

SKRIPSI

**KLASIFIKASI KONFLIK SOSIAL DI KABUPATEN WAJO
MENGUNAKAN METODE MODIFIED BALANCED
RANDOM FOREST (MBRF) PADA DATA TEXT BERBASIS
WEBSITE**

Disusun dan diajukan oleh:

**YANUARSYAH FITRAH INDRA
D121 18 1318**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024**

LEMBAR PENGESAHAN SKRIPSI

**KLASIFIKASI KONFLIK SOSIAL DI KABUPATEN WAJO
MENGUNAKAN METODE MODIFIED BALANCED RANDOM
FOREST (MBRF) PADA DATA TEXT BERBASIS WEBSITE**

Disusun dan diajukan oleh

**Yanuarsyah Fitrah Indra
D121181318**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian
Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada tanggal 24 Juli 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,

Dr.Eng. Zulkifli Tahir, S.T., M.Sc.
NIP 19840403 201012 1 004

Muhammad Alief Fahdal Imran
Oemar, ST., M.Sc
NIP 19940522 202204 3 001

Ketua Program Studi,

Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN. Eng.
NIP. 19750716 200212 1 004



PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;
Nama : Yanuarsyah Fitrah Indra
NIM : D121181318
Program Studi : Teknik Informatika
Jenjang : SI

Menyatakan dengan ini bahwa karya tulisan saya berjudul

KLASIFIKASI KONFLIK SOSIAL DI KABUPATEN WAJO MENGUNAKAN METODE MODIFIED BALANCED RANDOM FOREST (MBRF) PADA DATA TEXT BERBASIS WEBSITE

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, Juli 2024

Yang Menyatakan

Yanuarsyah Fitrah Indra

ABSTRAK

YANUARSYAH FITRAH INDRA. *KLASIFIKASI KONFLIK SOSIAL DI KABUPATEN WAJO MENGGUNAKAN METODE MODIFIED BALANCED RANDOM FOREST (MBRF) PADA DATA TEXT BERBASIS WEBSITE* (dibimbing oleh Dr.Eng. Zulkifli Tahir, S.T., M.Sc. dan Muhammad Alief Fahdal Imran Oemar, ST., M.Sc).

Negara Indonesia merupakan negara yang memiliki berbagai keberagaman sehingga sangat rentan untuk terjadinya sebuah konflik, baik konflik sesama individu, konflik terhadap individu dan kelompok, serta konflik sesama kelompok. Hal ini memberikan keresahan tersendiri bagi warga lainnya, sehingga sosial media menjadi salah satu sasaran empuk warga untuk melaporkan keresahan tersebut. Sistem yang dapat dibangun untuk meminimalisir dan mempercepat pemerintah mengambil tindakan ialah dengan analisis sentimen dengan memanfaatkan sentimen yang tersebar di media sosial.

Penelitian ini dilakukan dengan tujuan untuk memberikan masukan serta dasar yang kuat bagi pemerintah Kabupaten Wajo dalam pengambilan keputusan terkait konflik sosial yang terjadi. Diharapkan pemerintah juga dapat lebih memahami akar permasalahan serta mampu menentukan prioritas permasalahan yang perlu diselesaikan terlebih dahulu.

Model yang digunakan dalam penelitian ini adalah *Modified Balanced Random Forest* (MBRF) dimana model ini merupakan peningkatan dari model *Random Forest* dengan menambahkan metode *resampling* sebagai cara untuk mengatasi ketidak-seimbangan data. Penelitian ini juga membandingkan kinerja algoritma MBRF dengan algoritma dasarnya, *Random Forest*.

Berdasarkan hasil pengujian, kata "bupati" dan "anggota DPRD" mendominasi pada *wordcloud* dataset, sehingga mencerminkan sentimen yang kuat terkait hal tersebut di internet. Selain itu, terdapat juga kata "bunuh" dan "pukul" yang signifikan sehingga menunjukkan adanya kasus-kasus penganiayaan dan penghilangan nyawa yang diduga berkaitan dengan konflik sengketa tanah. Hasil pengujian analisis sentimen menunjukkan bahwa model *Modified Balanced Random Forest* menunjukkan akurasi prediksi yang signifikan, mencapai 87.9%, melampaui model *Random Forest* yang mencapai 83.2%. Ini menunjukkan keunggulan kinerja pada model *Modified Balanced Random Forest* jika dibandingkan dengan *Random Forest* dalam analisis sentimen terhadap dataset konflik sosial.

Kata Kunci: Konflik Sosial, Wajo, *Modified Balanced Random Forest*, *Resampling*, *Particle Swarm Optimization*, Analisis Sentimen.

ABSTRACT

YANUARSYAH FITRAH INDRA. *CLASSIFICATION ON SOCIAL CONFLICTS IN WAJO DISTRICT USING MODIFIED BALANCED RANDOM FOREST (MBRF) METHOD ON WEB-BASED USING TEXT DATA* (supervised by Dr.Eng. Zulkifli Tahir, S.T dan Muhammad Alief Fahdal Imran Oemar, ST., M.Sc)

Indonesia is a country with a lot of diversity, which makes it prone to various types of conflicts, whether between individuals, between individuals and groups, or between groups. This causes concern among citizens, who often turn to social media to express their worries. A useful system to help reduce these conflicts and enable the government to act quickly is sentiment analysis, which uses the sentiments shared on social media.

