

## ANALISA HOAX TERHADAP *ISSUE* PEMINDAHAN IBU KOTA NEGARA (IKN) INDONESIA MENGGUNAKAN *ALGORITMA* *SUPPORT VECTOR MACHINE (SVM)* DAN *RANDOM FOREST (RF)*

Tri yandi<sup>1</sup>, Hoiriyah<sup>2</sup>, Anwari<sup>3</sup>

<sup>1,2,3</sup>Sistem informasi, Teknik, Universitas Islam Madura

<sup>1</sup>[triyandi2604001922@gmail.com](mailto:triyandi2604001922@gmail.com), <sup>2</sup>[hoiriyah@gmail.com](mailto:hoiriyah@gmail.com) <sup>3</sup>[anwari.uim@gmail.com](mailto:anwari.uim@gmail.com)

### ABSTRAK

Seiring dengan perkembangan teknologi informasi, media berita juga turut berkembang menyajikan informasi dalam media online. Namun, sangat disayangkan pada penyebarannya masih banyak ditemukan berita hoax atau tidak benar. Berita *hoax* yang paling mudah beredar adalah *hoax* tentang isu pemindahan ibu kota negara (IKN) Indonesia. Agar terhindar dari berita *hoax* ialah dengan lebih teliti membaca judul berita pada situs yang terpercaya seperti Kompas. Karena itu penelitian ini akan mengembangkan dan menganalisis model klasifikasi berita *hoax* tentang isu pemindahan ibu kota (IKN) indonesia dengan menerapkan algoritma *Support Vector Machine (SVM)* dan *Random Forest (RF)*. Studi kasus penelitian ini dibagi dalam 2 kategori yaitu berita *hoax* yang didapat dari situs *Trunbackhoax & Hoax buster* sedangkan berita bukan *hoax* diambil dari situs berita Kompas. Hasil penelitian menyatakan bahwa hasil analisa deteksi *hoax* menggunakan algoritma *support vector machine* dan *random forest* terhadap *issue* pemindahan ibu kota negara (IKN) Indonesia, algoritma *support vector machine* memberikan hasil terbaik dengan nilai performa akurasi sebesar 77.52%. Algoritma *support vector machine* memberikan kinerja terbaik pada proses klasifikasi data.

**Kata kunci:** *Hoax, Ibu kota negara, Algoritma SVM, Algoritma Random Forest.*

### ABSTRACT

*information in online media. However, it is very popular that the spread is still a lot of hoax or untrue news. The most easily circulated hoax news is the hoax about the issue of moving the state capital (IKN) of Indonesia. In order to avoid hoax news, that is to more carefully read news titles on trusted sites such as Kompas. Therefore, this study will develop and analyze a classification model for hoax news about the issue of moving the Indonesian capital city (IKN) by applying the Support Vector Machine (SVM) and Random Forest (RF) algorithms. The case studies of this research are divided into 2 categories, namely news obtained from the Trunbackhoax & Hoax buster site, while the non-hoax news was taken from the Kompas news site. The results of the study stated that the results of the hoax detection analysis using the support vector machine and random forest algorithms on the problem of moving the national capital city (IKN) of Indonesia, the support vector machine algorithm gave the best results with an accuracy performance value of 77.52%. The support vector machine algorithm provides the best performance in the data classification process.*

**Keywords:** *Hoax, National Capital, SVM algorithm, Random Forest algorithm.*

### 1. PENDAHULUAN

Fenomena hoax belakangan ini sedang marak tersebar diberbagai media, baik di media cetak maupun media online [9]. Mirisnya, tidak adanya atensi khusus dari masyarakat akan adanya hal tersebut, sehingga mayoritas

masyarakat bisa dengan mudah mempercayai berita hoax dan dengan mudah menyebarkan kepada publik. Berita hoax adalah berita palsu yang manipulasi dari realitas sesungguhnya [1]. Banyak kasus atau peristiwa yang sebenarnya tidak terjadi namun diangkat menjadi sebuah berita dan dikemas

sebaik mungkin sehingga publik tertarik untuk membacanya[2]. Dengan banyaknya berita hoax yang beredar, tentu menimbulkan dampak negatif dikalangan masyarakat, beberapa pakar media menjelaskan beberapa dampak negatif yang dihasilkan oleh berita hoax :

1. Hoax akan menysasar emosi masyarakat.
2. Menimbulkan opini negative sehingga terjadi disintergratif bangsa.
3. Memberikan provokasi dan agitasi negatif.

Perkembangan teknologi tidak hanya berupa memberikan dampak positif saja, namun juga memberikan dampak negatif, tindak pidana penghinaan atau ujaran kebencian (*hate speech*), serta penyebaran informasi bohong atau hoax di media sosial yang ditujukan untuk menimbulkan rasa kebencian atau permusuhan antar individu dan/atau kelompok masyarakat tertentu berdasarkan Suku, Agama, Ras dan antar golongan (SARA)[3].

Pada tanggal 16 agustus 2019, presiden negara kesatuan republik indonesia menyatakan resmi terkait rencana pemindahan ibu kota negara dalam sidang bersama DPD dan DPR. Dalam pernyataan, presiden Jokowi menyampaikan bahwa ibu kota akan dipindahkan ke pulau kalimantan[4]. Selain pernyataan resmi melalui sidang bersama DPD dan DPR, presiden telah mengumumkan hal tersebut sebelumnya, yaitu pada tanggal 08 agustus 2019 melalui akun instagram pribadinya. Alasan pemindahan ibu kota telah diumumkan pemerintah melalui konferensi pers pada tanggal 30 april 2019, yaitu sebagai berikut:

1. Mengurangi beban jakarta dan jabodetabek.
2. Mendorong pemerataan pembangunan ke wilayah indonesia bagian timur.
3. Mengubah maindset pembangunan dari jawa centris menjadi indonesia centris.
4. Memiliki ibu kota negara yang mempresentasikan identitas bangsa, kebhinekaan, dan penghayatan terhadap pancasila.
5. Memiliki ibu kota yang memiliki daya saing tinggi secara regional maupun internasional[7].

