IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI KATEGORI BERITA ONLINE

SKRIPSI

Oleh:
AULIA ISTIANI
NIM. 19650041

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024

IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI KATEGORI BERITA ONLINE

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
AULIA ISTIANI
NIM. 19650041

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024

HALAMAN PERSETUJUAN

IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI KATEGORI BERITA ONLINE

SKRIPSI

Oleh :
AULIA ISTIANI
NIM. 19650041

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 07 Juni 2024

Pembimbing I,	Pembimbing II,
Dr. M. Amin Hariyadi, M.T.	Roro Inda Melani, M.T., M.Sc.
NIP. 19670018 200501 1 001	NIP. 19780925 200501 2 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM.
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI KATEGORI BERITA ONLINE

SKRIPSI

Oleh :
AULIA ISTIANI
NIM. 19650041

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 19 Juni 2024

Susunan Dewan Penguji

Ketua Penguji	: Dr. Ririen Kusumawati, S.Si., M.Kom. NIP. 19720309 200501 2 002	( )
Anggota Penguji I	: Tri Mukti Lestari, M.Kom NIP. 19911108 202012 2 005	( )
Anggota Penguji II	: Dr. M. Amin Hariyadi, M.T. NIP. 19670018 200501 1 001	( )
Anggota Penguji III	: Roro Inda Melani, M.T., M.Sc. NIP. 19780925 200501 2 008	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM.
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Aulia Istiani
NIM	: 19650041
Fakultas	: Sains dan Teknologi
Program Studi	: Teknik Informatika
Judul Skripsi	: Implementasi Metode Random Forest Untuk Klasifikasi Kategori Berita Online

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Juni 2024
Yang membuat pernyataan,

Aulia Istiani
NIM.19650041

MOTTO

"Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain."

(QS. Al-Insyirah: 6-7)

“Mulai aja dulu”

HALAMAN PERSEMBAHAN

Puji syukur kehadirat Allah SWT yang telah memberikan rahmatnya sehingga penulis bisa menyelesaikan skripsi ini dengan baik. Semoga Allah SWT senantiasa meridhoi.

Sholawat serta salam semoga tetap tercurah kepada junjungan kita, Rosulullah Muhammad SAW. Dengan segenap hati, penulis mempersembahkan hasil karya ini kepada:

Ibu, Ayah, Kakak, dan Adik tercinta yang telah menjadi motivasi terbesar bagi penulis, yang selalu mendukung, mendoakan dan memberikan limpahan kasih sayang yang tak terhingga.

Seluruh teman penulis yang telah memberikan energi positif serta dukungan dalam menyelesaikan penyusunan skripsi ini.

Diri sendiri yang telah berusaha dan berjuang sampai sejauh ini.

KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT, karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi ini dengan judul "Implementasi Metode Random Forest Untuk Klasifikasi Kategori Berita Online". Skripsi ini disusun untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Dalam proses penyusunan skripsi ini, penulis banyak mendapatkan bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. H.M. Zainuddin, MA selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Hariani, M.Si selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT. IPM selaku Ketua program studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. M. Amin Hariyadi, M.T. selaku dosen pembimbing 1 yang telah memberikan dukungan dan bimbingan dalam penyusunan skripsi ini.
5. Ibu Roro Inda Melani, M.T., M.Sc. selaku dosen pembimbing 2 dan dosen wali yang senantiasa memberikan arahan, bimbingan, serta motivasi selama pendidikan.
6. Dr. Ririen Kusumawati, S.Si., M.Kom. selaku dosen penguji 1 dan Ibu Tri Mukti Lestari, M.Kom selaku dosen penguji 2 yang telah memberikan arahan dalam menyelesaikan skripsi ini.

7. Seluruh Dosen Teknik Informatika yang telah memberikan ilmu dan pengetahuan selama masa perkuliahan.
8. Ibu, Ayah, Kakak dan Adik tercinta yang telah memberikan dukungan, doa, dan motivasi sehingga penulis dapat menelesaikan penyusunan skripsi ini.
9. Teman-teman angkatan 2019 Program Studi Teknik Informatika, yang selalu memberikan semangat, bantuan, dan kebersamaan selama masa studi.
10. Seluruh pihak yang telah terlibat secara langsung maupun tidak langsung dalam membantu menyelesaikan skripsi ini.
11. Penulis sendiri yang telah berusaha dan bertanggung-jawab untuk menyelesaian pilihan yang telah dimulai.

Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang berkepentingan dan dapat menambah wawasan serta pengetahuan di bidang ilmu teknik informatika.

Malang, 31 Mei 2024

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Istiani, Aulia. 2024. **Implementasi Metode Random Forest Untuk Klasifikasi Kategori Berita Online.** Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M.Amin Hariyadi, M.T. (II) Roro Inda Melani, M.T., M.Sc.

.

Kata kunci: *Random Forest,* Klasifikasi teks, Berita

Penelitian ini dilakukan untuk membantu editor berita online yang masih mengklasifikasikan kategori berita online secara manual. Pengklasifikasian kategori berita online secara manual tidak efektif karena membutuhkanwaktu lama untuk membaca keseluruhan berita dan memungkinkan terjadinya kesalahan karena kemiripan isi dari beberapa berita. Metode yang digunakan dalam penelitian ini adalah metode *Random Forest* dengan pembobotan TF-IDF. Data sebanyak 1115 artikel berita terlebih dahulu dilakukan proses *preprocessing* untuk mengubah data mentah menjadi data yang bisa diproses oleh sistem meliputi *case folding, stopword removal, stemming* dan *tokenizing*. Seluruh kata dari proses *tokenizing* dilakukan proses TF-IDF untuk mengubah kata menjadi bentuk numerik dan dijadikan inputan pada sistem. Hasil uji coba dengan tiga skenario berbeda menunjukkan bahwa skenario pertama dengan rasio 90:10 menghasilkan performa terbaik dengan akurasi 79%., skenario kedua dengan rasio 80:20 menghasilkan akurasi 66% dan skenario ketiga dengan rasio 70:30 menghasilkan akurasi 77%.

ABSTRACT

Istiani, Aulia. 2024. **The Implementation of Random Forest Method for Online News Category Classification**. Thesis. Informatics Engineering. Faculty of Science and Technology. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Dr. M.Amin Hariyadi, M.T. (II) Roro Inda Melani, M.T., M.Sc.

Keywords: *Random Forest, Text Classification, News*

The research aims to help online news editors who still classify online news manually. The manual classification is ineffective since it takes a longer time to read the whole news and may miss mistakes due to news similarity. The researcher employed the Random Forest method with TF-IDF weighing. She pre-processed data from 1,115 news articles to change the raw materials into data the system can process, including case folding, stopword removal, stemming, and tokenizing. All words from the tokenizing process went through the TF-IDF process to change them into numeric to be inputted into the system. The trial result using three different scenarios shows that the first scenario with a ratio of 90:10 induces the best performance with an accuracy of 79%. The second scenario, with a ratio of 80:20, induces 66% accuracy. Meanwhile, the third scenario with a ratio of 70:30 induces 77% accuracy.

مستخلص البحث

إستياني، أوليا. 2024. تنفيذ طريقة الغابة العشوائية لتصنيف فئات الأخبار الإلكترونية. البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. محمد أمين هريادي، الماجستير. المشرف الثاني: رورو إنداه ميلاني، الماجستيرة.

الكلمات الرئيسية: غابة عشوائية، تصنيف النص، أخبار.

تم إجراء هذا البحث لمساعدة محرري الأخبار الإلكترونية الذين ما زالوا يصنفون فئات الأخبار الإلكترونية يدويا. التصنيف اليدوي لفئات الأخبار الإلكترونية غير فعال لأنه يستغرق وقتا طويلا لقراءة الأخبار بأكملها ويسمح بالأخطاء بسبب تشابه محتوى بعض الأخبار. الطريقة المستخدمة في هذا البحث هي طريقة الغابة العشوائية مع ترجيح TF-IDF. تمت معالجة بيانات 1115 مقالة إخبارية أولا لتحويل البيانات الأولية إلى بيانات يمكن معالجتها بواسطة النظام، بما في ذلك طي الحالة وإزالة الكلمات الموقوفة والتوقف والترميز. تم تنفيذ جميع الكلمات من عملية الترميز بواسطة عملية TF-IDF لتحويل الكلمات إلى أشكال رقمية واستخدامها كمدخلات للنظام. أظهرت نتائج اختبار بثلاثة سيناريوهات مختلفة أن السيناريو الأول بنسبة 90:10 نتج عنه أفضل أداء بدقة 79٪، أما السيناريو الثاني بنسبة 80:20 فقد نتج عنه دقة 66٪ ونتج السيناريو الثالث بنسبة 70:30 دقة 77٪.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi saat ini telah mendorong perkembangan media massa terutama media online sehingga informasi berupa berita dapat cepat tersampaikan. Di Indonesia sendiri media online menjadi salah satu media yang banyak digunakan masyarakat karena harganya yang relatif murah dan informasi yang dibutuhkan lebih cepat didapatkan. Saat ini banyak media cetak yang beralih menggunakan media online seperti Detik.com, Tempointeraktif.com, dan berbagai portal berita local seperti radar malang maupun radar tulungagung. Dengan munculnya berbagai portal berita online berakibat banyaknya berita yang tersebar secara cepat dan aktual dengan selisih waktu permenit maupun perdetik sehingga ketersediaan jumlah dokumen berita semakin tidak terkendali.

Pada umumnya, portal berita memiliki beberapa kategori berita untuk memudahkan masyarakat dalam mencari informasi sesuai kebutuhan. Kategori tersebut bisa berupa olahraga, hiburan, politik, ekonomi hingga berita kesehatan. Dengan banyaknya kategori dan jumlah dokumen berita menyulitkan editor untuk mengelompokkan berita sesuai kategori yang ada, jika masih dilakukan secara manual. Pengelompokkan secara manual membutuhkan waktu yang lama dan memungkinkan terjadinya kesalahan karena editor harus membaca keseluruhan isi berita. Permasalahana lain yang biasa muncul ketika terdapat kemiripan isi dari dokumen berita yang akan diunggah, misalnya suatu berita membahas mengenai teknologi yang berhubungan dengan kesehatan, maka berita tersebut bisa

diklasifikasikan sebagai kategori teknologi atau kesehatan. Oleh karena itu dibutuhkan proses klasifikasi artikel berita secara otomatis.

Mengenai penyebaran berita secara cepat, Ummul Mukminin ‘Aisyah r.a. pernah dituduh dengan tuduhan palsu oleh orang munafik di mana tuduhan tersebut dengan cepat menyebar di kalangan kaum muslim.

Aisyah r.a. berkata *“Sesampainya di kota Madinah, saya jatuh sakit selama satu bulan, sementara orang-orang masih marak menangapi isu yang disebarluaskan oleh para pembuat berita bohong sedangkan saya sendiri tidak merasa berbuat apa-apa”.* (HR. Muslim)

Berdasarkan penafsiran Ibnu Katsir, hadist tersebut berkenaan tentang tuduhan orang munafik kepada Aisyah r.a. mengenai berita perzinaan yang disebarkan oleh Abdullah bin Ubay bin Salul, pemimpin orang-orang munafik. Hadist tersebut memberikan pelajaran bahwa dalam menerima suatu kabar atau berita, sebaiknya dipastikan dulu kebenaran dari berita tersebut dengan tabayyun atau mencari kejelasan atau kebenaran dari seorang ahli atau sumber terpercaya. Dalam konteks penelitian ini, sumber terpercaya bisa berupa situs berita online yang mana dalam situs berita online, berita telah diklasifikasikan oleh sistem. Sistem memungkinkan pengklasifikasian berita secara tepat karena sistem mempelajari pola berita yang diberikan.

Dalam firman Allah surah Al-Maidah ayat 2 yakni:

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوْا لَا تُحِلُّوْا شَعَاۤىِٕرَ اللّٰهِ وَلَا الشَّهْرَ الْحَرَامَ وَلَا الْهَدْيَ وَلَا الْقَلَاۤىِٕدَ وَلَآ آۤمِّيْنَ الْبَيْتَ الْحَرَامَ يَبْتَغُوْنَ فَضْلًا مِّنْ رَّبِّهِمْ وَرِضْوَانًا ۗوَاِذَا حَلَلْتُمْ فَاصْطَادُوْا ۗوَلَا يَجْرِمَنَّكُمْ شَنَاٰنُ قَوْمٍ اَنْ صَدُّوْكُمْ عَنِ الْمَسْجِدِ الْحَرَامِ اَنْ تَعْتَدُوْا ۘوَتَعَاوَنُوْا عَلَى الْبِرِّ وَالتَّقْوٰى ۖوَلَا تَعَاوَنُوْا عَلَى الْاِثْمِ وَالْعُدْوَانِ ۠وَاتَّقُوا اللّٰهَ ۗاِنَّ اللّٰهَ شَدِيْدُ الْعِقَابِ

“Wahai orang-orang yang beriman! Janganlah kamu melanggar syiar-syiar kesucian Allah, dan jangan (melanggar kehormatan) bulan-bulan haram, jangan (mengganggu) hadyu (hewan-hewan kurban) dan qala'id (hewan-hewan kurban

yang diberi tanda), dan jangan (pula) mengganggu orang-orang yang mengunjungi Baitulharam; mereka mencari karunia dan keridaan Tuhannya. Tetapi apabila kamu telah menyelesaikan ihram, maka bolehlah kamu berburu. Jangan sampai kebencian(mu) kepada suatu kaum karena mereka menghalang-halangimu dari Masjidilharam, mendorongmu berbuat melampaui batas (kepada mereka). Dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah, sungguh, Allah sangat berat siksaan-Nya." (Q.S. Al-Maidah: 2)

Potongan ayat وَتَعَاوَنُوْا عَلَى الْبِرِّ وَالتَّقْوٰى berdasarkan penafsiran Ibnu Katsir dalam Rulli Hastuti (2022) bermakna bahwa seorang hamba diperintahkan oleh Allah untuk selalu tolong-menolong dalam suatu kebaikan. Maka apabila dalam kehidupan bermasyarakat terdapat seseorang atau sekelompok orang yang mengalami kesulitan dan membutuhkan pertolongan alangkah baiknya untuk segera ditolong. Dalam konteks ini, membuat sistem klasifikasi berita online secara otomatis merupakan perbuatan baik karena dapat mempermudah editor untuk mengelompokkan kategori berita secara cepat dan tepat sehingga informasi yang disampaikan kepada masyarakat menjadi lebih terstruktur dan mudah diakses.

Klasifikasi merupakan proses menemukan model yang diperoleh dari analisis training data yang digunakan untuk menentukan atribut atau kelas data yang relevan (Yunial, 2020). Proses klasifikasi berita secara otomatis digunakan untuk menentukan isi berita terkait apakah tergolong ke dalam kategori olahraga, politik, pendidikan atau yang lainnya. Sebelum berita diklasifikasikan, terlebih dahulu berita harus diproses dengan membandingkan pada data yang lain. Berita tergolong sebagai data tidak terstruktur dan sulit ditangani sehingga dibutuhkan text mining agar data teks dapat diolah dan menghasilkan analisis yang bermanfaat (Ariadi & Fithriasari, 2015). Text mining melakukan identifikasi pola terhadap sekumpulan

teks berskala besar untuk mendapatkan informasi dari data tersebut (Feldman & Sanger dalam Morama *et al.*, 2022). Proses penemuan pola dapat dilakukan dengan berbagai metode, salah satunya menggunakan metode Random Forest.