This research is conducted with the aim of providing input and a strong foundation for the Wajo Regency government in making decisions related to ongoing social conflicts. It is hoped that the government will also better understand the root causes of the problems and be able to prioritize which issues need to be resolved first.

The model used in this research is the Modified Balanced Random Forest (MBRF), which is an enhancement of the Random Forest model by adding a resampling method to address data imbalance. This research also compares the performance of the MBRF algorithm with its basic algorithm, Random Forest.

Based on the test results, the words "bupati" (regional head) and "anggota DPRD" (members of regional legislative council) dominate the dataset's word cloud, reflecting strong sentiments related to these topics on the internet. Additionally, significant occurrences of the words "bunuh" (kill) and "pukul" (hit) suggest cases of violence and loss of life likely related to land dispute conflicts. Sentiment analysis results indicate that the Modified Balanced Random Forest model achieved significant prediction accuracy, reaching 87.9%, surpassing the Random Forest model which achieved 83.2%. This demonstrates the superior performance of the Modified Balanced Random Forest model compared to the Random Forest in sentiment analysis of social conflict datasets.

Keywords: Social Conflict, Wajo, Modified Balanced Random Forest, Resampling, Particle Swarm Optimization, Sentiment Analysis.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	Error! Bookmark not defined.
PERNYATAAN KEASLIAN.....	Error! Bookmark not defined.
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
DAFTAR SINGKATAN DAN ARTI SIMBOL	ix
DAFTAR LAMPIRAN.....	x
KATA PENGANTAR	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Ruang Lingkup.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Konflik Sosial	5
2.2 Media Sosial.....	5
2.3 <i>Crawling</i> Data	6
2.4 <i>Text Mining</i>	6
2.5 Analisis Sentimen.....	6
2.6 Algoritma <i>Random Forest</i>	7
2.7 Algoritma <i>Modified Balanced Random Forest</i> (MBRF)	7
2.8 <i>Term Frequency — Inverse Document Frequency</i> (TF-IDF)	9
2.9 <i>Selection Feature Particle Swarm Optimization</i> (PSO).....	10
2.10 <i>Confusion Matrix</i>	11
BAB III METODE PENELITIAN.....	13
3.1 Waktu dan Lokasi Penelitian	13
3.2 Instrumen Penelitian	13
3.3 Tahapan Penelitian	14
3.4 Rancangan Sistem	15
3.5 Teknik Pengambilan Data	16
3.6 Preprocessing Data.....	17
3.7 Perhitungan <i>Extraction Feature</i> TF-IDF	24
3.8 <i>Selection Feature Particle Swarm Optimization</i>	26
3.9 Analisis Sentimen dengan <i>Random Forest</i>	28
3.10 Analisis Sentimen dengan <i>Modified Balanced Random Forest</i> (MBRF).....	30
3.11 Sistem Informasi Pelaporan Konflik Sosial	33
BAB IV HASIL DAN PEMBAHASAN	35
4.1 Pengumpulan Data	35
4.2 Pelabelan Data.....	36
4.3 <i>Preprocessing</i> Data	38
4.4 Analisis Sentimen dengan <i>Random Forest</i>	43
4.5 Analisis Sentimen dengan <i>Modified Balanced Random Forest</i>	45

4.6 Analisis Perbandingan Data Uji	47
BAB V KESIMPULAN DAN SARAN.....	51
5.1 Kesimpulan	51
5.2 Saran.....	52
DAFTAR PUSTAKA	53

DAFTAR GAMBAR

Gambar 1. Lokasi Penelitian	13
Gambar 2. Tahapan penelitian	14
Gambar 3. Rancangan sistem analisis sentimen	15
Gambar 4. Rancangan sistem pembangunan sistem informasi.....	16
Gambar 5. <i>Flowchart Selection Feature PSO</i>	27
Gambar 6. <i>Flowchart algoritma Random Forest</i>	29
Gambar 7. <i>Flowchart algoritma Modified Balanced Random Forest (MBRF)</i>	30
Gambar 8. <i>Dataset (a) sebelum dan (b) setelah menjalani undersampling</i>	31
Gambar 9. <i>Dataset (a) sebelum dan (b) setelah menjalani oversampling</i>	31
Gambar 10. Halaman utama website pelaporan konflik sosial.....	33
Gambar 11. Halaman form pelaporan konflik sosial.	34
Gambar 12. Hasil <i>scraping</i> pada sosial media Twitter dan Youtube.....	35
Gambar 13. Grafik dataset berdasarkan label	36
Gambar 14. <i>Confusion Matrix</i> dari <i>Random Forest</i>	44
Gambar 15. <i>Confusion Matrix</i> dari <i>Modified Balanced Random Forest</i>	45
Gambar 16. Wordcloud keseluruhan dataset	49