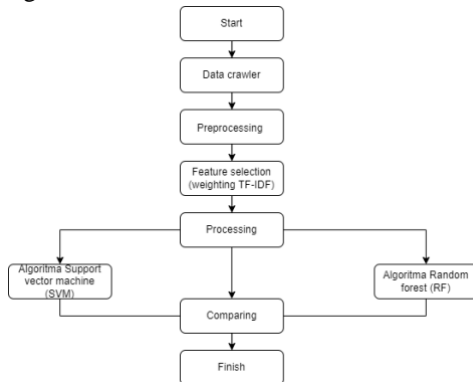
Belakangan ini marak muncul fenomena berita bohong (*hoax*) tentang isu pemindahan ibu kota negara (*IKN*) Indonesia yang beredar di dunia maya[8]. Adanya akses internet dan semakin menjamurnya sosial media menjadikan hoax begitu mudah dibuat dan disebarkan tanpa ada hukuman yang tegas kepada para pembuat dan penyebar hoax[5]. Sederhananya dapat kita artikan bahwa hoax adalah istilah untuk menggambarkan suatu berita bohong, fitnah, atau sejenisnya. Adapun maraknya hoax karena mudahnya akses informasi di media sosial dan banyaknya jenis media sosial seperti facebook, twitter, whatsapp, instagram, dan lain lain[6].

Dalam penelitian sebelumnya,[6] "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter" oleh Zidna Alhaq, Ali Mustopa, Sri Mulyatun, Joko Dwi Santoso pada tahun 2021 menjelaskan tentang analisa sentiment publik dalam aplikasi twitter dengan satu metode saja. Namun, dalam hal ini menjadi komparasi dengan penelitian ini yaitu, Deteksi Hoax terhadap *Issue* Pemindahan Ibu Kota Negara (*IKN*) Indonesia Dengan Menggunakan Algoritma Support Vector Machine (*SVM*) Dan Random Forest (*RF*). Dalam penelitian ini,peneliti menjelaskan berbagai bahaya hoax dan memberikan stimulasi cara melakukan deteksi dan analisa berita hoax dengan metode Machine learning.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan teknologi data *mining* seperti mengklasifikasikan analisisa sentimen publik pada postingan *tweet* tersebut bersifat positif, negatif. Data yang diperoleh dari hasil *Crawling* data *tweet* pada aplikasi Twitter kemudian diproses dan di input kedalam *database*, kemudian diklasifikasikan menggunakan dua algoritma yaitu algoritma *Support vector machine* (*SVM*) dan *Random forest* (*RF*). Metode pengumpulan data menggunakan data yang didapatkan dari hasil *crawling* di twitter berdasarkan dengan *keyword* yang dimasukkan dengan memanfaatkan *search API* (*Application Programming Interface*) yang disediakan oleh Twitter. Pada penelitian ini, metode yang digunakan untuk mendapatkan data informasi

dan untuk menyelesaikan masalah adalah sebagai berikut :



Gambar 1. 1 Alur Metodologi penelitian

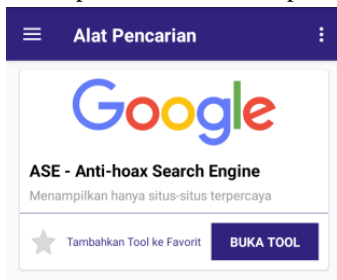
## 2.1 Studi literatur

Metode pengumpulan data dengan metode studi literatur dilakukan dengan pengumpulan dari berbagai jurnal, makalah, buku, maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya mendeteksi hoax menggunakan algoritma *support vector machine* (SVM) dan *random forest* (RF).

## 2.2 Teknik pengumpulan data

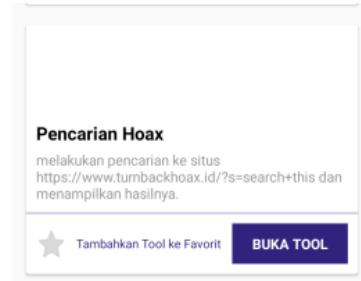
Pengumpulan data *Tweet* dari Twitter menggunakan aplikasi *Rapid Miner* dengan query #Ibukotanegara dan #IKNNusantara. Pengumpulan dataset *hoax* dan *non hoax* dilakukan dengan cara mengumpulkan data menggunakan aplikasi *hoax buster*, dengan pemanfaatan beberapa operator dan *tols* yang tersedia pada aplikasi *Hoax buster tols* (HBT). Berikut ini merupakan *tols* yang digunakan:

- 1.Operator alat pencarian pada *tols* ASE (Anti- Hoax Search Engine) untuk menampilkan informasi terpercaya.



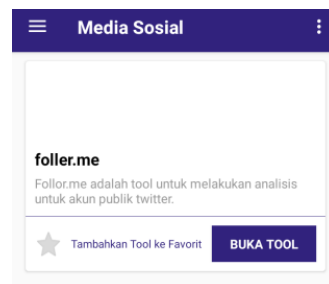
Gambar 1. 2 Tools Anti hoax search engine

- 2.Operator alat pencarian pada *tols* pencari hoax yang digunakan untuk pencarian situs hoax dan menampilkan hasilnya.



Gambar 1. 3 Tools Pencarian hoax

- 3.Operator media sosial pada *tols* foller.me untuk melakukan analisa akun publik twitter penyebar berita hoax



Gambar 1. 4 Tols analisa penyebar hoax

## 2.3 Teknik analisa data

1. Pelabelan data

Pelabelan data dilakukan dengan mengumpulkan 2 (dua) dataset berupa tabel yaitu, pertama melakukan pelabelan data tweet dengan label hoax dan non hoax secara manual berdasarkan hasil analisa informasi data tweet menggunakan *hoax buster tols* (HBT). Kemudian selanjutnya melakukan pelabelan sentimen tweet dengan label positif dan negatif secara acak (*random*).