Random Forest merupakan salah satu algoritma dengan tingkat akurasi yang baik (Baskoro *et al.*, 2021). Metode Random Forest merupakan metode yang biasa digunakan pada proses klasifikasi dan regresi yang mana didasarkan pada konsep *ensemble learning* yakni menggabungkan beberapa pengklasifikasian untuk meningkatkan kinerja model. Konsep dari metode *Random Forest* yakni menggabungkan beberapa pohon keputusan (*Decision Tree*) yang mana setiap tree diberikan data secara acak menggunakan metode *bagging* atau *Boostrap Aggregating* dari dataset training (Rohman *et al.*, 2018). Keputusan akhir diperoleh dari major voting pada pohon keputusan yang telah dibuat. Menurut Pamuji & Ramadhan (2021) *Random Forest* memiliki beberapa kelebihan yakni memiliki performa baik dalam melakukan klasifikasi, menghasilkan tingkat eror yang relatif rendah, bekerja secara efisien dalam data dengan jumlah besar, dan efektif dalam menangani *missing* data. Metode *Random Forest* dapat meningkatkan akurasi atau performa terhadap model klasfikasi dengan mengambil fitur terbaik dalam proses seleksi fitur yang dilakukan (Supriyadi *et al.*, 2020).

Penelitian klasifikasi teks menggunakan Random Forest dilakukan oleh Basar *et al.* (2022) untuk menganalisis sentiment pengguna twitter terhadap pembayaran *cashless*. Akurasi yang didapatkan cukup tinggi yakni 95% dengan kedalaman tree 55 dan jumlah tree 300. Penelitian lain dilakukan oleh Morama *et al.* (2022) mengenai analisis sentiment terhadap ulasan hotel. Uji coba sistem

dilakukan dengan tiga skenario berdasarkan parameter jumlah tree dan kedalaman tree yang berbeda. Hasil yang didapat, sistem memperoleh akurasi sebesar 90% pada kedalaman tree sebanyak 10 dan jumlah tree 300. Morama *et al.* (2022) menyimpulkan bahwa jumlah tree dan kedalaman tree berperan penting dalam meningkatkan akurasi prediksi.

Berdasarkan permasalahan yang ada, maka peneliti melakukan penelitian untuk mengklasifikasikan kategori berita online menggunakan metode *Random Forest*. Metode *Random Forest* dinilai cocok untuk melakukan klasififkasi data dengan jumlah sampel yang banyak. Diharapkan pengklasifikasian ini dapat digunakan portal atau situs berita online dalam menentukan kategori berita secara cepat dan efisien.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka masalah yang diangkat dalam penelitian ini yakni bagaimana nilai *accuracy, precision, recall* dan *f-measure* dalam klasifikasi kategori berita online menggunakan metode *Random Forest*?

1.3 Batasan Masalah

1. Data diambil dari situs berita online yakni detik.com dan kompas.com.
2. Berita diklasifikasikan berdasarkan kategori edukasi, keuangan, kesehatan, olahraga dan teknologi.

1.4 Tujuan Penelitian

Penelitian ini dilakukan untuk mengetahui nilai *accuracy, precision, recall dan f-measure* dalam klasifikasi kategori artikel berita online menggunakan metode *Random Forest*

1.5 Manfaat Penelitian

Manfaat yang didapat dalam penelitian ini:

1. Membantu editor berita online untuk mengklasifikasikan kategori berita secara mudah dan cepat
2. Membantu masyarakat dalam mencari berita online sesuai kategori yang diinginkan

BAB II

STUDI PUSTAKA

Pada bab ini peneliti membahas mengenai beberapa penelitian sebelumnya yang mana dilakukan sebagai bahan perbandingan dan kajian, serta membahas mengenai teori-teori dasar yang berhubungan dengan penelitian klasifikasi kategori artikel berita online menggunakan metode Random Forest.

2.1 Pengertian Terkait dengan Penelitian

Larasati *et al.* (2022) melakukan implementasi metode *Random Forest* dalam menganalisis sentiment ulasan aplikasi dana. Data yang digunakan sebanyak 1354 diambil menggunakan teknik scraping dengan hasil klasifikasi kelas positif sebanyak 577 ulasan, negatif 279 ulasan dan netral sebanyak 509 ulasan. Hasil yang diperoleh pada penelitian ini mendapat nilai akurasi 84%, *precision* 84%, *recall* 84% dan *f1-Score* 84% dengan perbandingan data latih dan data uji sebanyak 80% : 20% dan parameter kedalaman tree 65 dan jumlah tree 400.

Penelitian menggunakan metode Random Forest dilakukan oleh *Baskoro et al.* (2021) mengenai analisis sentiment ulasan pelanggan hotel. Data sebanyak 1166 komentar diperoleh dari web tripadvisor.co.id menggunakan teknik crawling data dengan kata kunci "hotel purwokerto". Kelas pada dataset dibagi ke dalam kelas positif dan negatif berdasarkan nilai rating pelayanan sehingga pelabelan kelas dilakukan secara otomatis menggunakan fungsi if else. Pengujian dilakukan dengan membagi data testing dan data uji sebesar 80%:20% dan menggunakan tiga skenario *preprocessing* yang berbeda. Hasil yang didapat bahwa skenario tiga memperoleh akurasi terbaik sebesar 87,23% yang didapat dengan cara mengoreksi

komentar negative pada rating 30 secara manual serta ditambahkan 3628 kata pada tahap *stopword*.

Morama *et al.* (2022) Melakukan penelitian menggunakan metode *Random Forest* untuk menganalisis ulasan hotel berbasis aspek. Data yang digunakan sebanyak 1428 ulasan dari hasil scrapping web pada situs TripAdvisor dengan pembagian kelas positif, negatif dan netral. Pengujian dilakukan dengan menggunakan tiga skenario berdasarkan parameter jumlah tree dan kedalaman tree yang bertujuan untuk mengetahui apakah parameter tersebut berpengaruh pada hasil prediksi. Hasil yang didapat menunjukkan bahwa semakin banyak jumlah tree dan kedalaman tree maka semakin baik hasil prediksi yang diperoleh. Skenario tiga dengan menggunakan jumlah tree 300 memperoleh akurasi sebesar 90%, presisi 91%, recall 90% dan f1-score 90%, sedangkan kedalaman tree sebesar 10 mendapat akurasi 90%, presisi 92%, recall 90% dan f1-score 90%.

Penelitian lain dilakukan oleh Basar *et al.* (2022) mengenai analisis sentiment pengguna twitter. Penelitian ini menggunakan algoritma *Random Forest*. Data yang digunakan dibagi ke dalam kelas positif, negative dan netral yang diambil menggunakan API twitter mengenai opini pengguna shopeepay terhadap pembayaran *cashless*. Hasil yang didapatkan diketahui bahwa sistem dapat menganalisis data dengan baik saat jumlah tree 300 dan kedalaman tree 55 dan menghasilkan nilai akurasi 95%, precision 95%, recall 94% dan F1-score 95%.

Penelitian dengan topik terkait telah dilakukan oleh Munirul *et al.* (2020) yang membahas tentang analisis dan deteksi konten hoax menggunakan perbandingan Metode *Multilayer Perceptron*, *Naïve Bayes*, *Support Vector*

Machine dan *Random Forest*. Penelitian tersebut digunakan untuk mengetahui algoritma terbaik yang bisa digunakan untuk membedakan berita hoax dan berita asli. Data yang digunakan diambil dari situs berita online sebanyak 50 artikel hoax dan 100 artikel non-hoax yang kemudian dibagi menjadi 80% data *training* dan 20% data *testing*. Hasil yang diperoleh menunjukkan bahwa metode *Random Forest* mendapatkan tingkat akurasi terbaik yakni 75.37%. Sedangkan *Multilayer Perceptron* mendapat tingkat akurasi 68.63%, *Naïve Bayes* mendapat tingkat akurasi 74.51% dan *Support Vector Machine* mendapat tingkat akurasi 60.00%.

Ramadhan *et al.* (2022) melakukan penelitian dengan membandingkan metode *Random Forest* dan *logistic regression* untuk mendeteksi berita palsu. Data yang digunakan diperoleh dari kaggle sebanyak 6560 data yang terbagi dalam dua kelas yakni *fake* dan *real*. Pengujian dilakukan dengan membagi data *training* 70% dan data *testing* 30% yang mana menghasilkan nilai akurasi tertinggi 84% dengan metode *Random Forest* (*entropy*), 83% menggunakan metode *Random Forest* (*gini*) dan 77% menggunakan *Logistic Regression*.

Penelitian mengenai klasifikasi berita dilakukan oleh Habib *et al.* (2022) menggunakan metode *Naïve Bayes*. Data yang digunakan sebanyak 510 data yang dikategorikan ke dalam kelas demokrasi, kemiskinan dan ketenagakerjaan. Penelitian ini bertujuan untuk membantu BPS Riau dalam mengklasifikasikan jenis berita sesuai fenomena yang ada di lingkungan daerah Riau. Skenario penelitian dilakukan dengan tiga perbandingan data latih dan data uji yakni 70:30, 80:20 dan 90:10. Hasil yang didapatkan, skenario ke-tiga mendapatkan akurasi tertinggi

sebesar 94% dengan perbandingan data latih 90% sebanyak 153 data dan data uji 10% sebanyak 17 data.

Penelitian lain dilakukan oleh Ridwan *et al.* (2021) mengenai klasifikasi kalimat berita olahraga secara otomatis. Metode yang digunakan adalah *Artificial Neural Network* (ANN) dengan backpropagation sebagai metode pembelajaran dan penambahan teknik pembobotan kata TF-IDF pada *preprocessing*. Tujuan penelitian ini dilakukan untuk membangun sistem klasifikasi kalimat berita olahraga secara otomatis berdasarkan isi pembahasan berita guna membantu pembaca mendapatkan bacaan secara ringkas. Data yang digunakan sebanyak 1000 kalimat berita olahraga dari situs bolasport.com yang mana diambil 800 kalimat sebagai data latih dan 200 kalimat sebagai data testing untuk diklasifikan kedalam 10 kelas berbeda. Percobaan pada penelitian ini dilakukan dengan 4 *neuron hidden* yang berbeda yakni 2, 4, 6, 8 *neuron hidden*. Hasil penelitian yang dilakukan mendapatkan akurasi optimal dari percobaan yang menggunakan *hidden layer* berjumlah 6 buah *neuron* dengan tingkat akurasi 99% untuk data latih dan 57% untuk data uji.

Penelitian tentang klasifikasi berita online juga dilakukan oleh Putra *et al*. (2023) dengan menggunakan metode *Support Vector Machine* dan seleksi fitur *chi square*. Data yang digunakan diperoleh dari situs kompas.com pada tanggal 22 Februari 2023 sampai 7 April 2023 dengan jumlah 2400 judul berita yang kemudian dibagi dengan perbandingan 70% data latih dan 30% data uji. Penelitian ini bertujuan untuk mengelompokkan judul berita secara otomatis kedalam 6 kategori berita yakni olahraga, kesehatan, berita nasional, kuliner, keuangan, dan teknologi.

Hasil dari penelitian ini memperoleh akurasi sebesar 93.06%, presisi 92.11%, recall 93.06% dan fI-score sebesar 93.04%.

Nathaniel Chandra *et al.* (2019) melakukan penelitian menggunaka metode *Naïve Bayes* dengan fitur N-gram mengenai klasifikasi berita lokal radar Malang. Data yang digunakan sebanyak 471 berita yang diambil dari kompas.com sebagai data latih dan 293 berita sebagai data uji yang diambil dari radarmalang.co.id. Penelitian dilakukan dengan 5 kali uji coba pengambilan secara acak dari data latih dan data uji. Akurasi tertinggi didapatkan pada percobaan pertama sebesar 78,66% dengan pengambilan data latih sebanyak 61,65% dan sisanya 38,35% dijadikan data uji.

Penelitian lain dilakukan oleh Briliansyah (2020) mengenai pembuatan sistem klasifikasi berita menggunakan metode *K-Nearest Neighbor*. Data yang digunakan sebanyak 400 data yang diambil dari website detik.com dan kompas.com dengan 4 kategori berupa olahraga, teknologi, ekonomi, dan politik. Uji coba dilakukan pada data sebanyak 20 berita mendapat akurasi sebesar 74%.

Huda (2020) melakukan penelitian menggunakan metode *Learning Vector Quantization* untuk mengklasifikasikan berita. Data yang dikumpulkan dari website kompas.com, detik.com, halodoc.com, cnnindonesia.com dan kumparan.com sebanyak 300 data berita dengan 3 kategori berupa politik, kesehatan dan edukasi. Pengujian dilakukan dengan tiga skenario berbeda yang menghasilkan akurasi terbaik sebesar 96,67% pada skenario ketiga dengan perbandingan data latih sebesar 90% dan data uji sebesar 10%.

Penelitian lain mengenai klasifikasi berita dilakukan oleh Hasanah (2022) menggunakan metode *Artificial Neural Network* (ANN). Data yang digunakan sebesar 360 data berita yang diambil dari website detik.com, kompas.com, cnnindonesia.com, sindonews.com, dan republic.com. Kategori yang digunakan berupa politik, teknologi, travel, hiburan, olahraga, pendidikan, ekonomi, hukum dan kesehatan. Pengujian dilakukan dengan 3 skenario berbeda dan menghasilkan akurasi terbaik sebesar 96,29% pada skenario pertama dengan perbandingan 90:10.

Tabel 2.1 Pengertian terkait dengan penelitian

No.	Peneliti	Judul	Metode	Hasil	Batasan Penelitian
1.	(Larasati et al., 2022)	Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest	Random Forest	Memperoleh akurasi 84%	— 1354 data ulasan aplikasi dana dari google play store — Kelas positif, negatif dan netral
2.	(Baskoro et al., 2021)	Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)	Random Forest	Menghasilkan akurasi 87,23%	— 1166 data ulasan hotel dari website tripadvisor.co.id — Kelas positif dan kelas negative
3.	Morama *et al.* (2022)	Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier	Random Forest	Menghasilkan akurasi 90%	— 1166 data ulasan hotel dari web tripadvisor.co.id — Kelas positif, negatif dan netral
4.	Basar *et al*. (2022)	Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest	Random Forest	Menghasilkan akurasi 95%	— Data opini pengguna shopeepay dari twitter — Kelas positif, negatif dan netral

Lanjutan pengertian terkait dengan penelitian

No.	Peneliti	Judul	Metode	Hasil	Batasan Penelitian
5.	Munirul *et al.* (2020)	Analisa Dan Deteksi Konten Hoax Pada Media Berita Indonesia Menggunakan Machine Learning	Multilayer Perceptron, Naïve Bayes, Support Vector Machine, Random Forest	Mendapat akurasi terbaik pada metode Random Forest sebesar 75,37%	— 150 artikel berita dari situs berita online — Kelas hoax dan non-hoax
6.	Ramadhan *et al.* (2022)	Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression	Random Forest dan logistic regression	Mendapat akurasi 84% pada metode Random Forest menggunakan *entropy*	— 6560 data berita dari kaggle — Kelas fake dan real
7.	Habib *et al.* (2022)	Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier	Naïve Bayes	Mendapat akurasi tertinggi sebesar 94%	— 510 data berita dari portal berita daerah Riau — Kelas demokrasi, kemiskinan, dan ketenagakerjaan
8.	(Ridwan et al., 2021)	Klasifikasi Kalimat Pada Berita Olahraga Secara Otomatis Menggunakan Metode Artificial Neural Network	Artificial Neural Network (ANN)	Mendapat akurasi sebesar 99% untuk data pengujian data latih dan 57% pengujian data uji	— 1000 data berita olahraga dari bolasport.com — Kelas manajemen, cedera, jadwal, preview, review, klasemen, statistik, juara, pemain/atlet, dan kelas lainnya
9.	(Putra et al., 2023)	Klasifikasi Judul Berita Online menggunakan Metode Support Vector Machine (SVM) dengan Seleksi Fitur Chi-square	Support Vector Machine	Akurasi terbaik sebesar 93,06%	— 2400 data judul berita dari kompas.com — Kelas kelas olahraga, kesehatan, berita nasional, kuliner, keuangan, dan teknologi
10.	(Nathaniel Chandra et al., 2019)	Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram	Naïve Bayes	Mendapat akurasi terbaik sebesar 78,66%	— 764 data berita dari kompas.com dan radarmalang.com — Kelas politik, ekonomi, news, edukasi, kesehatan, travel, dan olahraga