DAFTAR TABEL

Tabel 1. <i>Confusion Matrix</i>	11
Tabel 2. Sampel data	17
Tabel 3. Sampel data setelah melalui tahapan <i>cleaning</i>	18
Tabel 4. Sampel data setelah melalui tahapan <i>case folding</i>	19
Tabel 5. Sampel data setelah melalui tahapan <i>tokenization</i>	19
Tabel 6. Contoh kata pada kamus normalisasi.....	20
Tabel 7. Sampel data setelah melalui tahapan <i>normalization</i>	21
Tabel 8. Sampel data setelah melalui tahapan <i>stopwording</i>	23
Tabel 9. Sampel data setelah melalui tahapan <i>stemming</i>	24
Tabel 10. Sampel Data	24
Tabel 11. Sampel data hasil TF-IDF	25
Tabel 12. Rincian <i>dataset</i> setelah dilakukan pelabelan.	37
Tabel 13. Sentimen sebelum dan sesudah proses <i>cleaning</i>	38
Tabel 14. Sentimen sebelum dan sesudah <i>proses case folding</i>	39
Tabel 15. Sentimen sebelum dan sesudah proses <i>tokenization</i>	39
Tabel 16. Sentimen sebelum dan sesudah proses <i>normalization</i>	40
Tabel 17. Sentimen sebelum dan sesudah proses <i>stopwording</i>	41
Tabel 18. Sentimen sebelum dan sesudah proses <i>stemming</i>	42
Tabel 19. <i>Dataset</i> akhir setelah melalui proses pembersihan data.....	43
Tabel 20. Evaluasi model <i>Random Forest</i>	44
Tabel 21. Evaluasi model <i>Modified Balanced Random Forest</i>	46
Tabel 22. Perbandingan hasil kinerja antara dua metode klasifikasi.	47
Tabel 23. Contoh <i>dataset</i> yang ambigu untuk diprediksi.....	48

DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
BRF	<i>Balanced Random Forest</i>
gBest	<i>Global Best</i>
lBest	<i>Local Best</i>
MBRF	<i>Modified Balanced Random Forest</i>
NLP	<i>Natural Language Processing</i>
NLTK	<i>Natural Language Tool Kit</i>
PSO	<i>Particle Swarm Optimization</i>
RF	<i>Random Forest</i>
TF-IDF	<i>Term-Frequency – Inverse Document Frequency</i>

DAFTAR LAMPIRAN

Lampiran 1. Dataset	57
Lampiran 2. Source Code Pre-processing	58
Lampiran 3. Source Code <i>Scratch Modified Balanced Random Forest</i>	61

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT atas berkat dan karunia-Nya sehingga skripsi dengan judul “**Klasifikasi Konflik Sosial di Kabupaten Wajo Menggunakan Metode Modified Balanced Random Forest (MBRF) Pada Data Text Berbasis Website**” ini dapat terselesaikan dengan baik sebagai salah satu syarat agar dapat menyelesaikan jenjang studi Strata-1 pada Departemen Teknik Informatika Universitas Hasanuddin.

Penulis menyadari secara penuh bahwa pada penulisan skripsi ini terdapat berbagai macam kendala yang dihadapi. Namun, berkat bantuan, bimbingan, dan dukungan dari berbagai pihak sehingga kendal tersebut dapat diatasi. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih kepada :

1. Allah SWT, atas berkat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini,
2. Ibu saya tercinta, ibu Kursiyah, yang sudah senantiasa mendukung, mendoakan, dan membantu memfasilitasi kebutuhan penulis selama menjalani perkuliahan. Terima kasih atas kesabarannya selama ini.
3. Almarhum ayah saya, Dr. Ir. Indra Jaya Mansyur M.T, yang meski telah berpulang kerahmatullah, tapi tetap menjadi inspirasi dan motivasi bagi penulis untuk menyelesaikan studinya. Semoga amal ibadahnya diterima disisi-Nya.
4. Bapak Dr.Eng. Zulkifli Tahir, S.T selaku Dosen Pembimbing 1, dan Bapak Muhammad Alief Fahdal Imran Oemar, ST., M.Sc selaku Dosen Pembimbing 2 yang senantiasa perhatian, dan meluangkan waktu, tenaga, serta pikirannya untuk memberikan masukan dan arahan kepada penulis hingga skripsi ini selesai.
5. Teman – teman Synchronous 2018 atas dukungan, bantuan, serta semangatnya selama ini.
6. Teman seperjuangan penulis di SMAN 5 Makassar, Muti, Idham, Ihsan, Ichsan, Nisten, Winda, Fauzan, Alif, Nabil, Abdillah, Rahma, Nida, Inda,

Fitri, Diva, Azizah, Chassy, dan Ainun yang hingga saat ini senantiasa memberikan dukungan dan menemani penulis.

7. Teman penulis, Fadhiah yang senantiasa menemani dan memberikan dukungan selama penulis menyusun tugas akhir ini.
8. Teman dekat penulis selama kuliah, Wira, Irzan, Maulana, Fiqri, Gusti, Tri Rezki, Arum, Maghfirah, Salahuddin, Ayi, Ayu, Alisyah yang selalu menemani penulis selama masa perkuliahan hingga saat ini.
9. Pihak – pihak yang tidak bisa saya sebutkan namanya satu persatu namun senantiasa membantu penulis baik dalam menyusun laporan, mendukung, maupun menemani penulis untuk mengerjakan sistem.

