolom nama responden, isi *tweet* responden dan hasil *sentiment*, apabila seluruh kolom digunakan maka terjadi *error* sehingga tidak terbaca pada *tool Tree* yang terdapat di aplikasi. Hasil yang didapat, hanya kolom-kolom tersebut yang tercantum dan membentuk pohon keputusan.

Hasil dan Pembahasan

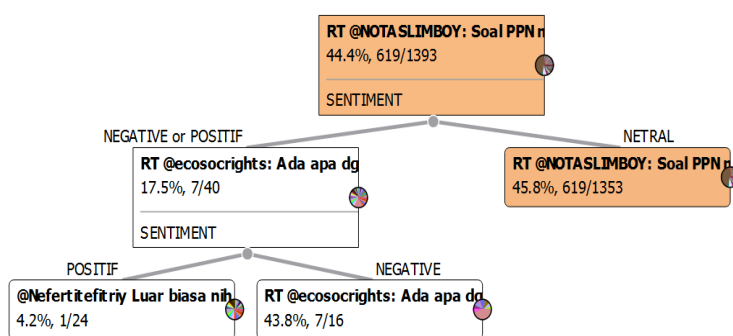
Pada bab ini terdapat dua hasil yang didapat yaitu pengujian algoritma dan pohon keputusan. Hasil pengujian algoritma dapat menjelaskan perbandingan antara dua model yang digunakan untuk mendapatkan model yang terbaik untuk penelitian ini. Model yang digunakan yaitu *Random Forest* (tidak berbasis waktu) dan *Logistic Regression* (berbasis waktu). Alasan menggunakan kedua model tersebut adalah untuk mengetahui apakah waktu berpengaruh terhadap klasifikasi tersebut, sedangkan pohon keputusan untuk mempermudah pembaca dalam melihat hasil dari penelitian.



Gambar 6. Hasil Pengujian Algoritma

Gambar 6 menunjukkan hasil pengujian algoritma menggunakan *Random Forest* dan *Logistic Regression*. Dari 1393 data didapat hasil *Random Forest* dan *Logistic Regression* adalah AUC yaitu *Area Under the Curve* yang dapat mempermudah dalam membandingkan model satu dengan yang lainnya. CA berfungsi untuk mengetahui tingkat akurasi suatu dataset, sedangkan F1 sebagai acuan perbandingan antara *Precision* dan *Recall*. *Precision* digunakan untuk mengetahui tingkat ketepatan suatu data, sedangkan *Recall* berfungsi untuk mengetahui tingkat keberhasilan sistem dalam menemukan kembali data tersebut. [11]

Dari hasil yang didapat, perbandingan antara hasil dari *Random Forest* dan *Logistic Regression* tidak terlalu jauh. *Random Forest* yang memiliki nilai tinggi pada setiap *Evaluation Result* tetap dianggap lebih baik dari model *Logistic Regression*. Data yang digunakan untuk pengujian algoritma juga memiliki akurasi yang tinggi sehingga dikatakan layak untuk melakukan penelitian ini. Data yang digunakan dapat dipastikan untuk akurasinya. *AUC*, *CA*, *F1*, *Precision* dan *Recall* berperan sebagai *baseline* pada penelitian ini, sehingga menjelaskan bahwa model tersebut menjadi acuan dalam mengukur keefektifan suatu data, sehingga dapat diputuskan jika waktu tidak akan mempengaruhi klasifikasi yang ada.



Gambar 7. Pohon Keputusan

Gambar 7 menjelaskan bahwa pohon keputusan yang dibuat menghasilkan tiga jawaban yaitu kontra (negatif), pro (positif), dan netral. Hasil kontra (negatif) sejumlah 16 responden yang berarti bahwa sesungguhnya banyak masyarakat Indonesia yang kurang setuju atau kontra terhadap pemindahan ibukota. Rata-rata masyarakat berpendapat jika ibukota pindah maka akan berpengaruh terhadap ekonomi negara. Anggaran yang digunakan untuk pemindahan ibukota ini tidak sedikit dan masyarakat beranggapan bahwa anggaran tersebut dapat digunakan untuk hal yang lebih penting saat ini yaitu *Covid-19*, dan menganggap pemindahan ibukota ini bukan menjadi hal yang prioritas. Hasil pro atau yang setuju dengan pemindahan ibukota sejumlah 24 responden, dimana responden berpendapat bahwa memang sudah saatnya ibukota pindah. Sebenarnya banyak hal positif yang didapat, contohnya pemerataan ekonomi, agar pulau-pulau Indonesia selain pulau Jawa dapat maju dari segala bidang dan mengurangi kepadatan penduduk di Pulau Jawa. Pada hasil yang netral yang didapat dengan jumlah 1353 responden, rata-rata responden hanya sekedar berpendapat atau memberi saran untuk pemerintah, agar dapat berlaku adil untuk semua program yang sedang berjalan dan tidak lupa terhadap *pandemic* yang sedang terjadi di Indonesia. Para responden lebih percaya bahwa pemerintah sudah memikirkan pemindahan ibukota tersebut secara baik, dan dapat mengatasi resiko apapun yang terjadi nantinya. [15]

Pada penelitian ini memang ditemukan banyak kelas netral dibanding dengan kelas positif atau negatif. Masyarakat yang tidak menyatakan setuju namun juga tidak menentang adanya program tersebut dan lebih menyerahkan seluruhnya

kepada pemerintah. Kelas netral lebih banyak masyarakat yang hanya memberi saran dan hanya *reply tweet* milik responden lainnya sehingga *tweet* tersebut dianggap netral. Perbandingan antara kelas positif, negatif, dengan netral terpaut cukup jauh 40 untuk kelas positif dan negative, kemudian 1953 untuk kelas netral. Kelas positif negatif ditentukan dari *hashtag* #DukungPemindahanIbukota dan #BatalkanPemindahanIbukota.