Lanjutan pengertian terkait dengan penelitian

No.	Peneliti	Judul	Metode	Hasil	Batasan Penelitian
11.	(Briliansyah, 2020)	Sistem Klasifikasi Kategori Berita Mengunakan Metode K-Nearest Neighbor	K-Nearest Neighbor	Akurasi sebesar 74%	— 400 data berita diambil dari detik.com dan kompas.com — Kelas olahraga, teknologi, ekonomi, dan politik
12.	(Huda, 2020)	Klasifikasi Berita Menggunakan Metode Learning Vector Quantization	Learning Vector Quantization	Akurasi sebesar 96,67%	— 300 data berita diambil dari website kompas, detik, halodoc, cnnindonesia dan kumparan — Kelas politik, kesehatan, dan edukasi
13.	(Hasanah, 2022)	Klasifikasi Berita Online Menggunakan Metode Artificial Neural Network (ANN)	Artificial Neural Network (ANN)	Akurasi sebesar 96,29%	— 360 data berita diambil dari website detik, kompas, cnnindonesia, sindonews, dan republic — Kelas politik, teknologi, travel, hiburan, olahraga, pendidikan, ekonomi, hukum dan kesehatan
14.	Istiani (2020)	Implementasi Metode Random Forest untuk Klasifikasi Kategori Berita Online	Random Forest	Akurasi sebesar 79%	— 1115 data berita diambil dari kompas.com dan detik.com — Kelas edukasi, keuangan, kesehatan, olahraga dan teknologi

Berdasarkan Tabel 2.1 terdapat beberapa metode yang bisa digunakan dalam klasifikasi teks seperti *Random Forest*, *Naïve Bayes*, SVM, ANN, LVQ, *Multilayer Perceptron*, dan *Logistic Regression*. Dari beberapa metode yang digunakan, beberapa metode memiliki tingkat akurasi yang tinggi seperti metode

Random Forest yang menghasilkan akurasi 95% pada penelitian Basar *et al*. 2022, metode LVQ dengan akurasi 96,67% pada penelitian Huda 2020, serta ANN dengan akurasi 96, 29% pada penelitian Hasanah 2022.

2.2 Klasifikasi

Klasifikasi merupakan salah satu tipe pemodelan pada data mining yang mana dapat membantu serta mengatur kumpulan data rumit dan besar. Teknik klasifikasi dapat mendeskripsikan data set dengan tipe biner maupun nominal (Nofriansyah & Nurcahyo, 2015). Menurut Jolyta *et al.* (2020) klasifikasi merupakan proses pengumpulan data yang didasarkan atas sekumpulan kesamaan yang telah ditentukan. Tujuan teknik klasifikasi adalah menemukan model atau fungsi untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Kursini dalam Annur, 2018)

Menurut Jolyta *et al.* (2020) proses klasifikasi dibagi menjadi 3 tahapan yakni:

1) Proses *learning* yakni membangun model dari kumpulan training set untuk mendapatkan deskripsi tentang model klasifikasi.
2) Menguji serangkaian data baru ke dalam model yang telah dibuat untuk mengevaluasi akurasi dari model tersebut.
3) Melakukan proses klasifikasi dengan menggunakan metode klasifikasi seperti *Neural Network*, *Decision Tree*, maupun *Random Forest*.

2.3 Preprocessing

Preprocessing merupakan tahapan awal penelitian yang bertujuan untuk merubah data mentah menjadi data yang bisa diolah pada proses selanjutnya dengan

melalui proses seperti mengurangi jumlah kosa kata, maupun menyamakan bentuk kata (Hidayat & Rizqi, 2020). Terdapat beberapa tahapan dalam *preprocessing* sebagai berikut.

1. Case Folding

   Case folding merupakan proses untuk menyeragamkan teks dengan mengubah semua huruf kapital yang ada dalam dokumen menjadi huruf kecil. Pengubahan karakter hanya dilakukan pada huruf a sampai z, selain karakter huruf dan angka seperti tanda baca, koma maupun spasi yang biasa disebut dengan delimiter atau karakter khusus bisa diabaikan atau dihapus menggunakan perintah yang ada di python.

2. Stopword Removal

   Stopword removal merupakan proses pemilihan kata-kata yang penting sehingga kata-kata yang dianggap tidak penting atau tidak mempunyai keuntungan pada pemrosesan teks akan dihapus. *Stopword* umumnya berupa kata yang sering muncul seperti kata ganti orang, “dan”, “atau”, “juga”, “dari”, “ke” dan sebagainya. Tahap ini dilakukan dengan membandingkan kata yang ada dalam dokumen dengan daftar stopword. Kata-kata yang cocok akan dihapus.

3. Stemming

   Stemming merupakan proses menghilangkan imbuhan yang ada pada kata menjadi kata dasar seperti kata “berlari” diubah menjadi “lari”, “memasak” diubah menjadi “masak”, dan lain-lain. Tujuan dari *stemming* untuk

menghilangkan variasi kata yang tidak diperlukan sehingga dapat meningkatkan efisiensi dalam pemrosesan teks.

4. Tokenizing

Tokenizing merupakan proses pemisahan kalimat menjadi unit-unit yang lebih kecil atau kata-kata terpisah. Pemisahan kata dilakukan berdasarkan spasi antar dua kata. Penghilangan angka, tanda baca maupun karakter juga dilakukan pada tahap ini karena dianggap tidak memiliki pengaruh pada pemrosesan teks (Zulkifli & Suadaa, 2018).

2.4 TF-IDF

TF-IDF merupakan proses pembobotan kata atau term yang bertujuan untuk mendapatkan nilai dari masing-masing kata (Rohman *et al.*, 2018). TF-IDF mengukur tingkat kepentingan kata dari suatu dokumen dinyatakan dalam variabel $W_{d,t}$ yang merupakan bobot kata dalam suatu dokumen. Metode ini menggabungkan dua konsep yaitu *Term Frequency* (TF) yang mengukur frekuensi kemunculan sebuah kata dalam dokumen yang ditunjukan pada rumus 2.1. Sedangkan *Inverse Document Frequency* (IDF) menunjukkan frekuensi dokumen yang mengandung suatu kata tertentu yang ditunjukan pada rumus 2.2. Berikut perhitungan TF-IDF yang ditunjukkan pada rumus 2.3.

$$TF_{d,t} = \frac{jumlah\ munculnya\ kata\ t\ dalam\ dokumen}{Total\ jumlah\ keseluruhan\ kata\ dalam\ dokumen} \quad (2.1)$$

$$IDF = \log_2\left(\frac{D}{df}\right) \quad (2.2)$$

$$W_{d,t} = TF_{d,t} * IDF \quad (2.3)$$

Di mana W merupakan bobot dokumen ke-n, d adalah dokumen, t adalah kata kunci, df adalah jumlah dokumen yang mengandung kata kunci dan D adalah total dokumen (Findawati & Rosid, 2020).

2.5 Random Forest

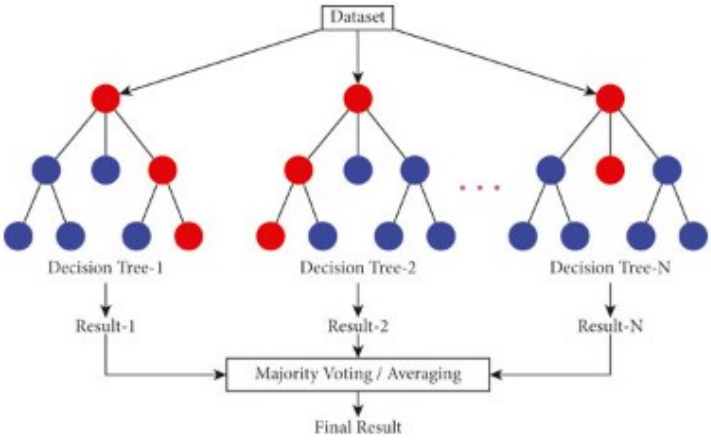
Random Forest pertama kali diperkenalkan oleh Breiman pada tahun 2001. Menurut Pamuji & Ramadhan (2021) random forest memiliki beberapa kelebihan yakni memiliki performa baik dalam melakukan klasifikasi, menghasilkan tingkat eror yang relatif rendah, bekerja secara efisien dalam data dengan jumlah besar, dan efektif dalam menangani *missing* data.

Metode *Random Forest* adalah salah satu jenis metode dari *ensemble learning* yang biasa digunakan untuk pengklasifikasian dan regresi *(Ramadhan et al.*, 2022). *Ensemble Learning* yakni menggabungkan beberapa pengklasifikasian untuk meningkatkan kinerja model. Menurut Dietterich dalam Willy *et al.* (2021) pada *ensemble learning*, pembelajaran dilakukan dengan voting dari sekumpulan hipotesa yang telah dibuat. *Output* class dipilih dari hipotesa dengan vote tertinggi.

Menurut Sadewo dalam Primajaya & Sari (2018) terdapat tiga aspek penting dalam metode *Random Forest* yakni:

1) Membangun pohon keputusan dengan teknik *bootstrap sampling*
2) Setiap pohon keputusan melakukan prediksi secara acak
3) *Random Forest* melakukan prediksi dengan proses voting. Untuk klasifikasi menggunakan nilai mayoritas atau nilai yang paling sering muncul, sedangkan untuk regresi dilakukan dengan menggunakan nilai rata-rata.

Berikut merupakan gambaran umum dari alur kerja Random Forest yang ditunjukkan pada gambar 2.1.



Gambar 2. 1 Alur kerja Random Forest

Model dalam *Random Forest* dibangun dengan menggunakan algoritma pohon keputusan atau (*Decision Tree*). Algoritma yang dapat digunakan yakni ID3 yang mana dalam pemisahan simpul didasarkan pada nilai *entropy* dan algoritma CART berdasarkan nilai gini (Ashfania *et al.*, 2023). Febriansyah *et al.* (2023) mendefinisikan gini sebagai ukuran ketidakseimbangan dalam distribusi atribut pada sekumpulan objek data. Sedangkan *entropy* merupakan ukuran impuritas atau ketidakpastian atribut pada sekumpulan data.

Berikut rumus pemisahan simpul didasarkan pada pencarian *entropy* yang ditunjukkan pada rumus 2.4.

$$Entropy\ (\mathrm{S}) = \sum_{i=1}^{n} -P_i * \log_2 P_i \qquad (2.4)$$

Di mana nilai S merupakan himpunan kasus, n adalah jumlah partisi S, dan P_i adalah proporsi himpunan kasus ke-i terhadap himpunan kasus.

Selanjutnya dilakukan pencarian *information gain*. *Information gain* merupakan perolehan informasi menggunakan *entropy* untuk menentukan atribut terbaik (Norhalimi & Siswa, 2022). Berikut rumus persamaan pencarian *information gain* yang ditujukan pada rumus 2.5.

$$Gain\ (\mathrm{S, A}) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy\ (Si) \qquad (2.5)$$

Di mana S merupakan himpunan kasus, A sebagai atribut, n merupakan jumlah partisi atribut A, $|S_i|$ adalah proporsi S_i terhadap S, $|S|$ adalah jumlah kasus dalam S, dan *entropy* (S_i) merupakan entropy untuk sampel memiliki nilai ke-*i* (Suntoro, 2019).

Berikut algoritma Random Forest menurut (Jackins *et al.*, 2021).

1. Pilih fitur secara acak sebanyak "n" dari total "k" fitur. n<k
2. Hitung simpul fitur "n" menggunakan titik pemisah terbaik
3. Kategorikan simpul kedalam simpul anak dengan pemisahan terbaik
4. Ulangi langkah 1 sampai 3 sampai jumlah simpul tercapai
5. Ulangi langkah 1 sampai 4 sebanyak "n" untuk membuat "n" pohon

2.6 Evaluasi Kinerja

Evaluasi kinerja dalam penelitian ini menggunakan *confusion matrix* yang digunakan untuk mengukur nilai *accuracy, precision, recall, f-measure,* TPR dan FPR. *Confusion matrix* merupakan metode yang sering digunakan pada machine learning untuk mengevaluasi atau mengukur kinerja sistem yang dibangun. Berikut merupakan konsep dari *confusion matrix* yang ditunjukkan pada tabel 2.2.

Tabel 2. 2 Konsep confusion matrix

Confusion Matrix		Kelas Aktual	
		Positif	Negatif
Kelas Prediksi	Positif	TP	FP
	Negatif	FN	TN

Metode confusion matrix ini digunakan untuk perbandingan antara data hasil prediksi dengan data aktual (Wiraswendro & Soetanto, 2022). Pengujian dilanjutkan dengan menghitung nilai *accuracy, precision, recall f-measure,* TPR dan FPR. *Accuracy* digunakan untuk mengatahui seberapa akurat model dapat mengklasifikasikan data target dengan benar yang ditunjukkan pada rumus 2.6. *Precision* menggambarkan sejauh mana keakuratan data yang diminta dengan hasil prediksi model yang ditunjukkan pada rumus 2.7. *Recall* menunjukkan tingkat keberhasilan sistem dalam mengidentifikasi semua kasus dan menemukan kembali suatu informasi yang ditunjukkan pada rumus 2.8. *F-measure* merupakan metrix gabungan dari *precision dan recall* yang ditunjukkan pada rumus 2.9. TPR (*True Positive Rate*) bisa disebut sebagai *Recall* atau *Sensitivity* mengukur sejauh mana model mengidentifikasi kasus positif yang diklasifikasikan dengan benar sebagai kasus positif yang ditunjukkan pada rumus 2.10. Sedangkan FPR (*False Positive Rate*) mengukur seberapa sering model salah mengidentifikasi kelas negatif yang diklasifikasikan positif yang ditunjukkan pada rumus 2.11 (Dinatha & Rakhmawati, 2020).

$$\text{Accuracy} = \frac{Jumlah\ klasifikasi\ benar}{Jumlah\ dokumen\ uji} = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (2.6)$$

$$\text{Precision} = \frac{TP}{TP+FP}\ x\ 100\% \quad (2.7)$$

$$\text{Recall} = \frac{TP}{TP+FN}\ x\ 100\% \quad (2.8)$$

$$\text{F-Measure} = \frac{2\ X\ Precision\ X\ Recall}{Precision+Recall} \quad (2.9)$$

$$\text{TPR} = \frac{TP}{TP+FN}\ x\ 100\% \quad (2.10)$$

$$\text{FPR} = \frac{FP}{FP+TN}\ x\ 100\% \quad (2.11)$$

Keterangan:

True Positive (TP)	: Jumlah data positif yang benar terklasifikasi sebagai data positif.
False Positive (FP)	: Jumlah data negatif yang salah terklasifikasi sebagai data positif.
False Negative (FN)	: Jumlah data positif yang salah terklasifikasi sebagai data negatif.
True Negative (TN)	: Jumlah data negatif yang benar terklasifikasi sebagai data negatif.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Pengumpulan Data

Data pada penelitian ini diambil dari media berita online yakni detik.com dan kompas.com menggunakan teknik web *scrapping*. Data yang diambil berupa link, judul, konten berita dan kategori. Data yang berhasil dikumpulkan disimpan dalam file excel dengan format csv.