Penulis berharap semoga Allah SWT berkenan membalas segala kebaikan yang telah diterima oleh penulis dari berbagai pihak yang telah membantu mempermudah penulis dalam mengerjakan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh karena itu penulis mengharapkan segala bentuk saran serta masukan yang membangun dari berbagai pihak. Semoga tugas akhir ini dapat memberikan pengetahuan dan manfaat bagi penulis dan pembaca

Makassar, Juli 2024

Penulis
Yanuarsyah Fitrah Indra

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam menjalani kehidupannya sebagai makhluk sosial, manusia tidak dapat berdiri sendiri dan terpisah dari anggota masyarakat lain. Setiap manusia memerlukan interaksi antar sesamanya, baik antar individu dengan individu, antar individu dengan kelompok, dan antar kelompok dengan kelompok. Pada dasarnya, hubungan interaksi sosial dapat dibagi menjadi interaksi positif dan negatif. Interaksi positif terjadi apabila terdapat hubungan timbal balik yang saling menguntungkan. Sementara interaksi negatif terjadi jika hubungan timbal balik merugikan satu pihak atau keduanya (bermusuhan). Interaksi negatif ini juga kemudian sering disebut sebagai konflik sosial dan merupakan salah satu permasalahan yang kerap muncul pada kehidupan bermasyarakat terutama di negara Indonesia. Adanya keberagaman suku, agama, adat istiadat, dan budaya semakin memberikan peluang besar terjadinya bentrokan sikap. Hal tersebut juga dikonfirmasi oleh para sosiolog yang mengatakan bahwa faktor terjadinya sebuah konflik sosial diantaranya karena adanya perbedaan karakter di setiap individu, perbedaan kebudayaan, hingga perbedaan kepentingan sehingga individu/kelompok saling bersaing untuk memperebutkan hal tersebut (Suhardin, 2020). Konflik sosial ini juga didukung oleh penyebaran informasi yang semakin bebas sehingga potensi gangguan ketentraman seperti informasi hoaks dan adu domba semakin marak terjadi.

Konflik sosial ini pun sering terjadi di Kabupaten Wajo, Sulawesi Selatan. Beberapa kejadian yang sering timbul akibat konflik sosial masyarakat di Kabupaten Wajo yaitu dalam pelaksanaan pemilihan kepala desa, tingkah laku anak punk yang mengganggu ketentraman warga, masalah kesehatan mental yang membuat takut masyarakat, penyalahgunaan perumahan untuk kegiatan prostitusi, konsumsi minuman beralkohol yang dapat menimbulkan ancaman kekerasan, dan konflik sosial yang sering terjadi akibat gesekan ekonomi, politik, sosial dan faktor lain (Sudirman, 2022). Selain itu, terdapat berbagai permasalahan terkait konflik agraria pada Kabupaten Wajo. Berbagai mediasi telah dilakukan, namun konflik

ini masih berlarut-larut dan kriminalisasi terhadap petani juga masih terus berlangsung hingga saat ini (Chandra W., 2020). Terlebih lagi akan diadakannya pemilihan umum (pemilu) pada tahun 2024 mendatang akan semakin memberikan kesempatan besar untuk terjadinya berbagai konflik sosial di masyarakat. Adanya indikasi maupun potensi seperti diatas yang dapat menimbulkan gangguan ketentraman dan ketertiban di masyarakat harus segera dicegah dan diatasi. Pemerintah Kabupaten Wajo telah melakukan beberapa upaya untuk mengatasi permasalahan ini, salah satunya dengan melakukan kegiatan rekonsiliasi melalui upaya-upaya mediasi persuasif dan melibatkan tokoh masyarakat. Pemerintah Kabupaten Wajo juga sedang berinisiatif untuk membangun sebuah sistem informasi berbasis *website* sebagai wadah masyarakat dalam melaporkan konflik sosial yang sedang terjadi di sekitarnya. Berdasarkan data yang dikumpulkan dari sistem informasi tersebut kemudian dapat berguna dan menjadi dasar dalam pengambilan keputusan atau tindakan sehingga dapat meminimalisir konflik sosial yang akan terjadi di Kabupaten Wajo.

Pengolahan data sentimen masyarakat yang dikumpulkan secara *online* tentunya berjumlah sangat banyak dan akan menyulitkan dan memakan waktu yang lama apabila dianalisis secara manual. Oleh karena itu, analisis sentimen hadir sebagai salah satu cara untuk dapat mengklasifikasikan data tersebut. Analisis Sentimen terhadap konflik sosial sudah diteliti oleh beberapa peneliti sebelumnya, seperti yang dilakukan oleh Effendi dkk (Effendi et al., 2022). dengan mengangkat konflik Rusia-Ukraina dimana peneliti melakukan analisis sentimen menggunakan algoritma Naïve Bayes dan *feature selection Particle Swarm Optimization* (PSO). Penelitian ini berhasil mengklasifikasikan sentimen negatif dan positif dengan tingkat *accuracy* sebesar 75,70%, *precision* 75,68%, dan *recall* 75,60%. Ahmad Santoso dkk (Santoso et al., 2022). pada penelitiannya terkait sentimen mobil listrik dengan menggunakan algoritma Support Vector Machine mendapatkan hasil bahwa metode SVM dengan penambahan *feature selection* PSO menambah akurasi analisis dengan nilai sebesar 86.07%, jika dibandingkan dengan penggunaan metode SVM saja tanpa adanya *feature selection* mendapatkan akurasi sebesar 82.51%. Miftahusalam dkk (Miftahusalam et al., 2022) juga melakukan analisis sentimen terhadap penghapusan tenaga honorer dengan membandingkan 3