Simpulan

Dari penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa *Testing Algoritma* yang menggunakan 1393 data memiliki tingkat akurasi yang tinggi dengan menggunakan model *Random Forest*. Hal tersebut dapat dilihat dari hasil *Testing Algoritma* pada bagian AUC yaitu 1.000 yang hasilnya lebih tinggi daripada *Logistic Regression* yang hanya 0.382, AUC sudah dapat menjadi acuan untuk *CA*, *F1*, *Precision*, dan *Recall* bahwa hasilnya akan lebih tinggi *Random Forest* daripada *Logistic Regression*.

Sesuai dengan fungsi pohon keputusan yaitu mengubah fakta besar menjadi sebuah pohon keputusan yang merepresentasikan *system* sehingga mudah dipahami. 1393 data diubah menjadi bentuk pohon keputusan dan dikelompokkan sesuai *Sentiment Analysis* sehingga dapat dijelaskan bahwa jumlah masyarakat pro dan masyarakat kontra tidak beda jauh yang berarti masih banyak masyarakat yang ragu akan keberhasilan program tersebut. Pemerintah sebaiknya lebih bisa menanamkan *mindset* kepada masyarakat agar percaya bahwa program ini memiliki nilai positif yang tinggi.

Untuk penelitian selanjutnya, penggunaan aplikasi, model dan metode lain diharapkan dapat dilakukan agar dapat menjadi perbandingan untuk tingkat relevansi dan keakuratan suatu penelitian.

Daftar Pustaka

- [1] I. Priyadi, J. Santony, and J. Na'am, "Data Mining Predictive Modeling for Prediction of Gold Prices Based on Dollar Exchange Rates, Bi Rates and World Crude Oil Prices," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 93, 2019, doi: 10.24014/ijaidm.v2i2.6864.
- [2] I. A. Safra *et al.*, "Analisa sentimen persepsi masyarakat terhadap pemindahan ibukota baru di kalimantan timur pada media sosial twitter 1,2," pp. 978–979, 2020.
- [3] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) SENTIMENT ANALYSIS ON THE DISCUSSION OF RELOCATING INDONESIA ' S CAPITAL CITY USING THE SUPPORT VECTOR MACHINE (SVM)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [4] P. Teknologi, J. Ilmiah, A. Lingkungan, and A. Taufiq, "MEDIA SOSIAL

- TWITTER MENGGUNAKAN ALGORITMA NAIVE Text Mining,” vol. 14, no. 1, 2020.
- [5] M. N. Rizaldi and S. Al Faraby, “Klasifikasi Argument Pada Teks dengan Menggunakan Metode Multinomial Logistic Regression Terhadap Kasus Pemindahan Ibu Kota Indonesia di Twitter,” vol. 4, no. L, pp. 904–913, 2020, doi: 10.30865/mib.v4i4.2348.
- [6] E. Mas, E. D. W. S. Kom, M. Kom, A. A. S. Kom, and M. Kom, “ANALISIS SENTIMEN: PEMINDAHAN IBU KOTA INDONESIA PADA TWITTER,” vol. 1, no. 2, pp. 397–401, 2020.
- [7] S. Akhmad, P. P. Adikara, and R. C. Wihandika, “Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan Algoritme Term-Based Random Sampling dan Metode Klasifikasi Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 10086–10094, 2019.
- [8] D. Statistika and U. Diponegoro, “3 1,2,3,” vol. 9, pp. 237–246, 2020.
- [9] T. Purwa, “Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017),” *J. Mat. Stat. dan Komputasi*, vol. 16, no. 1, p. 58, 2019, doi: 10.20956/jmsk.v16i1.6494.
- [10] A. Primajaya *et al.*, “Random Forest Algorithm for Prediction of Precipitation,” vol. 1, no. 1, pp. 27–31, 2018.
- [11] D. Derisma, “Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining,” *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 84–88, 2020, doi: 10.30871/jaic.v4i1.2152.
- [12] “Getting Started with Orange 07_ Model Evaluation and Scoring.” .
- [13] V. I. Santoso, G. Virginia, Y. Lukito, U. Kristen, and D. Wacana, “PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE,” vol. 14, no. 1, pp. 72–76, 2017.
- [14] P. Bidang, K. Sains, Y. Mardi, J. Gajah, M. No, and S. Barat, “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika.”
- [15] A. R. Febie Elfaladonna, “Analisa Metode Classification-Decision Tree Dan Algoritma,” *Sci. Inf. Technol.*, vol. 2, no. 1, pp. 10–17, 2019.