2. Data preprocessing

processing dilakukan dengan tujuan merubah tweet atau data teks tak terstruktur sehingga menjadi data terstruktur untuk kebutuhan analisis sentimen topik pindah ibu kota. Preprocessing pada penelitian ini,

menerapkan serangkaian tahap secara berurutan yaitu :

1. *Case folding* merupakan tahap untuk merubah unsur huruf kapital dalam pada dokumen menjadi standar, yakni *lowercase*.
2. *Cleaning* digunakan untuk membersihkan karakter seperti https:, @, # dan link URL.
3. *Tokenizing* dilakukan untuk memecah kalimat dalam dokumen menjadi kata.
4. *Stopword Removal* dilakukan untuk penghapusan kata dengan kandungan informasi rendah.
5. *Stemming* dilakukan untuk menghilangkan imbuhan baik awalan maupun akhiran sehingga kembali menjadi bentuk kata dasar.

## 2.4 Pembobotan TF-IDF

Dalam tahapan ini merupakan pembobotan kata dalam kata tweet yang tersusun menggunakan Term Frequency (TF) dengan Inverse Document Frequency (IDF). barometer bobot kata dilakukan dengan mencari hasil kata dalam tweet kemunculan masing-masing kata, kemudian dilakukan dengan visualisasi word cloud.

$$w_{i,j} = tf_{i,j} \times \log \left( \frac{N}{df_i} \right)$$

Keterangan:

$tf_{i,j}$  = Banyaknya kata  $-i$  pada dokumen ke- $j$

$N$  = Total dokumen

$df_i$  = Banyaknya dokumen yang mengandung kata ke- $i$

## 2.5 Word cloud

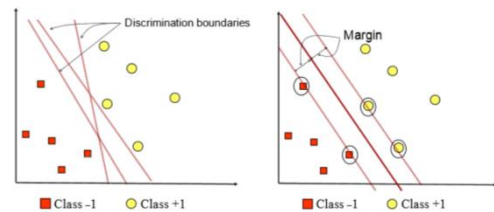
Visualisasi *word cloud* merupakan visualisasi data yang umumnya berupa teks. Visualisasi *word cloud* dapat menentukan bobot kemunculan data teks sesuai frekuensi munculnya data teks tersebut.



Gambar 1. 5 Visualisasi Wordcloud

## 2.6 Klasifikasi Support vector machine (SVM)

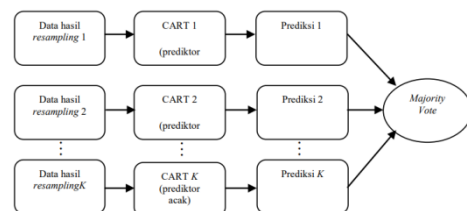
Tata cara *Support Vector Machine* (SVM) ialah tata cara yang bisa digunakan menganalisis sentimen pada penelitian ini. Hasil yang hendak ditetapkan dengan tata cara ini merupakan klasifikasi kelas positif serta kelas negatif yang didapat bersumber pada bobot pada tiap fitur dokumen teks.



Gambar 1. 6 Ilustrasi Hyperplan SVM

## 2.7 Klasifikasi Random forest (RF)

Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari simpul akar, simpul internal dan simpul daun dengan cara memperoleh atribut dan data secara acak sesuai dengan aturan yang berlaku, atribut dan informasi nilai perolehan.



Gambar 1. 7 Ilustrasi Random Forest

## 2.8 Komparasi performa klasifikasi

Proses komparasi klasifikasi ini dilakukan dengan membandingkan performa dari kedua algoritma yaitu *support vector machine* (SVM) dan *Random forest* (RF). Perbandingan ini dilakukan dengan beberapa tahapan, yaitu



Gambar 1. 8 Alur proses komparasi

### 1. Akurasi

Akurasi diartikan sebagai tingkat kedekatan nilai dari hasil prediksi yang benar dari nilai aktual. Dalam perhitungan akurasi menggunakan *confusion matrix* untuk

menghitung persen akurasi dari model klasifikasi pada Rumus berikut :

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

2. Presisi

Presisi adalah perbandingan untuk mengukur keakuratan seluruh hasil dataset dan mengetahui hasil klasifikasi kategori yang sebenarnya. Presisi merupakan hasil dari perhitungan yang benar (TP) dibagi dengan jumlah data yang teridentifikasi oleh sistem. Rumus presisi sebagai berikut :

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\%$$

3. Recall

Tingkat kesuksesan sebuah sistem dalam mendeteksi suatu kelompok menjadi sebuah parameter *recall*. *Recall* merupakan perhitungan dari pembagian antara jumlah data yang benar dengan jumlah yang seharusnya. Rumus *recall* sebagai berikut :

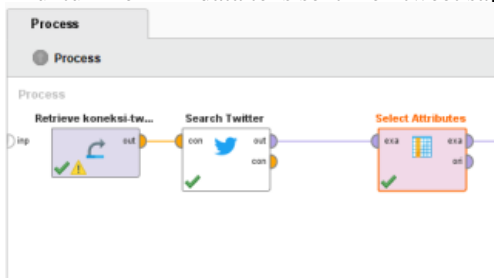
$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$$

Keterangan Rumus : TP (*True Positive*)  
TN (*True Negative*)  
FP (*False Positive*)  
FN (*False Negative*)

3. HASIL DAN PEMBAHASAN

3.1 Sumber data

Proses penambangan data dilakukan pada bulan januari-februari. Sebelum data penelitian ini dihasilkan, proses awal melakukan sinkronisasi untuk mengkoneksikan aplikasi twitter dengan aplikasi Rapid miner sehingga *clawer* data yang dihasilkan secara otomatis dengan menggunakan *query* #ibukotanegara dan #IKNNusantara yang kemudian menambahkan operator select attribute untuk memilih data teks sentimen tweet saja



Gambar 1. 9 Crawling data tweet

Data yang dihasilkan pada penelitian ini menggunakan sebanyak 1.500 (seribu lima ratus) data dengan menggabungkan hasil *crawling* data tweet dengan *query* yang berbeda kemudian disimpan dalam format *Comma Separated Values* (CSV). Operator yang digunakan pada proses ini ialah :

1. Search Twitter

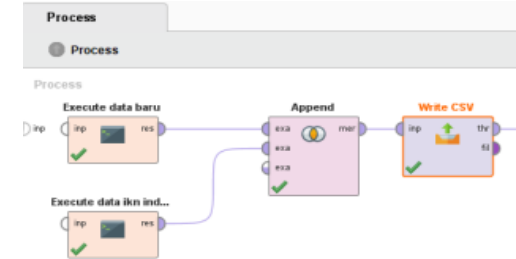
Pada operator ini merupakan operator utama dalam melakukan penambangan data (data *crawler*). Pada operator ini dapat terintegrasi pada aplikasi twitter dengan melakukan sinkronisasi melalui pembuatan koneksi pada view repository terdapat *tools connections* yang digunakan untuk membuat koneksi pada aplikasi twitter. Dalam *crawling* data ini menggunakan *input query* #ibukotanegara dan #IKNNusantara pada view parameters.