algoritma, yakni algoritma Support Vector Machine, *Random Forest*, dan Naïve Bayes. peneliti juga membandingkan penggunaan *resampling* untuk menyeimbangkan data, dan didapatkan bahwa penggunaan algoritma *Random Forest* dengan penanganan data *imbalanced* menggunakan *random oversampling* menghasilkan tingkat akurasi lebih tinggi yaitu sebesar 66,67% jika dibandingkan dengan penggunaan metode SVM dan Naïve Bayes.

Berdasarkan penjabaran beberapa jurnal studi diatas, penelitian ini bertujuan untuk menganalisis sentimen terhadap konflik sosial di Kabupaten Wajo dengan menggunakan algoritma *Modified Balanced Random Forest* (MBRF) dengan *feature extraction* TF-IDF dan *feature selection Particle Swarm Optimization* (PSO). Algoritma MBRF merupakan hasil modifikasi dari algoritma *Random Forest* dan *Balanced Random Forest* yang bertujuan untuk meningkatkan akurasi prediksi, mengurangi kompleksitas waktu dan penanganan *imbalanced* yang menjadi masalah utama dalam algoritma klasifikasi *Machine Learning* (Purbolaksono et al., 2021). Pengumpulan data konflik sosial Kabupaten Wajo akan dilakukan dengan dua cara, yaitu berdasarkan pengaduan masyarakat pada sistem informasi yang akan dibangun, dan juga menggunakan metode *crawling* data pada media sosial seperti twitter dan youtube. Diharapkan penelitian ini dapat menjadi salah satu solusi dan dasar oleh pemegang otoritas untuk mengambil keputusan terhadap permasalahan konflik sosial yang ada di Kabupaten Wajo.

1.2 Rumusan Masalah

- a. Bagaimana membangun sistem informasi berbasis *website* mengenai pelaporan konflik sosial Kabupaten Wajo?
- b. Bagaimana membangun model analisis sentimen terhadap konflik sosial di Kabupaten Wajo dengan menggunakan metode *Modified Balanced Random Forest*?
- c. Bagaimana menganalisis dan mengevaluasi performa dalam mengklasifikasikan sentimen terhadap konflik sosial di Kabupaten Wajo menggunakan metode *Modified Balanced Random Forest*?

1.3 Tujuan Penelitian

- a. Untuk membangun sistem informasi berbasis *website* mengenai pelaporan konflik sosial Kabupaten Wajo.
- b. Untuk membangun model analisis sentimen terhadap konflik sosial di Kabupaten Wajo dengan menggunakan metode *Modified Balanced Random Forest*
- c. Menganalisis dan mengevaluasi performa dalam mengklasifikasikan sentimen terhadap konflik sosial di Kabupaten Wajo menggunakan *metode Modified Balanced Random Forest*

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat membantu :

- a. Mahasiswa agar dapat menjadi informasi yang bermanfaat di kemudian hari, serta dapat dijadikan sebagai salah satu landasan dalam melakukan penelitian lainnya yang sejenis.
- b. Masyarakat Kabupaten Wajo agar mendapatkan wadah untuk menampung keresahan mereka mengenai konflik sosial yang terjadi di sekitarnya.
- c. Pemerintah Kabupaten Wajo agar dapat memberikan wadah untuk masyarakat dalam melakukan pelaporan konflik sosial secara mudah dan cepat. Selain itu hasil analisis sentimen juga dapat menjadi acuan dan dasar oleh para pemegang otoritas dalam menentukan prioritas konflik sosial yang harus diselesaikan terlebih dahulu sehingga dapat menambah efisiensi kerja pemerintah dalam memberantas konflik sosial antar masyarakat.

1.5 Ruang Lingkup

Ruang lingkup pada penelitian ini diantaranya:

- a. Data yang digunakan berasal dari *crawling* data melalui media sosial Twitter dan Youtube,
- b. Data yang digunakan hanya dalam lingkup Kabupaten Wajo.
- c. Metode yang digunakan adalah *Modified Balanced Random Forest* (MBRF)
- d. Menggunakan *feature extraction* TF-IDF dan *feature selection* PSO

BAB II TINJAUAN PUSTAKA

2.1 Konflik Sosial

Konflik dapat diartikan sebagai suatu keadaan dimana dua pihak atau lebih saling berbenturan karena kepentingan yang berbeda. Pihak yang dimaksud dapat mengacu pada individu dan kelompok. Konflik kepentingan ini biasanya mengarah pada berbagai tindakan di mana pihak-pihak yang terlibat menggunakan cara-cara kekerasan atau damai untuk mencapai tujuan mereka (Muliono, 2020). Konflik sendiri hadir sebagai manifestasi dari ketegangan sosial, politik, ekonomi dan budaya atau bisa juga disebabkan oleh perasaan ketidakpuasan umum, ketidakpuasan terhadap komunikasi, ketidakpuasan terhadap simbol-simbol sosial dan ketidakpuasan terhadap kemungkinan resolusi serta adanya sumber daya mobilisasi (Irwandi & Chotim, 2017).