Tabel 3. 1 Jumlah data masing-masing kategori

No	Kategori Berita	Jumlah
1.	Teknologi	225
2.	Edukasi	223
3.	Keuangan	219
4.	Olahraga	223
5.	Kesehatan	225
Total		1115

Tabel 3.1 merupakan jumlah masing-masing kategori berita yang telah didapatkan dengan jumlah total sebanyak 1115 data berita Terdapat 5 kategori berita yang digunakan dalam penelitian ini yakni teknologi, edukasi, keuangan, olahraga, dan kesehatan.

Tabel 3.2 Contoh dataset

D	Konten	Kategori
D1	Hari Guru Nasional atau HGN diperingati setiap tanggal 25 November	Edukasi
D2	Menteri Keuangan Sri Mulyani Indrawati mengatakan generasi milenial dan Z punya kisah perjuangan yang sama dengan perjalanan Indonesia	Keuangan
D3	China Masters 2023 sudah memasuki babak perempat final	Olahraga
D4	Pengembang sekaligus penerbit game, Toge Productions, mengakuisisi pengembang game asal Surabaya, Mojiken Studio	Teknologi
D5	Kementerian Kesehatan (Kemenkes) RI mencatat ada 57 kasus cacar monyet	Kesehatan

Lanjutan contoh dataset

D	Konten	Kategori
D6	Menanamkan rasa gemar membaca memberi dampak baik bagi anak	Edukasi
D7	Pramudya Kusumawardana/Yeremia Rambitan melaju ke perempat final China Masters 2023	Olahraga
D8	Konsol game berbasis Windows Asus ROG Ally versi yang lebih murah, kini sudah bisa dibeli di Indonesia	Teknologi
D9	Menteri Keuangan Sri Mulyani Indrawati memaparkan hingga bulan Oktober 2023 penerimaan pajak pemerintah sudah tembus Rp 1.523,7 triliun	Keuangan
D10	Apriyani Rahayu/Siti Fadia Silva Ramadhanti mundur dari babak kedua China Masters 2023	Olahraga
D11	Karies gigi menjadi masalah kesehatan gigi yang kerap terjadi pada banyak orang	Kesehatan
D12	ADHD pada anak sering kali berlanjut sampaai dewasa	Kesehatan

Tabel 3.2 merupakan contoh kalimat dalam dokumen pada dataset yang dijadikan contoh implementasi sistem. D merupakan dokumen dan D1 merupakan dokumen 1, D2 merupakan dokumen 2 dan seterusnya.

3.2 Akuisisi Data

Pada penelitian ini, sebelum data dilakukan beberapa skenario pengujian, terlebih dahulu kategori data diubah kedalam bentuk numerik menggunakan label encoding untuk dapat diterapkan dalam sistem yang telah dibuat.

Tabel 3. 3 Label Encoding

No.	Kategori	Label Encoding
1.	Edukasi	0
2.	Kesehatan	1
3.	Keuangan	2
4.	Olahraga	3
5.	Teknologi	4

Tabel 3.3 merupakan hasil label yang telah dilakukan proses *encoding* menggunakan *library* yang telah tersedia di phyton yakni *LabelEncoder.* Dalam label *encoding*, pelabelan kategori diurutkan sesuai abjad sehingga kategori edukasi 0, kesehatan 1, keuangan 2, olahraga 3, dan teknologi 4.

Sebelum dilakukan pengujian, terlebih dahulu 1115 data dibagi ke dalam data *training* dan data *testing* dengan tiga skenario perbandingan yakni 90:10, 80:20 dan 70:30. Berikut pembagian data *training* dan data *testing* setiap kategori yang ditunjukkan pada tabel 3.4.

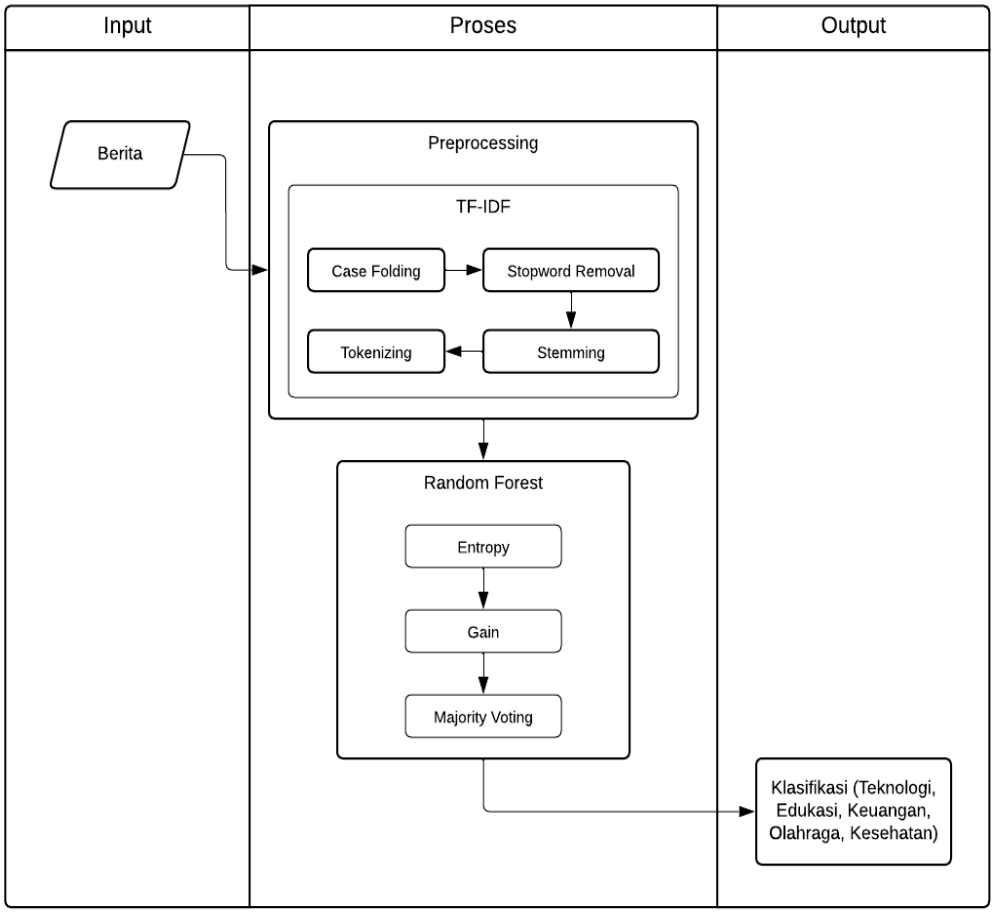
Tabel 3. 4 Perbandingan data training dan data testing

Perbandingan 70:30			
No.	**Kategori**	**Training**	**Testing**
1.	Edukasi	156	67
2.	Kesehatan	157	68
3.	Keuangan	153	66
4.	Olahraga	156	67
5.	Teknologi	157	68
Total		779	336
Perbandingan 80:20			
No.	**Kategori**	**Training**	**Testing**
1.	Edukasi	178	45
2.	Kesehatan	180	45
3.	Keuangan	175	44
4.	Olahraga	178	45
5.	Teknologi	180	45
Total		891	224
Perbandingan 90:10			
No.	**Kategori**	**Training**	**Testing**
1.	Edukasi	200	23
2.	Kesehatan	202	23
3.	Keuangan	197	22
4.	Olahraga	200	23
5.	Teknologi	202	23
Total		1001	114

3.3 Desain Sistem

Desain sistem pada penelitian ini ditunjukkan pada gambar 3.1 yang diawali dengan penginputan data berita, kemudian melakukan *preprocessing* menggunakan TF-IDF yang didalamnya terdapat tahapan berupa *case folding, stopword removal, stemming dan tokenizing*. Selanjutnya dilakukan klasifikasi menggunakan *Random Forest* dengan membentuk pohon keputusan yang didalamnya terdapat perhitungan

entropy dan *gain* serta pengambilan keputusan menggunakan majority voting untuk mendapatkan hasil klasifikasi kategori berita.



Gambar 3.1 Desain sistem

3.4 Preprocessing

Preprocessing merupakan tahapan awal penelitian yang bertujuan mengubah data mentah menjadi data yang dapat diolah. Proses ini memiliki 4 tahapan yakni *case folding, stopword removal, stemming*, dan *tokenizing*. Dalam contoh perbedaan hasil dokumen sebelum dan sesudah dilakukan *preprocessing*, peneliti mengambil satu sampel dari dokumen 1 (D1).

3.4.1 Case Folding

Berikut perbedaan hasil dokumen sebelum dan setelah dilakukannya *case folding* yang ditunjukkan pada table 3.5.

Tabel 3. 5 Contoh hasil case folding

Sebelum	Setelah
Hari Guru Nasional atau HGN diperingati setiap tanggal 25 November	hari guru nasional atau hgn diperingati setiap tanggal November

3.4.2 Stopword Removal

Berikut perbedaan hasil dokumen sebelum dan setelah dilakukannya *stopword removal* yang ditunjukkan pada tabel 3.6.

Tabel 3. 6 Contoh hasil stopword removal

Sebelum	Sesudah
hari guru nasional atau hgn diperingati setiap tanggal November	hari guru nasional hgn diperingati tanggal November

3.4.3 Stemming

Berikut perbedaan hasil dokumen sebelum dan setelah dilakukannya *stemming* yang ditunjukkan pada tabel 3.7.

Tabel 3. 7 Contoh hasil stemming

Sebelum	Sesudah
hari guru nasional hgn diperingati tanggal November	hari guru nasional hgn ingat tanggal November

3.4.4 Tokenizing

Berikut perbedaan hasil dokumen sebelum dan setelah dilakukannya *tokenizing* yang ditunjukkan pada tabel 3.8.

Tabel 3. 8 Contoh hasil tokenizing

Sebelum	Sesudah
hari guru nasional hgn ingat tanggal November	['hari', 'guru', 'nasional', 'hgn', 'ingat', 'tanggal', 'november']

Tabel 3.9 merupakan hasil dokumen setelah dilakukan tahap preprocessing yang selanjutnya akan dilakukan pembobotan kata menggunakan TF-IDF.

Tabel 3. 9 Contoh data bersih hasil preprocessing

D	Konten	Kategori
D1	hari guru nasional hgn ingat tanggal November	0
D2	menteri uang sri mulyani indrawati kata generasi milenial punya kisah juang sama jalan Indonesia	2
D3	china masters pasuk babak empat final	3
D4	kembang sekaligus terbit game toge productions akuisisi kembang game asal surabaya mojiken studio	4
D5	menteri sehat kemenkes catat kasus cacar monyet	1
D6	tanam rasa gemar baca beri dampak baik anak	0
D7	pramudya kusumawardana yeremia rambitan laju empat final china masters	3
D8	konsol game bas windows asus rog ally versi lebih murah kini bisa beli Indonesia	4
D9	menteri uang sri mulyani indrawati papar hingga bulan oktober terima pajak perintah tembus triliun	2
D10	apriyani rahayu siti fadia silva ramadhanti mundur babak dua china masters	3
D11	karies gigi jadi masalah sehat gigi kerap jadi banyak orang	1
D12	adhd anak sering kali lanjut sampai dewasa	1

3.5 TF-IDF

Pembobotan kata dalam penelitian ini menggunakan metode TF-IDF untuk mengetahui seberapa penting dan sering kemunculan suatu kata dalam suatu dokumen. Perhitungan dilakukan pada semua term atau atribut kata setelah dilakukan *preprocessing*. Pada penelitian ini, perhitungan manual menerapkan 5 sampel yang mencakup 5 kategori berita dari data pada tabel 3.9 yakni D1, D2, D3, D4 dan D5.

Tabel 3.10 Perhitungan nilai term frequency (TF)

No.	Term	Frekuensi					TF				
		D1	D2	D3	D4	D5	D1	D2	D3	D4	D5
1.	hari	1	0	0	0	0	0,143	0	0	0	0
2.	guru	1	0	0	0	0	0,143	0	0	0	0
3.	nasional	1	0	0	0	0	0,143	0	0	0	0
4.	hgn	1	0	0	0	0	0,143	0	0	0	0
5.	ingat	1	0	0	0	0	0,143	0	0	0	0

Lanjutan Perhitungan nilai term frequency (TF)

No.	Term	Frekuensi					TF				
		D1	D2	D3	D4	D5	D1	D2	D3	D4	D5
6.	Tanggal	1	0	0	0	0	0,143	0	0	0	0
7.	November	1	0	0	0	0	0,143	0	0	0	0
8.	menteri	0	1	0	0	1	0	0,071	0	0	0,143
9.	uang	0	1	0	0	0	0	0,071	0	0	0
10.	sri	0	1	0	0	0	0	0,071	0	0	0
11.	mulyani	0	1	0	0	0	0	0,071	0	0	0
12.	indrawati	0	1	0	0	0	0	0,071	0	0	0
13.	kata	0	1	0	0	0	0	0,071	0	0	0
14.	generasi	0	1	0	0	0	0	0,071	0	0	0
15.	milenial	0	1	0	0	0	0	0,071	0	0	0
16.	punya	0	1	0	0	0	0	0,071	0	0	0
17.	kisah	0	1	0	0	0	0	0,071	0	0	0
18.	juang	0	1	0	0	0	0	0,071	0	0	0
19.	sama	0	1	0	0	0	0	0,071	0	0	0
20.	jalan	0	1	0	0	0	0	0,071	0	0	0
21.	Indonesia	0	1	0	0	0	0	0,071	0	0	0
22.	china	0	0	1	0	0	0	0	0,167	0	0
23.	masters	0	0	1	0	0	0	0	0,167	0	0
24.	pasuk	0	0	1	0	0	0	0	0,167	0	0
25.	babak	0	0	1	0	0	0	0	0,167	0	0
26.	empat	0	0	1	0	0	0	0	0,167	0	0
27.	Final	0	0	1	0	0	0	0	0,167	0	0
28.	kembang	0	0	0	2	0	0	0	0	0,154	0
29.	sekaligus	0	0	0	1	0	0	0	0	0,077	0
30.	terbit	0	0	0	1	0	0	0	0	0,077	0
31.	game	0	0	0	2	0	0	0	0	0,154	0
32.	toge	0	0	0	1	0	0	0	0	0,077	0
33.	productions	0	0	0	1	0	0	0	0	0,077	0
34.	akuisisi	0	0	0	1	0	0	0	0	0,077	0
35.	asal	0	0	0	1	0	0	0	0	0,077	0
36.	surabaya	0	0	0	1	0	0	0	0	0,077	0
37.	mojiken	0	0	0	1	0	0	0	0	0,077	0
38.	Studio	0	0	0	1	0	0	0	0	0,077	0
39.	sehat	0	0	0	0	1	0	0	0	0	0,143
40.	kemenkes	0	0	0	0	1	0	0	0	0	0,143
41.	catat	0	0	0	0	1	0	0	0	0	0,143
42.	kasus	0	0	0	0	1	0	0	0	0	0,143
43.	cacar	0	0	0	0	1	0	0	0	0	0,143
44.	Monyet	0	0	0	0	1	0	0	0	0	0,143
Jumlah kata		7	14	6	13	7					

Tabel 3.10 merupakan hasil perhitungan *term frequency* (TF). Nilai *term frequency* (TF) diperoleh dari jumlah kemunculan kata dalam suatu dokumen dibagi dengan jumlah kata yang ada dalam dokumen. Jumlah kata yang dihitung didapatkan dari token kata hasil tokenisasi pada proses *preprocessing*. Sebagai contoh kata hari pada D1 hanya muncul 1 kali dan jumlah total kata yang ada dalam D1 sebanyak 7 maka hasil perhitungan TF pada kata hari di dalam D1 adalah 0,143 yang didapat dari 1 dibagi 7.