2. Select attributes

Digunakan untuk memilih data yang diperlukan. Dalam hal ini data yang diperlukan ialah *text* data sentimen.

3.2 Penggabungan data

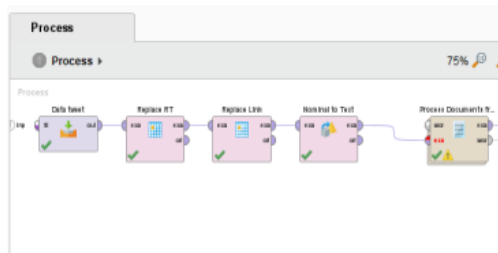
Tahap penggabungan data ini dilakukan dengan menggunakan *operator execute process*. Dalam operator ini terdapat data hasil *crawling* berupa *query* #ibukotabaru dan #IKNNusantara. Dalam penggabungan data hasil *crawling* tersebut membutuhkan *operator support* atau pendukung, yaitu operator *append* yang dalam fungsi utama dari operator ini merupakan operator penggabung data. Tahap akhir dalam proses penggabungan data ini memerlukan *operator write CSV* yang digunakan untuk menyimpan hasil penggabungan data dalam bentuk CSV. Gambar berikut ini merupakan simulasi proses penggabungan data dengan *query* berbeda yang.



Gambar 1. 10 Penggabungan Data

### 3.3 Preprocessing data

Proses pengolahan data dilakukan dengan beberapa tahapan, yakni proses sterilisasi data tweet dengan menggunakan serangkaian operator yang ada pada aplikasi Rapid miner. Pada proses hasil sterilisasi data ini menghasilkan data bersih dari beberapa elemen kata RT (Retweet), Hastag, mention, link website dan attribute kata yang dianggap tidak penting menggunakan operator Replace/Cleaning.



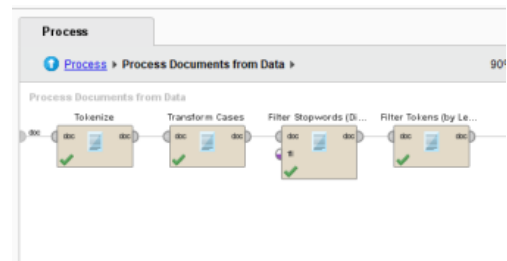
Gambar 1. 11 Preprocessing data

Selanjutnya dilakukan tahap sterilisasi data dengan menggunakan beberapa operator yang dikemas dalam operator sub proses untuk mengefisiensi tampilan pada view process di aplikasi Rapid miner. Berikut ini merupakan penjelasan beberapa operator yang dikemas pada operator sub proses yang digunakan :

1. *Tokenize* untuk pemisahan kata dalam kalimat. Proses tokenisasi pada teks ini merupakan pecahan kalimat tweet menjadi pemetakan karakter atau satuan kata yang sesuai dengan kebutuhan. Pada proses tokenisasi dilakukan filterisasi data untuk mengambil kata yang memiliki minimal 3 karakter dan maksimal 25 karakter, langkah ini dilakukan untuk memaksimalkan proses analisa. Untuk kata yang kurang dari 4 karakter atau lebih dari 25 karakter akan dihapus secara otomatis.
2. *Transform cases/case folding* untuk mengubah kata menjadi *lower case*. Proses ini akan mengubah text yang masih mengandung huruf uppercase atau kapital menjadi huruf lowercase atau huruf kecil semua. Dalam hal ini bertujuan untuk melakukan proses

model klasifikasi yang terdapat keseragaman huruf agar tidak terjadi kesalahan pada proses selanjutnya.

3. Filter stopwords/Stopword removal dengan menggunakan dictionary bahasa indonesia untuk menghapus kata berbobot rendah. Tahap ini mengambil kata-kata penting dari hasil tokenisasi dengan menggunakan algoritma stoplist (membuang kata kurang penting). Stopword adalah kata yang tidak deskriptif yang dibuang dalam pendekatan bahasa indonesia. Berikut contoh stopwords bahasa adalah “yang”, “dan”, “di”, “dari”.
4. Filter token (bylength) untuk menghapus kata-kata terlalu panjang dan terlalu pendek. Pada proses ini merupakan penghapusan kata terlalu pendek dan tidak dianggap penting. Seperti contoh “ya”, “aq”



Gambar 1. 12 Sub proses

### 3.4 Pelabelan data

Pelabelan data dilakukan secara manual dengan ketentuan label hoax dengan alasan mengandung kata atau isi berita yang tidak sesuai dengan fakta sebenarnya berasal dari situs TurnBackHoax.id, sedangkan berita bukan hoax diberi label yang berasal dari situs berita online kompas.com dan detik.com yang merupakan berita dengan isi berita mengandung fakta sebenarnya yang terjadi pada masyarakat. Dalam hal ini menggunakan tools ASE (Anti hoax search engine), tools Foller.Me dan tools pencari hoax pada aplikasi Hoax Buster Tools (HBT) yang terintegrasi dengan MAFINDO (masyarakat anti fitnah indonesia) atau website TurnBackHoax.id dan untuk mendeteksi hoax sehingga dalam hasil analisa ini dibentuk data set informasi atau

berita termasuk hoax atau tidak. Kemudian dalam pelabelan data sentimen tweet dilakukan secara manual dengan ketentuan sentimen positif berupa unsur dukungan dan tidak mengandung unsur kata kasar, begitupun sebaliknya dalam pelabelan sentimen negatif.