Kondisi geografis Indonesia yang sangat luas menjadikan negara ini sangat rawan akan konflik antar masyarakatnya, beberapa kejadian konflik di Indonesia seperti pada daerah Aceh akan berujung pada kekerasan, korban yang berjatuh, dan kericuhan, begitupun pada daerah Wajo, dimana berbagai konflik telah terjadi dan sangat meresahkan warga (Sudirman, 2022)

2.2 Media Sosial

Media sosial mengacu pada situs web dan *platform online* yang memungkinkan pengguna untuk membuat, berbagi konten, dan berinteraksi dengan orang lain (Dhiman, 2023). Dewasa ini media sosial sangat banyak dipakai untuk mengekspresikan diri dan cenderung tanpa batas di dunia maya layaknya di kehidupan nyata (Hasanah et al., 2022). Media sosial juga memungkinkan orang untuk tetap terinformasi dan terhubung dengan teman dan keluarga, memungkinkan bisnis untuk berkomunikasi dengan pelanggan, dan memfasilitasi pertukaran ide, pemikiran, dan informasi (Dhiman, 2023).

2.3 Crawling Data

Crawling dapat diartikan sebagai suatu metode pengunduhan atau pengambilan data yang bermaksud untuk mengumpulkan data dari suatu *database*, misalnya data yang diunduh dari server Twitter atau server media sosial lainnya yang menampung data teks (Normawati & Prayogi, 2021). Dalam penggunaan metode *crawling*, API (*Application Programming Interface*) digunakan untuk memperoleh *dataset* yang lebih besar (Kosasih et al., 2022).

2.4 Text Mining

Text Mining adalah proses analisis data yang melibatkan teks, dengan sumber data yang berasal dari dokumen seperti kumpulan kata-kata (Darwis et al., 2021). Tujuan dari *text mining* adalah menghasilkan informasi dari kumpulan dokumen. Teknik *text mining* mampu memproses, mengelompokkan, dan menganalisis data tidak terstruktur dalam jumlah besar untuk menghasilkan informasi yang dibutuhkan. *Text mining* dapat melakukan pengambilan informasi dan menganalisis perasaan dalam dokumen secara emosional untuk mengidentifikasi apakah suatu pernyataan bersifat positif atau negatif. Objek yang dianalisis dalam *text mining* adalah dokumen yang tidak terstruktur atau semi terstruktur. Teknik ini dapat mengambil informasi secara efektif dari sejumlah dokumen yang dibutuhkan. (Samsir et al., 2021).

Media sosial merupakan salah satu objek yang dapat diterapkan pada *text mining* dengan maksud untuk evaluasi serta penunjang organisasi dalam penyampaian informasi, adapun data yang ditambang pada media sosial dapat berupa sebuah komentar, ulasan, ataupun testimoni dari suatu produk atau objek yang akan dianalisis dengan tujuan untuk dapat mengetahui secara umum opini masyarakat terhadap objek tersebut (Normawati & Prayogi, 2021).

2.5 Analisis Sentimen

Opinion Mining atau analisis sentimen merupakan metode klasifikasi dokumen teks berdasarkan pendapat dan pandangan masyarakat dengan melibatkan pemahaman, ekstraksi, dan pengolahan data tekstual secara otomatis guna untuk

memperoleh sebuah informasi mengenai sentimen yang terdapat dalam sebuah kalimat (Ratino et al., 2020).

Analisis sentimen termasuk dalam salah satu bidang dari *Natural Language Processing* (NLP) dan merupakan proses untuk mengidentifikasi sebuah isu apakah cenderung negatif, positif, maupun netral (Fikri et al., 2020). Ada beberapa teknik pada metode klasifikasi sentimen, seperti *hybrid approach*, *lexicon based*, dan *machine learning* dimana seluruh metode tersebut akan melalui tahapan proses analisis yang mencakup teks pengumpulan data, *preprocessing* data yang akan terbagi menjadi beberapa proses, seperti tokenisasi, *stopword removal*, *stemming*, identifikasi sentimen, dan klasifikasi sentimen (Que et al., 2020). Saat ini bidang penelitian analisis sentimen juga telah menjadi bidang populer dan dianggap bermanfaat dalam banyak aspek. Secara umum, pengaplikasian analisis sentimen meliputi prediksi harga saham, isu politik, kepuasan terhadap produk atau layanan, analisis reputasi, dan lain sebagainya (Fikri et al., 2020).