Setelah perhitungan term frequency (TF) dilakukan perhitungan IDF untuk mengetahui jumlah dokumen yang mengandung kata tertentu. Berikut hasil perhitungan IDF.

Tabel 3.11 Perhitungan nilai inverse document frequency (IDF) dan TF-IDF

No.	Term	DF	IDF	TF-IDF				
				D1	D2	D3	D4	D5
1.	hari	1	0,699	0,1	0	0	0	0
2.	guru	1	0,699	0,1	0	0	0	0
3.	nasional	1	0,699	0,1	0	0	0	0
4.	hgn	1	0,699	0,1	0	0	0	0
5.	ingat	1	0,699	0,1	0	0	0	0
6.	Tanggal	1	0,699	0,1	0	0	0	0
7.	november	1	0,699	0,1	0	0	0	0
8.	menteri	2	0,398	0	0,028	0	0	0,057
9.	uang	1	0,699	0	0,05	0	0	0
10.	sri	1	0,699	0	0,05	0	0	0
11.	mulyani	1	0,699	0	0,05	0	0	0
12.	indrawati	1	0,699	0	0,05	0	0	0
13.	kata	1	0,699	0	0,05	0	0	0
14.	generasi	1	0,699	0	0,05	0	0	0
15.	milenial	1	0,699	0	0,05	0	0	0
16.	punya	1	0,699	0	0,05	0	0	0
17.	kisah	1	0,699	0	0,05	0	0	0
18.	juang	1	0,699	0	0,05	0	0	0
19.	sama	1	0,699	0	0,05	0	0	0
20.	jalan	1	0,699	0	0,05	0	0	0
21.	Indonesia	1	0,699	0	0,05	0	0	0
22.	china	1	0,699	0	0	0,117	0	0
23.	masters	1	0,699	0	0	0,117	0	0
24.	pasuk	1	0,699	0	0	0,117	0	0
25.	babak	1	0,699	0	0	0,117	0	0
26.	empat	1	0,699	0	0	0,117	0	0
27.	Final	1	0,699	0	0	0,117	0	0

Lanjutan perhitungan nilai inverse document frequency (IDF) dan TF-IDF

No.	Term	DF	IDF	TF-IDF				
				D1	D2	D3	D4	D5
28.	kembang	1	0,699	0	0	0	0,108	0
29.	sekaligus	1	0,699	0	0	0	0,054	0
30.	terbit	1	0,699	0	0	0	0,054	0
31.	game	1	0,699	0	0	0	0,108	0
32.	toge	1	0,699	0	0	0	0,054	0
33.	productions	1	0,699	0	0	0	0,054	0
34.	akuisisi	1	0,699	0	0	0	0,054	0
35.	asal	1	0,699	0	0	0	0,054	0
36.	surabaya	1	0,699	0	0	0	0,054	0
37.	mojiken	1	0,699	0	0	0	0,054	0
38.	Studio	1	0,699	0	0	0	0,054	0
39.	sehat	1	0,699	0	0	0	0	0,1
40.	kemenkes	1	0,699	0	0	0	0	0,1
41.	catat	1	0,699	0	0	0	0	0,1
42.	kasus	1	0,699	0	0	0	0	0,1
43.	cacar	1	0,699	0	0	0	0	0,1
44.	Monyet	1	0,699	0	0	0	0	0,1

Tabel 3.11 merupakan hasil perhitungan IDF dan TF-IDF. Perhitungan *Inverse Document Frequency* (IDF) dilakukan dengan menghitung df atau jumlah dokumen yang mengandung kata terkait. Seperti contoh kata menteri mempunyai nilai df 2 karena muncul di dua dokumen berbeda yakni D2 dan D5. Setelah mendapatkan nilai df setiap kata, maka dapat dilakukan perhitungan IDF menggunakan rumus 2.2 yakni log dari pembagian jumlah dokumen pada dataset dengan nilai df. Setelah nilai IDF didapatkan, maka perhitungan TF-IDF dapat dilakukan dengan mengalikan nilai TF dan IDF menggunakan rumus 2.3. Pembobotan TF-IDF dalam penelitian ini memanfaatkan *library* yang sudah tersedia pada bahasa pemrograman phyton yakni TfidfVectorizer. Sebelum dimasukkan ke dalam model, terlebih dahulu TF-IDF dijadikan array untuk bisa diolah model dengan menjalankan program yang ditunjukkan pada gambar 3.2.

```
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

print(X_train_tfidf)
print(X_test_tfidf)

# Menampilkan term yang dihasilkan oleh TfidfVectorizer
print("\nVocabulary:")
print(tfidf_vectorizer.get_feature_names_out())

X_train_array = X_train_tfidf.toarray()
X_test_array = X_test_tfidf.toarray()
```

Gambar 3. 2 Ubah TF-IDF menjadi array

3.6 Klasifikasi Random Forest

Setelah proses *preprocessing* dan pembobotan kata dilakukan, langkah selanjutnya mengklasifikasikan artikel berita menggunakan *Random Forest*. Inputan pada algoritma *Random Forest* diperoleh dari hasil TF-IDF yang sudah dijadikan array. Hasil array TF-IDF kemudian dimasukkan ke dalam model *Random Forest* seperti pada gambar 3.3.

```
clf = RandomForest(n_trees=10)
clf.fit(X_train_array, y_train)
```

```
predictions = clf.predict(X_test_array)
```

Gambar 3. 3 Memasukkan TF-IDF ke dalam model Random Forest

BAB IV

UJI COBA DAN PEMBAHASAN

Bab ini berisi hasil pengujian sistem dengan beragam skenario pengujian dan perhitungan confusion matrix untuk mendapatkan nilai akurasi, presisi, *recall* dan *f-measure*

4.1 Uji Coba

Dalam mengevaluasi kinerja model yang telah dibuat, peneliti melakukan beberapa skenario pengujian menggunakan pemisahan data menjadi data *training* dan data *testing* dengan variasi perbandingan yang berbeda. Data *training* digunakan untuk melatih model, sedangkan data *testing* digunakan untuk menguji kinerja model yang telah dibuat. Hasil evaluasi model nantinya dibandingkan dengan *ground truth* yang mengacu pada label atau kategori berita yang ssesungguhnya. *Ground truth* dalam penelitian ini diambil dari 5 kategori yang diperoleh dari proses scrapping dataset pada situs detik.com dan kompas.com.

Tabel 4. 1 . Skenario uji coba

No.	Data *Training*	Data *Testing*
1.	90%	10%
2.	80%	20%
3.	70%	30%

Tabel 4.1 merupakan tabel perbandingan data *training* dan data *testing* untuk dilakukan uji coba sistem. Penelitian dilakukan dengan membagi rasio data untuk mengetahui rasio mana yang memiliki tingkat akurasi tertinggi. Perbandingan data *training* dan data *testing* setiap kategori berita ditampilkan dalam tabel 3.4.

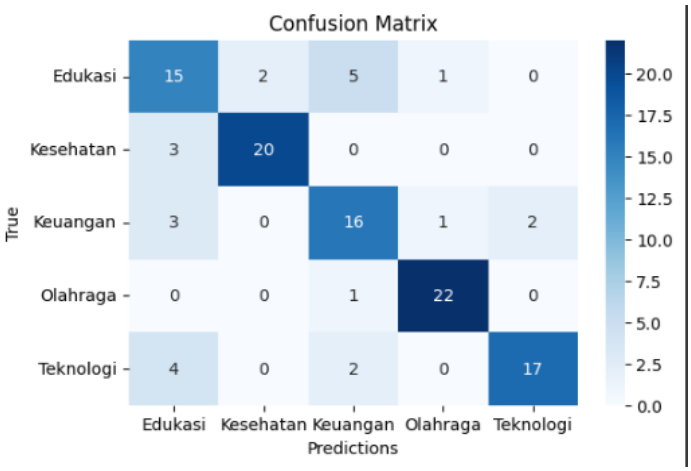
4.2 Hasil Uji Coba

Pada sub-bab ini peneliti memaparkan hasil dari skenario uji coba yang telah dijabarkan pada sub-bab 4.1. Pada pengujian pertama menggunakan perbandingan 90% data *training* sebanyak 1001 data berita dan 10% data *testing* sebanyak 114 data berita dengan masing-masing kategori yakni kategori edukasi 23 berita, kesehatan 23 berita, keuangan 22 berita, olahraga 23 berita, dan teknologi 23 berita. Hasil klasifikasi pengujian pertama menggunakan *confusion matrix* ditunjukkan pada gambar 4.1 dan tampilan *confusion matrix* ditunjukkan pada gambar 4.2. Sedangkan data hasil klasifikasi yang dibandingkan dengan *ground truth* ditunjukkan pada Lampiran 1.

Classification Report :

	precision	recall	f1-score	support
0	0.60	0.65	0.63	23
1	0.91	0.87	0.89	23
2	0.67	0.73	0.70	22
3	0.92	0.96	0.94	23
4	0.89	0.74	0.81	23
accuracy			0.79	114
macro avg	0.80	0.79	0.79	114
weighted avg	0.80	0.79	0.79	114

Gambar 4. 1 Hasil Classification Report 90:10



Gambar 4. 2 Confusion Matrix 90:10

$$\text{Akurasi} = \frac{15+20+16+22+17}{114} x100\% = 79\%$$

Tabel 4. 2 Confusion matrix 90:10

No.	Kategori	Confusion Matrix				Data Testing
		TP	FP	FN	TN	
1.	Edukasi	15	10	8	81	114
2.	Kesehatan	20	2	3	89	114
3.	Keuangan	16	8	6	84	114
4.	Olahraga	22	2	1	89	114
5.	Teknologi	17	2	6	89	114

Tabel 4.2 merupakan hasil penjumlahan TP, TN, FP, dan FN masing-masing kategori yang digunakan dalam perhitungan presisi, *recall*, *f-measure*, TPR (*True Positive Rate*) dan FPR(*False Positive Rate*).

Tabel 4. 3 Hasil presisi, recall, f-measure, tpr dan fpr

No.	Kategori	Presisi	Recall	F-measure	TPR	FPR
1.	Edukasi	60%	65%	63%	65%	11%
2.	Kesehatan	91%	87%	89%	87%	2%
3.	Keuangan	67%	73%	70%	73%	9%
4.	Olahraga	92%	96%	94%	96%	2%
5.	Teknologi	89%	74%	81%	74%	2%
Rata-rata		80%	79%	79%		

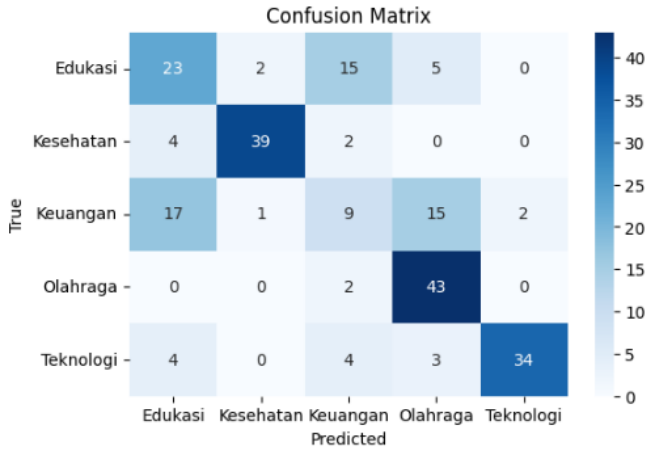
Tabel 4.3 merupakan hasil performa sistem yang meliputi nilai rata-rata *presisi* 80%, *recall* 79% dan *f-measure* sebesar 79%.

Pada pengujian kedua menggunakan perbandingan 80% data *training* sebanyak 891 data berita dan 20% data *testing* sebanyak 224 data berita dengan masing-masing kategori yakni kategori edukasi 45 berita, kesehatan 45 berita, keuangan 44 berita, olahraga 45 berita, dan teknologi 45 berita. Hasil klasifikasi pengujian kedua menggunakan *confusion matrix* ditunjukkan pada gambar 4.3 dan tampilan *confusion matrix* ditunjukkan pada gambar 4.4. Sedangkan data hasil klasifikasi yang dibandingkan dengan *ground truth* ditunjukkan pada Lampiran 2.

Classification Report :

	precision	recall	f1-score	support
0	0.48	0.51	0.49	45
1	0.93	0.87	0.90	45
2	0.28	0.20	0.24	44
3	0.65	0.96	0.77	45
4	0.94	0.76	0.84	45
accuracy			0.66	224
macro avg	0.66	0.66	0.65	224
weighted avg	0.66	0.66	0.65	224

Gambar 4. 3 Hasil Classification Report 80:20



Gambar 4. 4 Confusion Matrix 80:20

$$\text{Akurasi} = \frac{23+39+9+43+34}{224} x100\% = 66\%$$

Tabel 4. 4 Confusion matrix 80:20

No.	Kategori	Confusion Matrix				Data Testing
		TP	FP	FN	TN	
1.	Edukasi	23	25	22	154	224
2.	Kesehatan	39	3	6	176	224
3.	Keuangan	9	23	35	157	224
4.	Olahraga	43	23	2	156	224
5.	Teknologi	34	2	11	177	224

Tabel 4.4 merupakan hasil penjumlahan TP, TN, FP, dan FN masing-masing kategori yang digunakan dalam perhitungan presisi, *recall*, *f-measure*, TPR (*True Positive Rate*) dan FPR(*False Positive Rate*).

Tabel 4. 5 Hasil presisi, recall, f-measure, tpr dan fpr

No.	Kategori	Presisi	Recall	F-measure	TPR	FPR
1.	Edukasi	48%	51%	49%	51%	14%
2.	Kesehatan	93%	87%	90%	87%	2%
3.	Keuangan	28%	20%	24%	20%	13%
4.	Olahraga	65%	96%	77%	96%	13%
5.	Teknologi	94%	76%	84%	76%	1%
Rata-rata		66%	66%	65%		

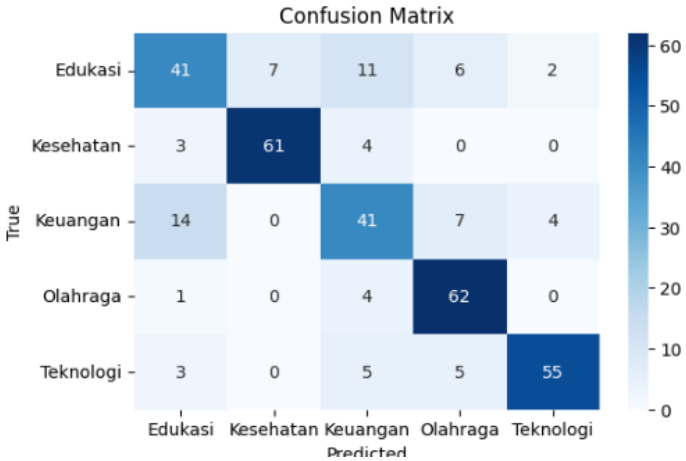
Tabel 4.5 merupakan hasil performa sistem yang meliputi nilai rata-rata *presisi* sebesar 66%, *recall* 66% dan *f-measure* 65%.

Pada pengujian ketiga menggunakan perbandingan 70% data *training* sebanyak 779 data berita dan 30% data *testing* sebanyak 336 data berita dengan masing-masing kategori yakni kategori edukasi 67 berita, kesehatan 68 berita, keuangan 66 berita, olahraga 67 berita, dan teknologi 68 berita. Hasil klasifikasi pengujian ketiga menggunakan *confusion matrix* ditunjukkan pada gambar 4.5 dan tampilan *confusion matrix* ditunjukkan pada gambar 4.6. Sedangkan data hasil klasifikasi yang dibandingkan dengan *ground truth* ditunjukkan pada Lampiran 3.