**Table 1 Pelabelan hoax dan non hoax**

No	Tweet	Label
1	Peneliti inggris sebut calon ibu kota baru RI rawan Tsunami	Non Hoax
2	Dana haji untuk pembangunan IKN	Hoax
3	Rocky gerung mengatakan indonesia akan bangkrut jika memaksakan bangun IKN	Non hoax
4	Pemindahan IKN ke Kaltim Perbanyak Kesempatan Kerja Bagi Masyarakat	Hoax
5	766 persen masyarakat Indonesia puas dengan kinerja Presiden Jokowi IKN Untuk Negeri	Non Hoax
6	Hasil Survei APSSI soal IKN, Sebanyak 48,2 Persen Masyarakat Minta Ditunda, Dominan Dampak Negatif	Non Hoax
7	Ide Pemindahan Ibu Kota Sudah Ada Sejak Era Soekarno	Non Hoax
8	IKN Nusantara akan gunakan kendaraan tanpa awak sebagai transportasi publik	Non Hoax
9	Peringati hari Kebangkitan Nasional, 8 organisasi pemuda lintas agama ikrar kebangsaan di Titik Nol IKN	Non Hoax
10	Terdapat enam kluster ekonomi sebagai penggerak utama untuk mewujudkan visi Superhub Ekonomi IKN	Non Hoax

**Table 2 Pelabelan Sentimen tweet**

No	Tweet	Label
1	Melalui pembangunan kelestarian lingkungan tetap akan dijaga dan pemerataan ekonomi akan terwujud di Kalimantan.	Positif
2	Lanjutkan untuk Indonesia Maju ibu kota Baru.	Positif
3	Dukung IKN Nusantara Penggerak Ekonomi Indonesia di Masa Depan.	Negatif
4	Pemindahan IKN ke Kaltim Perbanyak Kesempatan Kerja Bagi Masyarakat.	Negatif
5	Ratusan rakyat tak butuh ibu kota negara yg baru. Rakyat hanya butuh Rezim baru.	Negatif
6	Pantas Elon Musk Kepincut Investasi di Indonesia Pemerintah Tawarkan Kawasan dekat IKN.	Negatif
7	Pembangunan IKN Nusantara di Kalimantan Timur merupakan wujud betapa pentingnya posisi Kaltim secara geografis ekonomis hingga identitas sosial budaya bagi Indonesia.	Positif
8	Minyak naik itu karna maksa pingin bangun IKN investor kabur karna indonesia indeks korupsinya naik terus demokrasi menurun.	Negatif
9	Iya njir masih ada aja yang nolak IKN Untuk Negeri padahal udah jelas keren banget dan manfaat untuk indonesia.	Positif
10	Alhamdulillah ditangan Jokowi Indonesia menjadi lebih baik IKN Untuk Negeri.	Positif

### 3.6 Pembobotan TF-IDF

Dalam pembobotan ini dilakukan untuk mengetahui banyaknya kemunculan kata pada data tweet tersebut. Berikut merupakan hasil dari proses pembobotan TF-IDF.

**Table 3 Hasil pembobotan TF-IDF**

Word	Attribute name	Total occurences	Documen occurences
Indonesia	Indonesia	532	468
Ibu kota negara	Ibukotanegara	422	422
Pimpin	Pimpin	229	225
Pembangunan	Pembangunan	156	149
Nusantara	Nusantara	145	140
Negeri	Negeri	113	112
Kota	Kota	56	48
Dukung	Dukung	47	47
Maju	Maju	47	47
Semangat nusantara	Semangat nusantara	45	45
Pemindahan	Pemindahan	44	44
Kalimantan	Kalimantan	38	32
Masyarakat	Masyarakat	38	37
Pemerataan	Pemerataan	38	38
Dana	Dana	35	35

Berdasarkan hasil pembobotan TF-IDF ini ditemukan hasil kata indonesia dengan total *occurences* (kemunculan data) sebanyak 532 dari total *documen occurences* (total dokumen) sebanyak 468, kemudian pada posisi terbanyak kedua yakni kata ibu kota negara dengan *total*

*occurences* (kemunculan data) sebanyak 422 dari total *documen occurences* (total dokumen) sebanyak 422.

### 3.7 Visualisasi wordcloud

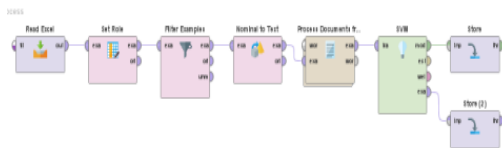
Visualisasi dilakukan untuk mengetahui teks kemuculan kata pada data hasil proses pembobotan TF-IDF. Visualisasi wordcloud ini merupakan hasil olah data pada pembobotan TF-IDF yang dilakukan sebelumnya. Berdasarkan hasil worcloud ini di setting dengan menggunakan parameter netral, sehingga *size text* yang muncul setara.



Gambar 1. 13 visualisasi wordcloud

### 3.8 Klasifikasi Support vector machine (SVM)

Implementasi algoritma SVM dilakukan untuk mengklasifikasi data sentimen tweet dan data tweet berupa unsur berita kelas hoax dan non hoax dalam penelitian ini dilanjutkan dengan menambahkan operator *apply model* agar dapat membaca data. Dalam proses klasifikasi dilakukan penentuan prediksi data dengan melakukan pembelajaran pada algoritma sehingga data prediksi valid. Berikut merupakan road map proses pembelajaran pada algoritma dalam menentukan prediksi pada label yang dibutuhkan.