2.6 Algoritma *Random Forest*

Algoritma *Random Forest* adalah pengembangan dari metode *Classification and Regression Tree* (CART) dengan menerapkan teknik *bootstrap aggregating* (*bagging*) dan pemilihan fitur secara acak (Warsito B. et al., 2020). Metode *Random Forest* merupakan sebuah esamble (kumpulan) metode pembelajaran menggunakan pohon keputusan sebagai *base classifier* yang dibangun dan dikombinasikan (Fitri E. et al., 2020). Setiap pohon keputusan menggunakan prediktor acak dan *random forest* sendiri melakukan prediksi dengan mengkombinasikan hasil dari setiap pohon keputusan dengan cara *majority vote* untuk klasifikasi dan juga rata-rata untuk regresi (Afdhal I, et al., 2022). Ada beberapa kelebihan dari metode *Random Forest* yaitu hasil akurasi yang bagus. Metode ini juga mampu mengatasi *missing value* dan *noise* yang ada pada data, serta algoritma ini cocok untuk mengklasifikasikan data dalam jumlah yang besar (Pahrul, 2023).

2.7 Algoritma *Modified Balanced Random Forest* (MBRF)

Metode *Modified Balanced Random Forest* (MBRF) merupakan sebuah metode peningkatan dari metode sebelumnya, yakni *Random Forest* (RF) dan

Balanced Random Forest (BRF) (Purbolaksono et al., 2021). Metode BRF menangani data yang tidak seimbang ke dalam proses algoritma dimana BRF menerapkan teknik *undersampling* untuk setiap proses pembentukan pohon keputusan dalam algoritma *Random Forest*, itulah mengapa dikenal sebagai *Balanced Random Forest*, karena menggabungkan teknik *sampling* dengan ide *ensemble*. Namun, BRF memiliki beberapa kelemahan dimana proses *random undersampling* akan mengungkap data penting yang terbuang sehingga dapat mempengaruhi hasil klasifikasi. Oleh karena itu, metode *Modified Balanced Random Forest* hadir sebagai solusi dari kekurangan pada metode BRF, metode MBRF tidak hanya dapat meningkatkan akurasi tetapi juga mampu mengurangi kompleksitas waktu. Metode MBRF ini mengubah proses algoritma BRF yang membuang sebagian besar data. Dengan kata lain, *undersampling* acak BRF digantikan dengan teknik pengelompokan. Teknik distribusi data training juga disesuaikan dengan jumlah parameter *Random Forest* yang digunakan. (Agusta & Adiwijaya, 2019).

Metode MBRF dimulai dari pengambilan *dataset* dari *data training* (D) yang akan dipecah sebanyak masukan dari N *tree* pada D_i (D_1, D_2, \dots, D_n) *tree*, yang memiliki X_i dan Y_i , di mana X_i data vektor dan Y_i label sebuah kelas, lalu algoritma *Random Forest* dijalankan yang diberi label (-) dan (+) sehingga membentuk pohon dua baris dan akan dilakukan proses *undersampling* untuk menyeimbangkan penyebaran data dan menghilangkan *imbalanced* data (Agusta & Adiwijaya, 2019). Data kemudian di partisi kedalam satu atau lebih *cluster* yang memiliki karakteristik data yang sama satu sama lain. Data yang memiliki kemiripan didalam *cluster* diminimalkan dan mempertahankan keterwakilan variasi antar data, sehingga terjadi pengurangan data yang cenderung sama di tiap *cluster*. Data yang dihasilkan oleh proses *undersampling* ini kemudian masuk kedalam proses klasifikasi dengan menggunakan algoritma *Random Forest* (Purbolaksono et al., 2021).

Klasifikasi diawali dengan menentukan jumlah k atau jumlah pohon yang akan digunakan, Lalu, dari sejumlah k tadi dilakukan *Bootstrap sampling* untuk membangun pohon prediksi, Selanjutnya dilakukan penentuan kriteria pemisahan *node* dalam pohon prediksi dengan menggunakan *entropy*. *Entropy* bekerja dengan

melakukan perhitungan untuk mencari kesamaan pada *dataset* dan membagi menjadi beberapa kelas, jika kelas yang dihasilkan berisi data yang serupa, maka *entropy* bernilai nol dan jika kelas yang dihasilkan dapat dibagi menjadi 2 maka *entropy* akan menjadi satu. Lalu selanjutnya dengan melakukan klasifikasi menggunakan *majority voting* dari hasil setiap pohon keputusan. Terakhir, dilakukan pengukuran akurasi ketepatan metode klasifikasi (Purbolaksono et al., 2021)