Classification Report :

	precision	recall	f1-score	support
0	0.66	0.61	0.64	67
1	0.90	0.90	0.90	68
2	0.63	0.62	0.63	66
3	0.78	0.93	0.84	67
4	0.90	0.81	0.85	68
accuracy			0.77	336
macro avg	0.77	0.77	0.77	336
weighted avg	0.77	0.77	0.77	336

Gambar 4. 5 Hasil Classification Report 70:30



Gambar 4. 6 Confusion Matrix 70:30

$$\text{Akurasi} = \frac{41+61+41+62+55}{336} x100\% = 77\%$$

Tabel 4. 6 Confusion matrix 70:30

No.	Kategori	Confusion Matrix				Data Testing
		TP	FP	FN	TN	
1.	Edukasi	41	21	26	248	336
2.	Kesehatan	61	7	7	261	336
3.	Keuangan	41	24	25	246	336
4.	Olahraga	62	18	5	251	336
5.	Teknologi	55	6	13	262	336

Tabel 4.6 merupakan hasil penjumlahan TP, TN, FP, dan FN masing-masing kategori yang digunakan dalam perhitungan presisi, *recall*, *f-measure*, TPR (*True Positive Rate*) dan FPR(*False Positive Rate*).

Tabel 4. 7 Hasil presisi, recall, f-measure, tpr dan fpr

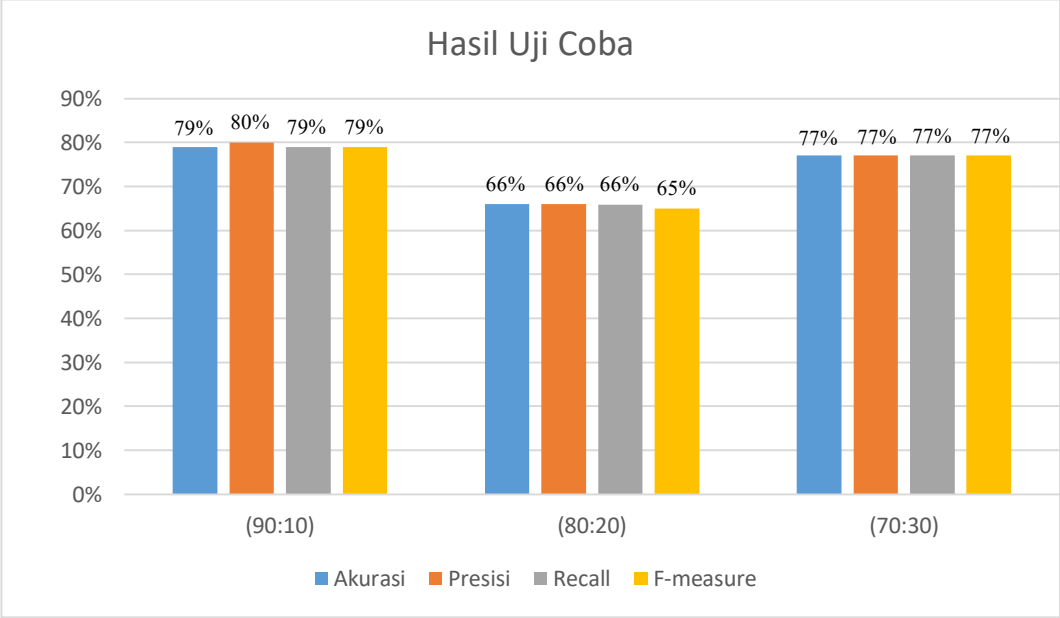
No.	Kategori	Presisi	Recall	F-measure	TPR	FPR
1.	Edukasi	66%	61%	64%	61%	8%
2.	Kesehatan	90%	90%	90%	90%	3%
3.	Keuangan	63%	62%	63%	62%	9%
4.	Olahraga	78%	93%	84%	93%	7%
5.	Teknologi	90%	81%	85%	81%	2%
Rata-rata		77%	77%	77%		

Tabel 4.7 merupakan hasil performa sistem yang menunjukkan nilai rata-rata *presisi* 77%, *recall* 77% dan *f-measure* sebesar 77%.

4.3 Pembahasan

Uji coba yang telah dilakukan menggunakan data sebanyak 1115 data berita diperoleh dari proses scrapping web pada situs berita online yakni kompas.com dan detik.com. Data hasil scrapping terlebih dahulu diolah pada proses *preprocessing* meliputi *case folding, stopword removal, stemming* dan *tokenizing* yang selanjutnya dilakukan ekstraksi fitur untuk mengubah kata menjadi numerik dengan menggunakan TF-IDF. TF-IDF terlebih dahulu diubah ke dalam bentuk array untuk bisa diproses sistem.

Berdasarkan hasil uji coba yang dilakukan pada sub-bab 4.2 dengan 3 skenario berbeda, uji coba pertama dengan perbandingan data 90:10 menghasilkan nilai akurasi sebesar 79%, presisi 80%, *recall* 79% dan *f-measure* 79%. Uji coba kedua pada perbandingan data 80:20 menghasilkan nilai akurasi sebesar 66%, presisi 66%, *recall* 66% dan *f-measure* 65%. Sementara itu, pada uji coba ketiga dengan perbandingan 70:30 menghasilkan nilai akurasi sebesar 77%%, presisi 77%, *recall* 77% dan *f-measure* 77%. Nilai akurasi, presisi, *recall*, *f-measure* dari ketiga skenario pengujian ditunjukkan pada gambar grafik 4.7.



Gambar 4. 7 Grafik hasil uji coba tiga scenario

Uji coba pertama dengan menggunakan rasio 90:10 mendapat nilai akurasi tertinggi dari ketiga skenario yakni dengan nilai 79%. Hasil akurasi tinggi ini menunjukkan bahwa model memiliki kinerja yang baik dalam mengklasifikasikan data uji ke dalam kategori yang tepat. Dengan rasio data *training* paling tinggi yakni 90% memungkinkan model dapat melakukan pembelajaran pola yang lebih baik dan lebih lengkap dari data yang dilatih. Berdasarkan hasil TPR dan FPR yang telah dilakukan, diketahui bahwa model secara umum memiliki kinerja yang baik dalam mengidentifikasi kasus positif yang berhasil diidentifikasi dengan benar yakni pada kelas 2,4 dan 5 di mana TPR tinggi dan FPR rendah. Kelas 2 memperoleh hasil TPR tinggi yakni 87% *true positive* 20 dan FPR rendah yakni 2% *false positive 2*. Kelas 4 memperoleh hasil TPR tertinggi yakni 96% dengan *true positive* 22 data dan FPR rendah 2% dengan *false positive* 2. Kelas 5 memperoleh TPR cukup tinggi yakni 74% *true positive* 17 dan FPR rendah yakni 2% *false positive* 2. Presisi dan *recall*

mendapat nilai 79% yang menunjukkan model dapat bekerja efektif dalam memprediksi kasus positif. *F-measure* dengan nilai 79% menunjukkan bahwa model memiliki keseimbangan yang baik dan kinerja yang konsisiten antara presisi dan *recall*.

Uji coba kedua dengan perbandingan 80:20 memperoleh akurasi sebesar 66%. Pada skenario kedua ini rasio data *training* lebih sedikit dari skenario pertama memungkinkan model memiliki informasi yang lebih terbatas dalam mempelajari pola data sehingga dapat mengurangi performa pada data pengujian. Berdasarkan hasil TPR dan FPR menunjukkan bahwa model efektif dalam mengidentifikasi kasus positif yang diklasifikasikan benar pada kelas 2 dan 5. Hal itu diketaui dari hasil TPR kelas 2 yakni 87% *true positive* 39 dengan FPR rendah yakni 2% *false positive* 3 dan pada kelas 5 nilai TPR 76% *true positive* 34 dan FPR rendah dengan nilai 1% *false positive* 2. Pada kelas 4 meskipun TPR mendapat nilai tertinggi yakni 96% namun FPR pada kelas ini tingggi yakni 13% yang berarti model mendapatkan banyak kesalahan dalam mengidentifikasi kasus negatif. Kelas 3 menunjukkan performa terburuk dari model untuk mengidentifikasi kasus positif yang hanya mendapat nilai 20% *true positive* 9 dan FPR tinggi yakni 13% *false positive* 23. Presisi dan *recall* mendapat nilai 66% di mana cukup rendah dari skenario pertama. Hal ini menunjukkan bahwa model cukup banyak memprediksi kasus positif yang salah (*false positive*) dengan total 76 data. *F-measure* dengan nilai 65% menunjukkan bahwa model memiliki keseimbangan antara presisi dan *recall* namun tidak optimal.

Uji coba ketiga terjadi kenaikan akurasi dengan perbandingan 70:30 memperoleh akurasi sebesar 77%. Pada skenario ketiga dengan data latih yang lebih kecil daripada skenario pertama dan kedua memungkinkan model dapat melakukan generalisasi yang lebih baik pada data uji yang besar. Berdasarkan hasil TPR dan FPR, kelas 2, 4, dan 5 menunjukkan TPR tinggi dan FPR rendah. Pada kelas 2 TPR mendapat nilai 90% *true positive* 61 dan FPR rendah 3% *false positive* 7. Pada kelas 4 TPR mendapat nilai 93% *true positive* 62 dan FPR 7% *false positive* 18. Pada kelas 5 TPR mendapat nilai 81% *true positive* 55 dan FPR rendah 2% *false positive* 6. Hal ini menunjukkan model efektif dalam mengidentifikasi kasus positif dengan benar dan jarang membuat kesalahan dalam mengidentifikasi kasus negatif sebagai kasus positif, meskipun pada kelas 4 model mengidentifikasi *false positive* sebanyak 18 namun masih terbilang cukup rendah. Kelas 1 dan 3 memiliki TPR yang lebih rendah dan nilai FPR yang lebih tinggi dari kelas lainnya menunjukkan bahwa model masih sering membuat kesalahan dalam mengidentifikasi kasus negatif sebagai positif. Presisi dan *recall* mendapat nilai 77% yang menunjukkan model dapat bekerja efektif dalam memprediksi kasus positif. *F-measure* dengan nilai 77% menunjukkan bahwa model memiliki keseimbangan yang baik dan kinerja yang konsisten antara presisi dan *recall*.

Dari hasil skenario pengujian yang dilakukan menunjukkan bahwa sistem berhasil secara efektif melakukan klasifikasi berita ke dalam 5 kategori yakni edukasi, keuangan, kesehatan, olahraga dan teknologi. Penggunaan data *training* yang lebih besar dapat menghasilkan tingkat akurasi yang lebih tinggi Azmi *et al.* (2023) yang ditunjukkan pada gambar grafik 4.7. Pada percobaan kedua model

mengalami *overfitting* karena menghasilkan akurasi *training* tinggi namun menghasilkan akurasi *testing* rendah yakni 66%. Hal ini dikarenakan faktor kualitas data *training* terpengaruh *noise* sehingga model kesulitan untuk generalisasi dengan baik pada data testing. Pada percobaan ketiga dengan data *training* yang lebih sedikit daripada percobaan kedua menghasilkan akurasi lebih tinggi yakni 77%. Hal ini dikarenakan faktor kualitas data lebih baik sehingga dapat membantu model belajar pola lebih efektif tanpa terpengaruh *noise* meskipun jumlah data *training* lebih sedikit. Hasil uji coba dapat disimpulkan bahwa jumlah dataset, data yang tidak berimbang di tiap kategori, dan rasio perbandingan data *training* dan data *testing* berpengaruh terhadap hasil akurasi.

Perspektif islam mengenai klasifikasi tidak dibahas secara langsung, namun Allah SWT menjelaskan dalam firman-Nya mengenai pemisahan antara kebenaran dan kebatilan yang mana menjadi dasar dalam konsep klasifikasi pada penelitian ini.

وَلَا تَلْبِسُوا الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُوا الْحَقَّ وَأَنْتُمْ تَعْلَمُون

"Jangan kalian mencampur kebenaran dengan kebatilan. Jangan juga kalian menyembunyikan kebenaran. Padahal kalian menyadarinya," (Q.S Al-Baqarah : 42).

Imam Jalaludin dalam kitab Tafsirul Jalalain mengemukakan bahwa yang dimaksud dengan "al-haqq" atau kebenaran dalam surah Al-Baqarah ayat 42 ini adalah kitab suci yang diturunkan kepada ahli kitab. Sedangkan yang dimaksud dengan kebatilan merupakan keterangan dusta yang mereka ada-adakan. Tafsir lain yang dikemukakan oleh Imam Al-Baidhawi dalam Kitab Anwarut Tanzil wa Asrarut Ta'wil menjelaskan bahwa kata "*talbisu*" yang berarti mencampur merujuk

pada suatu tindakan membuat sesuatu menjadi mirip dengan yang lain sehingga bisa diartikan dengan “janganlah kalian mencampur kebenaran dengan kebatilan yang direkayasa dan menyembunyikan kebenaran tersebut sehingga keduanya tidak dapat dibedakan”.

Dalam hal klasifikasi atau pengelompokan, surah Al-Baqarah menekankan pentingnya untuk pemisahan kebenaran dan kebatilan. Dalam hal ini, pemisahan atau pengelompokan kategori berita sebaiknya dilakukan secara otomatis menggunakan sistem karena lebih akurat dan untuk menghindari kesalahan yang dibuat oleh editor.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian yang telah dilakukan menggunakan metode *Random Forest* untuk klasifikasi kategori berita online dengan data sejumlah 1115 data berita dimana tiap kategori dibagi ke dalam beberapa rasio pembagian data *training* dan data *testing*. Uji coba model dilakukan dengan tiga skenario berbeda untuk mengetahui perbedaan performa model. Uji coba pertama dilakukan dengan perbandingan 90:10 menghasilkan akurasi sebesar 79%, presisi 80%, *recall* 79%, *f-measure* sebesar 79%. Uji coba kedua dilakukan dengan perbandingan 80:20 menghasilkan akurasi sebesar 66%, presisi 66%, *recall* 66% dan *f-measure* sebesar 65%. Uji coba ketiga dengan perbandingan data *training* dan *testing* 70:30 menghasilkan akurasi sebesar 77%, presisi 77%, *recall* 77% dan *f-measure* sebesar 77%. Dari uji coba yang dilakukan skenario pertama mendapat hasil akurasi terbaik yakni 79% pada uji coba dengan perbandingan data *training* sebesar 90% dan data *testing* sebesar 10% dapat disimpulkan bahwa jumlah dataset dan rasio perbandingan data *training* dan data *testing* berpengaruh terhadap performa model.

5.2 Saran

Setelah dilakukan penelitian terhadap klasifikasi kategori berita online menggunakan metode Random Forest. Peneliti menyadari adanya kekurangan pada penelitian ini, sehingga dibutuhkan kritik dan saran membangun untuk perbaikan penelitian di masa mendatang. Berikut perbaikan yang disarankan:

1. Mencoba dan membandingkan penelitian dengan metode klasifikasi lain seperti *Support Vector Machine* (SVM), *Naive Bayes*, atau model berbasis *neural network* seperti LSTM atau BERT.
2. Menggunakan metode *K-Fold Cross Validation* untuk mengevaluasi model sehingga performa model dapat lebih akurat karena metode ini dapat memilih model terbaik dan dapat menghindari *overfitting*.
3. Menggunakan seleksi fitur yang lain seperti *Bag of Words* (BoW), *Word2Vec*, GloVe dan seleksi lainnya.
4. Menggunakan teknik seperti *Grid Search* atau *Random Search* untuk menentukan *hyperparameter* terbaik. *Hyperparameter* yang optimal dapat meningkatkan performa model.

DAFTAR PUSTAKA

Annur, H. (2018). Klasifikasi Masyarakat Miskin menggunakan Metode Naïve Bayes. *ILKOM Jurnal Ilmiah*, *10*(2), 160–165.