Gambar 1. 14 proses pembelajaran algoritma

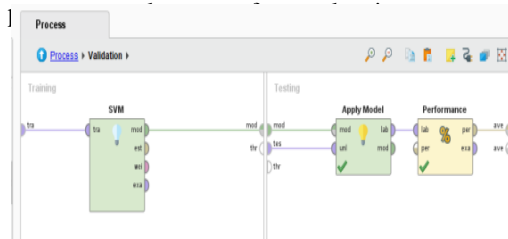
Pada proses pembelajaran pada algoritma SVM tersebut terdapat komposisi operator sistem seperti gambar diatas. Proses ini memerlukan operator store yang digunakan untuk menyimpan hasil belajar dari algoritma tersebut. Hasil belajar algoritma ini disimpan pada repository proses dan repository data. Dalam melakukan input data training dari hasil

pelabelan hoax manual melalui hasil deteksi pada aplikasi hoax buster, data dalam pelabelan manual tersebut sebanyak 40% data. Dalam proses ini melakukan pembelajaran pada algoritma svm sehingga dapat di temukan hasil prediksi terbaik pada labeling data dan proses



Gambar 1. 15 proses klasifikasi prediksi data

Gambar ini merupakan road map proses pengujian data training dan data testing untuk menemukan hasil prediksi labeling yang tepat pada data sentimen dan data hoax. Pada proses ini hasil prediksi klasifikasi SVM pada data yang digunakan kemudian disimpan dengan format CSV sehingga dapat dilanjutkan pada



Gambar 1. 16 Road Map Klasifikasi SVM

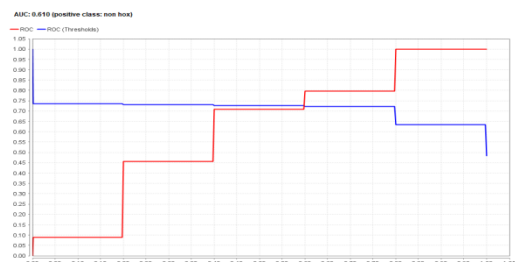
Hasil dari klasifikasi data hoax dari sentimen tweet dari total data sebanyak 169 data pada algoritma Support vector machine dengan menggunakan data training sebanyak 30% dan data testing sebanyak 70% sehingga mendapatkan hasil Performance Vector dengan bobot nilai accuracy: 95.24%, bobot nilai precision: 95.18% (positive class: non hoax), bobot nilai recall: 100.00% (positive class: non hoax), bobot nilai AUC (optimistic): 0.610 (positive class: non hoax) AUC: 0.610 (positive class: non hoax), bobot nilai AUC (pessimistic): 0.610 (positive class: non hoax). Dalam proses klasifikasi ini di temukan data hoax sebanyak 10 berita dan 159 berita non hoax.

Hasil dari klasifikasi data sentimen tweet dari total data sebanyak 366 data pada algoritma Support vector machine dengan menggunakan

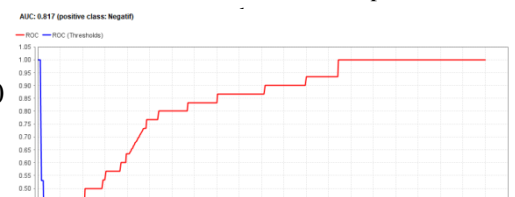
data training sebanyak 60% dan data testing sebanyak 40% sehingga mendapatkan hasil Performance Vector dengan bobot nilai accuracy: 84.07%, bobot nilai precision: 66.67% (positive class: Negatif), bobot nilai recall: 6.67 % (positive class: Negatif), bobot nilai AUC (optimistic): 0.819 (positive class: Negatif) AUC: 0.817 (positive class: Negatif), bobot nilai AUC (pessimistic): 0.816 (positive class: Negatif) dengan susunan Confusion Matrix dalam tabel berikut :

Table 4 Confusion Matrix

Klasifikasi hoax	Accuracy	True	Hoax	Non hoax
		Hoax	1	0
		Non hoax	4	79
	Precision	True	Hoax	Non hoax
		Positif	1	0
		Non hoax	4	79
	Recall	True	Hoax	Non hoax
		Hoax	1	0
	Non hoax	4	79	
Klasifikasi sentiment tweet	Accuracy	True	Positif	Negatif
		Hoax	151	28
		Non hoax	1	2
	Precision	True	Positif	Negatif
		Positif	151	28
		Negatif	1	2
	Recall	True	Positif	Negatif
		Positif	151	28
	Negatif	1	2	



Gambar 1. 17 ROC Hasil klasifikasi prediksi berita

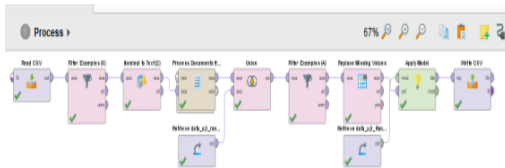


\

Gambar 1. 18 ROC hasil klasifikasi sentiment SVM

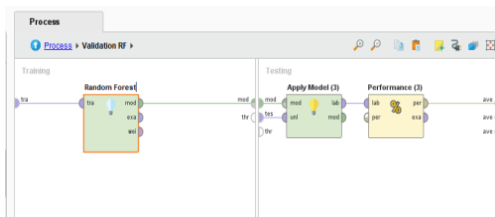
### 3.9 Klasifikasi Random forest (RF)

Implementasi algoritma SVM dilakukan untuk mengklasifikasi data penelitian dilanjutkan dengan menambahkan operator apply model agar dapat membaca data.



Gambar 1. 19 proses klasifikasi prediksi data

Gambar ini merupakan road map proses pengujian data training dan data testing untuk menemukan hasil prediksi labeling yang tepat pada data sentimen dan data hoax. Pada proses ini hasil prediksi klasifikasi Random forest qpada data yang digunakan kemudian disimpan dengan format CSV sehingga dapat dilanjutkan pada proses pengukuran performa algoritma.



Gambar 1. 20 Road Map Klasifikasi Random Forest

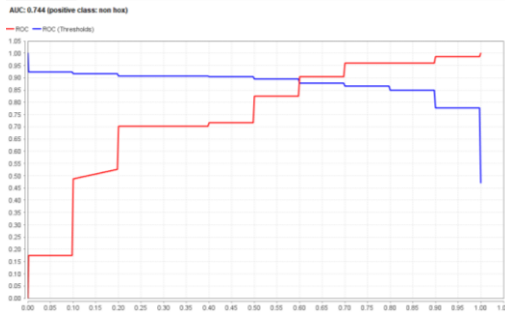
Hasil dari klasifikasi data hoax dari sentimen tweet dari total data sebanyak 169 data pada algoritma Support vector machine dengan menggunakan data training sebanyak 30% dan data testing sebanyak 70% sehingga mendapatkan hasil Performance Vector dengan bobot nilai accuracy: 86.90%, bobot nilai

precision: 87.95% (positive class: non hoax), bobot nilai recall: 98.65% (positive class: non hoax), bobot nilai AUC (optimistic): 0.746 (positive class: non hoax) AUC: 0.744 (positive class: non hoax), bobot nilai AUC (pessimistic): 0.742 (positive class: non hoax). Dalam proses klasifikasi ini di temukan data hoax sebanyak 21 berita dan 148 berita non hoax.