2.8 Term Frequency — Inverse Document Frequency (TF-IDF)

Term weighting adalah suatu proses pembobotan setiap kata agar bisa mengoptimalkan kemampuan analisis sentimen pada proses *text mining* (Gifari et al., 2022). *Term Frequency – Inverse Document Frequency* atau TF-IDF salah satu *term weighting* yang sangat sering digunakan dalam *information retrieval*, dimana pembobotan ini dinilai memiliki akurasi dan *recall* yang cukup tinggi (Yutika et al., 2021). *Term Frequency* mempunyai proporsi kepentingan yang sebanding dengan total kemunculannya dalam sebuah teks atau dokumen (Gifari et al., 2022). $f(t_k d_j)$ mendefinisikan jumlah kemunculan term k pada sebuah dokumen j . Rumus TF dapat didefinisikan dengan (Yutika et al., 2021) :

$$TF(t_k d_j) = f(t_k d_j) \quad (1)$$

Inverse Document Frequency (IDF) ialah metode pembobotan token yang berperan sebagai monitor untuk kemunculan token dalam sebuah himpunan teks (Gifari et al., 2022). Jika sebuah *term* muncul secara konsisten di banyak dokumen, maka nilai IDF-nya akan semakin kecil. Rumus untuk menghitung IDF adalah sebagai berikut (Yutika et al., 2021):

$$IDF(t_k) = \log \frac{D}{df(t)} \quad (2)$$

Cara kerja TF-IDF adalah dengan mencari kata yang unik dan relevan dalam suatu komentar. Pengukuran TF-IDF akan mencatat seberapa sering kata tertentu

muncul dalam suatu komentar dan dalam keseluruhan *dataset* komentar yang digunakan (Baskoro et al., 2021). TF-IDF dapat dirumuskan sebagai berikut (Yutika et al., 2021) :

$$TF\ IDF (t_k d_j) = TF (t_k d_j) * IDF (t_k) \quad (3)$$

2.9 Selection Feature Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) atau biasa dikenal dengan PSO merupakan sebuah teknik optimasi yang sangat simpel dalam memodifikasi sejumlah parameter. Pengoptimasian PSO dapat dilakukan dengan menyeleksi atribut atau pemilihan fitur dan juga peningkatan bobot atribut pada semua variabel atau atribut yang digunakan. PSO memiliki keunggulan berupa kemudahan dalam penerapannya dan sifatnya yang simpel, serta mampu mencapai kecepatan konvergensi yang baik (Que et al., 2020). Berikut merupakan langkah-langkah dalam melakukan metode PSO (Darmawan et al., 2022) :

1. Inisialisasi

- a. Inisialisasi kecepatan awal dimana pada iterasi ke-0, sudah dipastikan bahwa nilai kecepatan awal semua partikel adalah 0.
- b. Inisialisasi posisi awal dimana pada iterasi ke-0, posisi awal partikel dibangkitkan dengan menerapkan persamaan berikut :

$$x = x_{min} + rand[0.1] \times (x_{max} - x_{min}) \quad (4)$$

- c. Inisialisasi lBest (*local best*) dan gBest (*global best*) dimana pada iterasi ke-0, lBest akan disamakan dengan nilai posisi awal target. Sedangkan gBest dipilih dari satu lBest dengan fitness.

2. Perbaharui Kecepatan

Untuk meng-update kecepatan, digunakan rumus sebagai berikut :

$$v_{i,j}^{t+1} = w \cdot v_{i,j}^t + c_1 \cdot r_1 (Pbest_{i,j}^t - x_{i,j}^t) + w \cdot v_{i,j}^t + c_2 \cdot r_2 (Gbest_{g,j}^t - x_{i,j}^t) \quad (5)$$

dimana :

$v_{i,j}$ = komponen kecepatan individu ke i pada d dimensi

w = parameter inertia *weight*

c_1, c_2 = konstanta akselerasi (*learning rate*), nilainya antara 0 sampai 1

r_1, r_2 = parameter *random* antara 0 sampai 1

$Pbest_{i,j}$ = Pbest (*local best*) individu i pada j dimensi

$Gbest_{g,j}$ = Gbest (*global best*) individu g pada j dimensi

3. Perbaharui posisi dan hitung *fitness*

Untuk meng-update posisi, digunakan rumus sebagai berikut :

$$x_{i,j}^{t+1} = v_{i,j}^t + x_{i,j}^{t+1} \quad (6)$$

dimana :

$x_{i,j}$ = posisi individu i pada j dimensi

4. Update IBest (*local best*) dan gBest (*global Best*)

Dilakukan perbandingan antara IBest (*local best*) pada iterasi sebelumnya dengan hasil dari update posisi. *Fitness* yang lebih tinggi akan menjadi IBest yang baru. IBest terbaru yang memiliki nilai *fitness* tertinggi akan menjadi gBest yang baru.

2.10 Confusion Matrix

Confusion Matrix atau biasa juga disebut *Error Matrix* merupakan metode yang digunakan untuk melakukan perhitungan akurasi dalam proses klasifikasi atau *Supervised Learning* (Purbolaksono et al., 2021). Contoh *Confusion Matrix* untuk klasifikasi biner ditunjukkan pada Tabel 1 (Normawati & Prayogi, 2021) :

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		1	0
Kelas sebenarnya	1	TP	FN
	0	FP	TN

Keterangan :

TP (<i>True Positive</i>)	= jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1
TN (<i>True Negative</i>)	= jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0
FP (<i>False Negative</i>)	= jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 0
FN (<i>False Negative</i>)	= jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 0

Rumus *Confusion Matrix* untuk menghitung *accuracy*, *precision*, dan *recall* seperti berikut (Wardani et al., 2022) :

$$accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (7)$$

$$recall = \frac{TP}{TP+FN} \quad (8)$$

$$precision = \frac{TP}{TP+FP} \quad (9)$$