Ariadi, D., & Fithriasari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *JURNAL SAINS DAN SENI ITS Vol. 4, No.2*, *4*(2), 248–253.

Ashfania, G. A. M., Prahasto, T., Widodo, A., & Warsokusumo, T. (2023). *Penggunaan Algoritma Random Forest untuk Klasifikasi berbasis Kinerja Efisiensi Energi pada Sistem Pembangkit Daya*. *24*(3), 14–21.

Azmi, B. N., Hermawan, A., & Avianto, D. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *4*(4), 281–290. https://doi.org/10.35746/jtim.v4i4.298

Basar, T. F., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(3), 1426–1433. http://j-ptiik.ub.ac.id

Baskoro, B. B., Susanto, I., & Khomsah, S. (2021). Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR). *INISTA (Journal of Informatics Information System Software Engineering and Applications)*, *3*(2), 21–29. https://doi.org/10.20895/INISTA.V3

Briliansyah, F. (2020). Sistem klasifikasi kategori berita menggunakan metode k-nearest neighbor. *Skripsi*. http://etheses.uin-malang.ac.id/18635/%0Ahttp://etheses.uin-malang.ac.id/18635/1/13650056.pdf

Dinatha, P. M. R. C., & Rakhmawati, N. A. (2020). Komparasi Term Weighting dan Word Embedding pada Klasifikasi Tweet Pemerintah Daerah. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *9*(2), 155–161. https://doi.org/10.22146/jnteti.v9i2.90

Febriansyah, F., Asti Dwiyanti, Z., & Diash Firdaus. (2023). Deteksi Serangan Low Rate Ddos Pada Jaringan Tradisional Menggunakan Machine Learning Dengan Algoritma Decision Tree. *Cyber Security Dan Forensik Digital*, *6*(1), 6–11. https://doi.org/10.14421/csecurity.2023.6.1.3951

Findawati, Y., & Rosid, M. A. (2020). *Buku Ajar*.

Habib, S. M., Haerani, E., Gusti, S. K., & Ramadhani, S. (2022). Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *5*(2), 248–258. https://doi.org/10.32672/jnkti.v5i2.4191

Hasanah, U. U. (2022). Klasifikasi Berita Online Menggunakan Metode Artificial Neural Network (ANN). *Skripsi*, *8.5.2017*, 2003–2005.

Hidayat, E. Y., & Rizqi, M. A. (2020). Klasifikasi Dokumen Berita Menggunakan

Algoritma Enhanced Confix Stripping Stemmer dan Naïve Bayes Classifier. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *6*(2), 90–99.

Huda, N. (2020). Klasifikasi Berita Menggunakan Metode Learning Vector Quantization. *Skripsi*, 107.

Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *Journal of Supercomputing*, *77*(5), 5198–5219. https://doi.org/10.1007/s11227-020-03481-x

Jolyta, D., Ramdhan, W., & Zarlis, M. (2020). *Konsep Data Mining dan Penerapan*. Deepublish.

Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. *... Teknologi Informasi Dan ...*, *6*(9), 4305–4313. http://j-ptiik.ub.ac.id

Morama, H. C., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(4), 1702–1708. http://j-ptiik.ub.ac.id

Munirul, U., Mahendra, M., Alvanof, & Triandi, R. (2020). Analisa Dan Deteksi Konten Hoax Pada Media Berita Indonesia Menggunakan Machine Learning. *Jurnal Teknologi Terapan and Sains 4.0*, *1*(2).

Nathaniel Chandra, D., Indrawan, G., & Nyoman Sukajaya. (2019). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram. *Jurnal Ilmu Komputer Indonesia (JIKI)*, *4*(2), 11–19.

Nofriansyah, D., & Nurcahyo, G. W. (2015). *Algoritma Data Mining dan Pengujian*. Deepublish.

Norhalimi, M., & Siswa, T. A. Y. (2022). Optimasi Seleksi Fitur Information Gain pada Algoritma Naïve Bayes dan K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *7*(3), 237–255. https://doi.org/10.14421/jiska.2022.7.3.237-255

Pamuji, F. Y., & Ramadhan, V. P. (2021). Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotheraphy. *Jurnal Teknologi Dan Manajemen Informatika*, *7*(1), 46–50. https://doi.org/10.26905/jtmi.v7i1.5982

Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, *1*(1), 27. https://doi.org/10.24014/ijaidm.v1i1.4903

Putra, P. R. B., Indriati, & Perdana, R. S. (2023). *Klasifikasi Judul Berita Online menggunakan Metode Support Vector Machine (SVM) dengan Seleksi Fitur Chi-square*. *7*(5), 2132–2141.

Ramadhan, N. G., Adhinata, F. D., Segara, A. J. T., & Rakhmadani, D. P. (2022). Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression. *JURIKOM (Jurnal Riset Komputer)*, *9*(2), 251. https://doi.org/10.30865/jurikom.v9i2.3979

Ridwan, A. S., Chrisnanto, Y. H., & Ilyas, R. (2021). Klasifikasi Kalimat Pada Berita Olahraga Secara Otomatis Menggunakan Metode Artificial Neural Network. *Jurnal Komputer Dan Informatika*, *9*(1), 88–97.

https://doi.org/10.35508/jicon.v9i1.3708
Rohman, T. B., Purwanto, D. D., & Santoso, J. (2018). Sentiment Analysis Terhadap Review Rumah Makan di Surabaya Memanfaatkan Algoritma Random Forest. *Fakultas Sistem Informasi*, *60284*.
Rulli Hastuti, U. (2022). Konsep Layanan Perpustakaan : Analisis Tafsir Surat Al-Maidah Ayat (2). *THE LIGHT : Journal of Librarianship and Information Science*, *2*(2), 88–93. https://doi.org/10.20414/light.v2i2.6182
Suntoro, J. (2019). *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*.
Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis*, *13*(2), 67–75. https://doi.org/10.51903/e-bisnis.v13i2.247
Willy, W., Rini, D. P., & Samsuryadi, S. (2021). Perbandingan Algoritma Random Forest Classifier, Support Vector Machine dan Logistic Regression Clasifier Pada Masalah High Dimension (Studi Kasus: Klasifikasi Fake News). *Jurnal Media Informatika Budidarma*, *5*(4), 1720. https://doi.org/10.30865/mib.v5i4.3177
Wiraswendro, P. E., & Soetanto, H. (2022). Application of Random Forest Classifier Algorithm in Indonesian Sign Language System (Sibi) Detection System. *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, *19*(2), 75. https://doi.org/10.36080/bit.v19i2.2043
Yunial, agus heri. (2020). Analisa Perbandingan Algoritma Klasifikasi Support Vector Machine, Decession Tree Dan Naive Bayes. *Prosiding Seminar Informatika Dan Sistem Informasi*, *5*(2), 138–156. http://openjournal.unpam.ac.id/index.php/SNISIS/article/view/9269
Zulkifli, U. C., & Suadaa, L. H. (2018). Pengembangan Modul PreprocessingTeks untuk Kasus Formalisasi dan Pengecekan Ejaan Bahasa Indonesia pada Aplikasi Web Mining Simple Solution (WMSS). *Jurnal Matematika Statistika Dan Komputasi*, *15*(2), 95. https://doi.org/10.20956/jmsk.v15i2.5718

LAMPIRAN

Lampiran 1
Data hasil klasifikasi 90:10

No.	Aktual	Prediksi	TP	TN	FP	FN
1.	0	3	0	3	1	1
2.	0	0	1	4	0	0
3.	0	2	0	3	1	1
4.	0	0	1	4	0	0
5.	0	0	1	4	0	0
6.	0	0	1	4	0	0
7.	0	1	0	3	1	1
8.	0	0	1	4	0	0
9.	0	0	1	4	0	0
10.	0	0	1	4	0	0
11.	0	1	0	3	1	1
12.	0	2	0	3	1	1
13.	0	2	0	3	1	1
14.	0	2	0	3	1	1
15.	0	0	1	4	0	0
16.	0	2	0	3	1	1
17.	0	0	1	4	0	0
18.	0	0	1	4	0	0
19.	0	0	1	4	0	0
20.	0	0	1	4	0	0
21.	0	0	1	4	0	0
22.	0	0	1	4	0	0
23.	0	0	1	4	0	0
24.	2	2	1	4	0	0
25.	2	2	1	4	0	0
26.	2	2	1	4	0	0
27.	2	2	1	4	0	0
28.	2	2	1	4	0	0
29.	2	0	0	3	1	1
30.	2	2	1	4	0	0
31.	2	0	0	3	1	1
32.	2	2	1	4	0	0
33.	2	0	0	3	1	1
34.	2	2	1	4	0	0
35.	2	3	0	3	1	1
36.	2	2	1	4	0	0
37.	2	2	1	4	0	0
38.	2	4	0	3	1	1
39.	2	2	1	4	0	0
40.	2	2	1	4	0	0
41.	2	2	1	4	0	0
42.	2	2	1	4	0	0
43.	2	2	1	4	0	0
44.	2	2	1	4	0	0

45.	2	4	0	3	1	1
46.	3	3	1	4	0	0
47.	3	3	1	4	0	0
48.	3	3	1	4	0	0
49.	3	2	0	3	1	1
50.	3	3	1	4	0	0
51.	3	3	1	4	0	0
52.	3	3	1	4	0	0
53.	3	3	1	4	0	0
54.	3	3	1	4	0	0
55.	3	3	1	4	0	0
56.	3	3	1	4	0	0
57.	3	3	1	4	0	0
58.	3	3	1	4	0	0
59.	3	3	1	4	0	0
60.	3	3	1	4	0	0
61.	3	3	1	4	0	0
62.	3	3	1	4	0	0
63.	3	3	1	4	0	0
64.	3	3	1	4	0	0
65.	3	3	1	4	0	0
66.	3	3	1	4	0	0
67.	3	3	1	4	0	0
68.	3	3	1	4	0	0
69.	1	0	0	3	1	1
70.	1	1	1	4	0	0
71.	1	1	1	4	0	0
72.	1	0	0	3	1	1
73.	1	1	1	4	0	0
74.	1	0	0	3	1	1
75.	1	1	1	4	0	0
76.	1	1	1	4	0	0
77.	1	1	1	4	0	0
78.	1	1	1	4	0	0
79.	1	1	1	4	0	0
80.	1	1	1	4	0	0
81.	1	1	1	4	0	0
82.	1	1	1	4	0	0
83.	1	1	1	4	0	0
84.	1	1	1	4	0	0
85.	1	1	1	4	0	0
86.	1	1	1	4	0	0
87.	1	1	1	4	0	0
88.	1	1	1	4	0	0
89.	1	1	1	4	0	0
90.	1	1	1	4	0	0
91.	1	1	1	4	0	0
92.	4	4	1	4	0	0
93.	4	4	1	4	0	0
94.	4	4	1	4	0	0
95.	4	4	1	4	0	0
96.	4	0	0	3	1	1
97.	4	2	0	3	1	1

98.	4	4	1	4	0	0
99.	4	4	1	4	0	0
100	4	4	1	4	0	0
101	4	4	1	4	0	0
102	4	0	0	3	1	1
103	4	4	1	4	0	0
104	4	4	1	4	0	0
105	4	4	1	4	0	0
106	4	4	1	4	0	0
107	4	4	1	4	0	0
108	4	0	0	3	1	1
109	4	4	1	4	0	0
110	4	4	1	4	0	0
111	4	4	1	4	0	0
112	4	0	0	3	1	1
113	4	4	1	4	0	0
114	4	2	0	3	1	1

Lampiran 2
Data hasil klasifikasi 80:20

No.	Aktual	Prediksi	TP	TN	FP	FN
1.	0	3	0	3	1	1
2.	0	0	1	4	0	0
3.	0	3	0	3	1	1
4.	0	0	1	4	0	0
5.	0	2	0	3	1	1
6.	0	0	1	4	0	0
7.	0	1	0	3	1	1
8.	0	0	1	4	0	0
9.	0	0	1	4	0	0
10.	0	0	1	4	0	0
11.	0	0	1	4	0	0
12.	0	0	1	4	0	0
13.	0	0	1	4	0	0
14.	0	0	1	4	0	0
15.	0	0	1	4	0	0
16.	0	2	0	3	1	1
17.	0	0	1	4	0	0
18.	0	0	1	4	0	0
19.	0	0	1	4	0	0
20.	0	0	1	4	0	0
21.	0	3	0	3	1	1
22.	0	2	0	3	1	1
23.	0	0	1	4	0	0
24.	0	2	0	3	1	1
25.	0	2	0	3	1	1
26.	0	1	0	3	1	1
27.	0	0	1	4	0	0
28.	0	2	0	3	1	1
29.	0	2	0	3	1	1

30.	0	3	0	3	1	1
31.	0	2	0	3	1	1
32.	0	0	1	4	0	0
33.	0	0	1	4	0	0
34.	0	0	1	4	0	0
35.	0	2	0	3	1	1
36.	0	2	0	3	1	1
37.	0	2	0	3	1	1
38.	0	0	1	4	0	0
39.	0	0	1	4	0	0
40.	0	2	0	3	1	1
41.	0	2	0	3	1	1
42.	0	0	1	4	0	0
43.	0	2	0	3	1	1
44.	0	3	0	3	1	1
45.	0	2	0	3	1	1
46.	2	2	1	4	0	0
47.	2	3	0	3	1	1
48.	2	3	0	3	1	1
49.	2	3	0	3	1	1
50.	2	2	1	4	0	0
51.	2	0	0	3	1	1
52.	2	3	0	3	1	1
53.	2	0	0	3	1	1
54.	2	3	0	3	1	1
55.	2	0	0	3	1	1
56.	2	3	0	3	1	1
57.	2	0	0	3	1	1
58.	2	0	0	3	1	1
59.	2	0	0	3	1	1
60.	2	4	0	3	1	1
61.	2	2	1	4	0	0
62.	2	0	0	3	1	1
63.	2	2	1	4	0	0
64.	2	3	0	3	1	1
65.	2	0	0	3	1	1
66.	2	2	1	4	0	0
67.	2	4	0	3	1	1
68.	2	1	0	3	1	1
69.	2	0	0	3	1	1
70.	2	0	0	3	1	1
71.	2	2	1	4	0	0
72.	2	3	0	3	1	1
73.	2	0	0	3	1	1
74.	2	0	0	3	1	1
75.	2	2	1	4	0	0
76.	2	0	0	3	1	1
77.	2	0	0	3	1	1
78.	2	3	0	3	1	1
79.	2	0	0	3	1	1
80.	2	2	1	4	0	0
81.	2	3	0	3	1	1
82.	2	3	0	3	1	1

83.	2	3	0	3	1	1
84.	2	0	0	3	1	1
85.	2	3	0	3	1	1
86.	2	3	0	3	1	1
87.	2	2	1	4	0	0
88.	2	3	0	3	1	1
89.	2	0	0	3	1	1
90.	3	3	1	4	0	0
91.	3	3	1	4	0	0
92.	3	3	1	4	0	0
93.	3	2	0	3	1	1
94.	3	3	1	4	0	0
95.	3	3	1	4	0	0
96.	3	3	1	4	0	0
97.	3	3	1	4	0	0
98.	3	3	1	4	0	0
99.	3	3	1	4	0	0
100.	3	3	1	4	0	0
101.	3	3	1	4	0	0
102.	3	3	1	4	0	0
103.	3	3	1	4	0	0
104.	3	3	1	4	0	0
105.	3	3	1	4	0	0
106.	3	3	1	4	0	0
107.	3	2	0	3	1	1
108.	3	3	1	4	0	0
109.	3	3	1	4	0	0
110.	3	3	1	4	0	0
111.	3	3	1	4	0	0
112.	3	3	1	4	0	0
113.	3	3	1	4	0	0
114.	3	3	1	4	0	0
115.	3	3	1	4	0	0
116.	3	3	1	4	0	0
117.	3	3	1	4	0	0
118.	3	3	1	4	0	0
119.	3	3	1	4	0	0
120.	3	3	1	4	0	0
121.	3	3	1	4	0	0
122.	3	3	1	4	0	0
123.	3	3	1	4	0	0
124.	3	3	1	4	0	0
125.	3	3	1	4	0	0
126.	3	3	1	4	0	0
127.	3	3	1	4	0	0
128.	3	3	1	4	0	0
129.	3	3	1	4	0	0
130.	3	3	1	4	0	0
131.	3	3	1	4	0	0
132.	3	3	1	4	0	0
133.	3	3	1	4	0	0
134.	3	3	1	4	0	0
135.	1	0	0	3	1	1