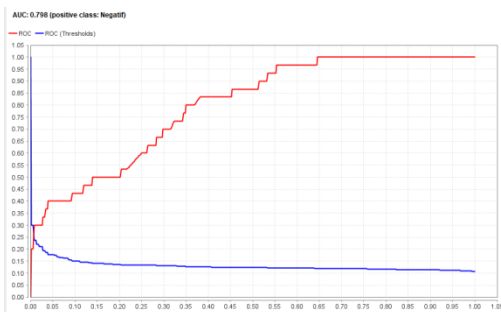
Hasil dari klasifikasi algoritma Random forest dengan menggunakan data training sebanyak 60% dan data testing sebanyak 40% sehingga mendapatkan hasil Performance Vector dengan bobot nilai accuracy: 83.52%, bobot nilai precision: *unknown* (positive class: Negatif), bobot nilai AUC (optimistic): 0.799 (positive class: Negatif) AUC: 0.798 (positive class: Negatif), bobot nilai AUC (pessimistic): 0.796 (positive class: Negatif) dengan susunan Confusion Matrix dalam tabel berikut :

Klasifikasi hoax	Accuracy	True	Hoax	Non hoax
		Hoax	0	1
		Non hoax	10	73
	Precision	True	Hoax	Non hoax
		Hoax	0	1
		Non hoax	10	73
Klasifikasi sentiment tweet	Recall	True	Hoax	Non hoax
		Hoax	0	1
		Non hoax	10	73
	Accuracy	True	Positif	Negatif
		Positif	152	30
		Negatif	0	0
Klasifikasi sentiment tweet	Precision	True	Positif	Negatif
		Positif	152	30
		Negatif	0	0
Klasifikasi sentiment tweet	Recall	True	Positif	Negatif
		Positif	152	30

		Negatif	0	0
--	--	---------	---	---



Gambar 1. 21 Hasil klasifikasi prediksi berita hoax

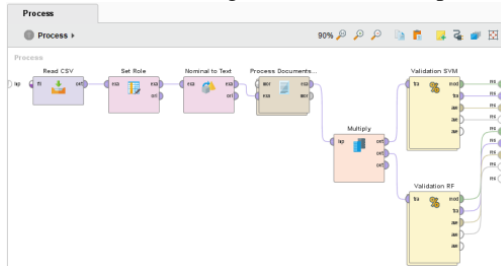


Gambar 1. 22 ROC Klasifikasi sentimen Random Forest

Performa akurasi, presisi, recall dan juga AUC (*Area under curve*) Pada perbandingan performa klasifikasi hoax dari kedua algoritma ini, algoritma Support vector machine mendapatkan nilai akurasi terbaik dengan bobot nilai *accuracy*: 95.24% sedangkan algoritma Random Forest mendapatkan bobot nilai *accuracy* 86.90%. bobot nilai *precision* dari algoritma SVM juga lebih baik dari pada algoritma Random forest yakni dengan bobot nilai sebesar 95.18 % begitupun bobot nilai recall dan AUC pada algoritma SVM lebih besar dari pada algoritma random forest. Pada performa klasifikasi sentiment tweet dari data yang digunakan, performa algoritma SVM lebih mendominasi bobot nilai yang dihasilkan, yaitu dengan bobot nilai akurasi 84.07%, presisi 66.67% , AUC 0.817. sedangkan algoritma random forest mendapat nilai akurasi sebesar 83.52%, bobot nilai precision: *unknown* AUC: 0.798. berikut ini merupakan diagram komparasi performa algoritma dalam proses klasifikasi hoax dan klasifikasi sentiment tweet pada data yang digunakan. Berikut merupakan diagram komparasi performa klasifikasi algoritma SVM dan Random forest.

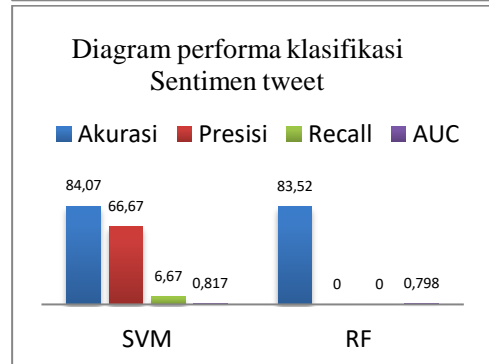
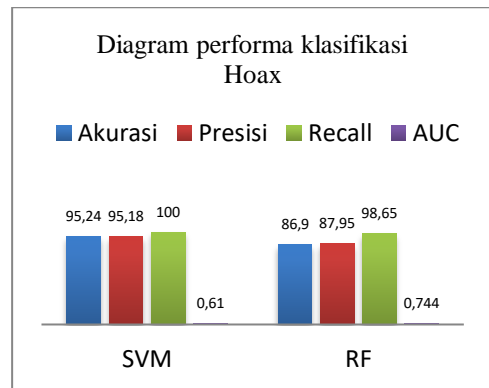
### 3.11 Komparasi performa algoritma

Pada proses komparasi algoritma SVM dan Random forest ini dilakukan dengan cara menyusul beberapa operator proses yang dalam hal ini operator utama yang digunakan ialah operator multiply. Multiply ini merupakan operator yang dapat melakukan penggabungan klasifikasi algoritma sehingga output yang dihasilkan sesuai dengan kebutuhan komparasi.



Gambar 1. 23 Road Map Komparasi Algoritma

Pada tahapan komparasi ini ditemukan hasil terbaik dari perbandingan performa dalam klasifikasi berita hoax dan performa klasifikasi sentiment tweet pada data yang digunakan.



#### 4. KESIMPULAN

Berdasarkan hasil analisa deteksi hoax menggunakan algoritma *support vector machine* (SVM) dan *random forest* (RF) terhadap *issue* pemindahan ibu kota negara (IKN) Indonesia, maka ada beberapa data yang dapat disimpulkan yaitu:

1. Berita yang ada pada sosial media twitter dapat diidentifikasi dan diklasifikasi apakah *fake* (palsu) atau *real* (bukan) dengan membuat klasifikasi TF-IDF pada algoritma *Support vector machine* (SVM) dan *Random forest* (RF).
2. Pemanfaatan API twitter dapat terintegrasi pada aplikasi Rapid miner sehingga dapat melakukan *crawling* data pada aplikasi twitter.
3. Algoritma yang dapat memberikan hasil terbaik dalam mengklasifikasi data berita hoax dengan nilai performa akurasi sebesar 95.24%, sedangkan algoritma *Random forest* mendapatkan hasil performa akurasi sebesar 86.90%.
4. Algoritma *support vector machine* memberikan kinerja terbaik pada proses klasifikasi data sentimen tentang *issue* pemindahan ibu kota negara (IKN) Indonesia dengan nilai performa akurasi sebesar 84.07%, sedangkan algoritma *Random forest* mendapatkan hasil performa akurasi sebesar 83.52%.
5. Data hasil crawler pada aplikasi Twitter bahwa 85% dalam sentimen tweet merupakan data sentimen positif. Sedangkan data berita hoax pada twitter bahwa 97% merupakan berita non hoax.

#### UCAPAN TERIMA KASIH

Dengan Nama Allah Yang Maha Pengasih lagi Maha Penyayang. Segala puja dan puji syukur kepada Allah SWT atas segala rahmat dan karunia yang telah dilimpahkan, sehingga penulisan proposal Tugas Akhir dengan judul: "ANALISA HOAX TERHADAP *ISSUE* PEMINDAHAN IBU KOTA NEGARA (IKN) INDONESIA MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *RANDOM FOREST* (RF)." dapat diselesaikan dengan baik. Proposal Tugas Akhir ini disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Program Studi Teknik

Informatika, Fakultas Teknik, Universitas Islam Madura.

Pada kesempatan ini penulis sampaikan terima kasih yang sedalam-dalamnya kepada :

1. Kedua orangtuaku Ayahanda (Moh. Yusuf) dan Ibunda (Juma'ati) tercinta yang telah mendidik penulis dari kecil hingga menjadi orang sukses.
2. Ibu Hoiriyah, M.kom dan Anwari, S.Kom.,MT atas bimbingan, kesabaran dan pendorong semangat dalam menyelesaikan tugas akhir ini.
3. Bapak dan Ibu dosen terima kasih atas bimbingan dan ilmu pengetahuan yang diberikan selama masa perkuliahan.
4. Rekan-rekan S1 di lingkungan Fakultas Teknik Universitas Islam Madura, terima kasih atas kebaikan dan kerjasamanya dalam penelitian ini.

Penulis menyadari bahwa dalam penulisan proposal Tugas Akhir ini masih jauh dari sempurna, untuk itu demi perbaikan dan penyempurnaan proposal Tugas Akhir, maka kritik dan saran sangat diharapkan. Besar harapan penulis bahwa buku Tugas Akhir ini dapat memberikan informasi dan manfaat bagi pembaca pada umumnya dan mahasiswa di lingkungan Fakultas Teknik pada khususnya.

#### DAFTAR PUSTAKA

- [1]soeyanto, olyn seftani. (2019). Berita Bohong ( hoax ) DI Media Sosial Waktu terjadinya kasus 22 mei dari Tindakan Manusia Dan Tatanan moral Subjektif. <https://doi.org/10.31227/osf.io/ek8y6>
- [2]Alhaq, Z., Mustopa, A., & Santoso, J. D. (n.d.). *PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER*.
- [6]Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 231–237. <https://doi.org/10.29207/resti.v5i2.2698>
- [5]Hani, ), Rochmanto, B., Mustikawati, E., & Hermanto, P. (n.d.). *Analisis Sentimen Pengguna Twitter terhadap Kebijakan Pemindahan Ibukota Indonesia Menggunakan Metode NBC dan SVM*. <http://website.com>

[4]Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, S., Zamachsari, F., Vangeran Saragih, G., Gata, W., Ilmu Komputer, M., Ilmu Komputer, F., & Nusa Mandiri Jakarta, S. (2017). Terakreditasi SINTA Peringkat 2 Analisis Sentimen Pindahan Ibu Kota Negara dengan Feature Selection Algoritma Naive Bayes dan Support Vector Machine. *Masa Berlaku Mulai*, 1(3), 504–512.

[8]Mutaqin, D. J., Muslim, M. B., & Rahayu, N. H. (2021). Analisis Konsep Forest City dalam Rencana Pembangunan Ibu Kota Negara. *Bappenas Working Papers*, 4(1), 13–29. <https://doi.org/10.47266/bwp.v4i1.87>

[9]Naufal Rizaldi, M., & Al Faraby, S. (n.d.). *JURNAL MEDIA INFORMATIKA BUDIDARMA Klasifikasi Argument Pada Teks dengan Menggunakan Metode Multinomial Logistic Regression Terhadap Kasus Pindahan Ibu Kota Indonesia di Twitter*. <https://doi.org/10.30865/mib.v4i4.2348>

[3]Safra, I. A., & Zuliarso, E. (n.d.). *ANALISA SENTIMEN PERSEPSI MASYARAKAT TERHADAP PEMINDAHAN IBUKOTA BARU DI KALIMANTAN TIMUR PADA MEDIA SOSIAL TWITTER*.

[7]Yuniasari, P., & Maspiyanti, F. (2021). Analisis Sentimen Data Tweet menggunakan Metode Support Vector Machine (Studi Kasus: Pindahan Ibukota Baru Republik Indonesia). In *Journal of Informatics and Advanced Computing* (Vol. 2, Issue 1).