136.	1	1	1	4	0	0
137.	1	1	1	4	0	0
138.	1	2	0	3	1	1
139.	1	1	1	4	0	0
140.	1	0	0	3	1	1
141.	1	1	1	4	0	0
142.	1	1	1	4	0	0
143.	1	1	1	4	0	0
144.	1	1	1	4	0	0
145.	1	1	1	4	0	0
146.	1	1	1	4	0	0
147.	1	2	0	3	1	1
148.	1	1	1	4	0	0
149.	1	1	1	4	0	0
150.	1	1	1	4	0	0
151.	1	1	1	4	0	0
152.	1	0	0	3	1	1
153.	1	1	1	4	0	0
154.	1	1	1	4	0	0
155.	1	1	1	4	0	0
156.	1	1	1	4	0	0
157.	1	1	1	4	0	0
158.	1	1	1	4	0	0
159.	1	1	1	4	0	0
160.	1	1	1	4	0	0
161.	1	1	1	4	0	0
162.	1	1	1	4	0	0
163.	1	1	1	4	0	0
164.	1	1	1	4	0	0
165.	1	1	1	4	0	0
166.	1	1	1	4	0	0
167.	1	1	1	4	0	0
168.	1	1	1	4	0	0
169.	1	1	1	4	0	0
170.	1	1	1	4	0	0
171.	1	1	1	4	0	0
172.	1	1	1	4	0	0
173.	1	1	1	4	0	0
174.	1	1	1	4	0	0
175.	1	1	1	4	0	0
176.	1	1	1	4	0	0
177.	1	1	1	4	0	0
178.	1	1	1	4	0	0
179.	1	0	0	3	1	1
180.	4	4	1	4	0	0
181.	4	4	1	4	0	0
182.	4	4	1	4	0	0
183.	4	4	1	4	0	0
184.	4	0	0	3	1	1
185.	4	2	0	3	1	1
186.	4	4	1	4	0	0
187.	4	4	1	4	0	0
188.	4	4	1	4	0	0

189.	4	4	1	4	0	0
190.	4	3	0	3	1	1
191.	4	4	1	4	0	0
192.	4	4	1	4	0	0
193.	4	4	1	4	0	0
194.	4	4	1	4	0	0
195.	4	4	1	4	0	0
196.	4	0	0	3	1	1
197.	4	4	1	4	0	0
198.	4	4	1	4	0	0
199.	4	4	1	4	0	0
200.	4	0	0	3	1	1
201.	4	4	1	4	0	0
202.	4	2	0	3	1	1
203.	4	4	1	4	0	0
204.	4	4	1	4	0	0
205.	4	3	0	3	1	1
206.	4	2	0	3	1	1
207.	4	2	0	3	1	1
208.	4	4	1	4	0	0
209.	4	4	1	4	0	0
210.	4	4	1	4	0	0
211.	4	4	1	4	0	0
212.	4	4	1	4	0	0
213.	4	4	1	4	0	0
214.	4	4	1	4	0	0
215.	4	4	1	4	0	0
216.	4	4	1	4	0	0
217.	4	4	1	4	0	0
218.	4	4	1	4	0	0
219.	4	3	0	3	1	1
220.	4	0	0	3	1	1
221.	4	4	1	4	0	0
222.	4	4	1	4	0	0
223.	4	4	1	4	0	0
224.	4	4	1	4	0	0

Lampiran 3
Data hasil klasifikasi 70:30

No.	Aktual	Prediksi	TP	TN	FP	FN
1.	0	0	1	4	0	0
2.	0	0	1	4	0	0
3.	0	0	1	4	0	0
4.	0	0	1	4	0	0
5.	0	0	1	4	0	0
6.	0	0	1	4	0	0
7.	0	1	0	3	1	1
8.	0	0	1	4	0	0
9.	0	0	1	4	0	0

10.	0	0	1	4	0	0
11.	0	1	0	3	1	1
12.	0	0	1	4	0	0
13.	0	2	0	3	1	1
14.	0	0	1	4	0	0
15.	0	0	1	4	0	0
16.	0	2	0	3	1	1
17.	0	0	1	4	0	0
18.	0	0	1	4	0	0
19.	0	0	1	4	0	0
20.	0	0	1	4	0	0
21.	0	3	0	3	1	1
22.	0	0	1	4	0	0
23.	0	0	1	4	0	0
24.	0	2	0	3	1	1
25.	0	2	0	3	1	1
26.	0	1	0	3	1	1
27.	0	0	1	4	0	0
28.	0	0	1	4	0	0
29.	0	2	0	3	1	1
30.	0	3	0	3	1	1
31.	0	2	0	3	1	1
32.	0	0	1	4	0	0
33.	0	0	1	4	0	0
34.	0	0	1	4	0	0
35.	0	0	1	4	0	0
36.	0	2	0	3	1	1
37.	0	2	0	3	1	1
38.	0	0	1	4	0	0
39.	0	0	1	4	0	0
40.	0	0	1	4	0	0
41.	0	0	1	4	0	0
42.	0	0	1	4	0	0
43.	0	2	0	3	1	1
44.	0	3	0	3	1	1
45.	0	0	1	4	0	0
46.	0	1	0	3	1	1
47.	0	0	1	4	0	0
48.	0	0	1	4	0	0
49.	0	0	1	4	0	0
50.	0	3	0	3	1	1
51.	0	4	0	3	1	1
52.	0	0	1	4	0	0
53.	0	0	1	4	0	0
54.	0	0	1	4	0	0
55.	0	1	0	3	1	1
56.	0	0	1	4	0	0
57.	0	2	0	3	1	1
58.	0	0	1	4	0	0
59.	0	1	0	3	1	1
60.	0	4	0	3	1	1
61.	0	0	1	4	0	0
62.	0	0	1	4	0	0

63.	0	0	1	4	0	0
64.	0	2	0	3	1	1
65.	0	3	0	3	1	1
66.	0	1	0	3	1	1
67.	0	3	0	3	1	1
68.	2	2	1	4	0	0
69.	2	2	1	4	0	0
70.	2	2	1	4	0	0
71.	2	2	1	4	0	0
72.	2	2	1	4	0	0
73.	2	0	0	3	1	1
74.	2	2	1	4	0	0
75.	2	2	1	4	0	0
76.	2	2	1	4	0	0
77.	2	0	0	3	1	1
78.	2	3	0	3	1	1
79.	2	0	0	3	1	1
80.	2	2	1	4	0	0
81.	2	4	0	3	1	1
82.	2	4	0	3	1	1
83.	2	2	1	4	0	0
84.	2	2	1	4	0	0
85.	2	2	1	4	0	0
86.	2	2	1	4	0	0
87.	2	2	1	4	0	0
88.	2	2	1	4	0	0
89.	2	4	0	3	1	1
90.	2	2	1	4	0	0
91.	2	0	0	3	1	1
92.	2	0	0	3	1	1
93.	2	2	1	4	0	0
94.	2	2	1	4	0	0
95.	2	0	0	3	1	1
96.	2	0	0	3	1	1
97.	2	2	1	4	0	0
98.	2	0	0	3	1	1
99.	2	0	0	3	1	1
100.	2	0	0	3	1	1
101.	2	0	0	3	1	1
102.	2	2	1	4	0	0
103.	2	0	0	3	1	1
104.	2	3	0	3	1	1
105.	2	3	0	3	1	1
106.	2	2	1	4	0	0
107.	2	3	0	3	1	1
108.	2	3	0	3	1	1
109.	2	2	1	4	0	0
110.	2	3	0	3	1	1
111.	2	0	0	3	1	1
112.	2	2	1	4	0	0
113.	2	2	1	4	0	0
114.	2	0	0	3	1	1
115.	2	2	1	4	0	0

116.	2	4	0	3	1	1
117.	2	2	1	4	0	0
118.	2	2	1	4	0	0
119.	2	2	1	4	0	0
120.	2	2	1	4	0	0
121.	2	2	1	4	0	0
122.	2	2	1	4	0	0
123.	2	2	1	4	0	0
124.	2	2	1	4	0	0
125.	2	2	1	4	0	0
126.	2	2	1	4	0	0
127.	2	2	1	4	0	0
128.	2	2	1	4	0	0
129.	2	2	1	4	0	0
130.	2	2	1	4	0	0
131.	2	2	1	4	0	0
132.	2	3	0	3	1	1
133.	2	2	1	4	0	0
134.	3	3	1	4	0	0
135.	3	3	1	4	0	0
136.	3	3	1	4	0	0
137.	3	3	1	4	0	0
138.	3	3	1	4	0	0
139.	3	3	1	4	0	0
140.	3	0	0	3	1	1
141.	3	3	1	4	0	0
142.	3	3	1	4	0	0
143.	3	3	1	4	0	0
144.	3	3	1	4	0	0
145.	3	3	1	4	0	0
146.	3	3	1	4	0	0
147.	3	3	1	4	0	0
148.	3	3	1	4	0	0
149.	3	3	1	4	0	0
150.	3	3	1	4	0	0
151.	3	2	0	3	1	1
152.	3	3	1	4	0	0
153.	3	3	1	4	0	0
154.	3	3	1	4	0	0
155.	3	3	1	4	0	0
156.	3	3	1	4	0	0
157.	3	3	1	4	0	0
158.	3	3	1	4	0	0
159.	3	3	1	4	0	0
160.	3	3	1	4	0	0
161.	3	3	1	4	0	0
162.	3	2	0	3	1	1
163.	3	3	1	4	0	0
164.	3	3	1	4	0	0
165.	3	2	0	3	1	1
166.	3	3	1	4	0	0
167.	3	3	1	4	0	0
168.	3	3	1	4	0	0

169.	3	3	1	4	0	0
170.	3	3	1	4	0	0
171.	3	3	1	4	0	0
172.	3	3	1	4	0	0
173.	3	3	1	4	0	0
174.	3	3	1	4	0	0
175.	3	3	1	4	0	0
176.	3	3	1	4	0	0
177.	3	3	1	4	0	0
178.	3	3	1	4	0	0
179.	3	3	1	4	0	0
180.	3	3	1	4	0	0
181.	3	3	1	4	0	0
182.	3	3	1	4	0	0
183.	3	3	1	4	0	0
184.	3	3	1	4	0	0
185.	3	3	1	4	0	0
186.	3	3	1	4	0	0
187.	3	3	1	4	0	0
188.	3	3	1	4	0	0
189.	3	2	0	3	1	1
190.	3	3	1	4	0	0
191.	3	3	1	4	0	0
192.	3	3	1	4	0	0
193.	3	3	1	4	0	0
194.	3	3	1	4	0	0
195.	3	3	1	4	0	0
196.	3	3	1	4	0	0
197.	3	3	1	4	0	0
198.	3	3	1	4	0	0
199.	3	3	1	4	0	0
200.	3	3	1	4	0	0
201.	1	0	0	3	1	1
202.	1	1	1	4	0	0
203.	1	1	1	4	0	0
204.	1	2	0	3	1	1
205.	1	1	1	4	0	0
206.	1	0	0	3	1	1
207.	1	1	1	4	0	0
208.	1	1	1	4	0	0
209.	1	1	1	4	0	0
210.	1	1	1	4	0	0
211.	1	1	1	4	0	0
212.	1	1	1	4	0	0
213.	1	2	0	3	1	1
214.	1	1	1	4	0	0
215.	1	1	1	4	0	0
216.	1	1	1	4	0	0
217.	1	1	1	4	0	0
218.	1	1	1	4	0	0
219.	1	1	1	4	0	0
220.	1	1	1	4	0	0
221.	1	1	1	4	0	0

222.	1	1	1	4	0	0
223.	1	1	1	4	0	0
224.	1	1	1	4	0	0
225.	1	1	1	4	0	0
226.	1	1	1	4	0	0
227.	1	1	1	4	0	0
228.	1	1	1	4	0	0
229.	1	1	1	4	0	0
230.	1	2	0	3	1	1
231.	1	1	1	4	0	0
232.	1	1	1	4	0	0
233.	1	1	1	4	0	0
234.	1	1	1	4	0	0
235.	1	1	1	4	0	0
236.	1	1	1	4	0	0
237.	1	1	1	4	0	0
238.	1	1	1	4	0	0
239.	1	1	1	4	0	0
240.	1	1	1	4	0	0
241.	1	1	1	4	0	0
242.	1	1	1	4	0	0
243.	1	1	1	4	0	0
244.	1	1	1	4	0	0
245.	1	1	1	4	0	0
246.	1	1	1	4	0	0
247.	1	1	1	4	0	0
248.	1	1	1	4	0	0
249.	1	1	1	4	0	0
250.	1	1	1	4	0	0
251.	1	1	1	4	0	0
252.	1	1	1	4	0	0
253.	1	1	1	4	0	0
254.	1	1	1	4	0	0
255.	1	1	1	4	0	0
256.	1	1	1	4	0	0
257.	1	1	1	4	0	0
258.	1	1	1	4	0	0
259.	1	1	1	4	0	0
260.	1	1	1	4	0	0
261.	1	2	0	3	1	1
262.	1	1	1	4	0	0
263.	1	1	1	4	0	0
264.	1	1	1	4	0	0
265.	1	0	0	3	1	1
266.	1	1	1	4	0	0
267.	1	1	1	4	0	0
268.	1	1	1	4	0	0
269.	4	4	1	4	0	0
270.	4	4	1	4	0	0
271.	4	4	1	4	0	0
272.	4	4	1	4	0	0
273.	4	0	0	3	1	1
274.	4	3	0	3	1	1

275.	4	4	1	4	0	0
276.	4	4	1	4	0	0
277.	4	4	1	4	0	0
278.	4	4	1	4	0	0
279.	4	3	0	3	1	1
280.	4	4	1	4	0	0
281.	4	4	1	4	0	0
282.	4	4	1	4	0	0
283.	4	4	1	4	0	0
284.	4	4	1	4	0	0
285.	4	0	0	3	1	1
286.	4	4	1	4	0	0
287.	4	4	1	4	0	0
288.	4	4	1	4	0	0
289.	4	0	0	3	1	1
290.	4	4	1	4	0	0
291.	4	2	0	3	1	1
292.	4	2	0	3	1	1
293.	4	4	1	4	0	0
294.	4	3	0	3	1	1
295.	4	2	0	3	1	1
296.	4	2	0	3	1	1
297.	4	4	1	4	0	0
298.	4	4	1	4	0	0
299.	4	4	1	4	0	0
300.	4	4	1	4	0	0
301.	4	4	1	4	0	0
302.	4	4	1	4	0	0
303.	4	4	1	4	0	0
304.	4	4	1	4	0	0
305.	4	4	1	4	0	0
306.	4	4	1	4	0	0
307.	4	4	1	4	0	0
308.	4	3	0	3	1	1
309.	4	2	0	3	1	1
310.	4	4	1	4	0	0
311.	4	4	1	4	0	0
312.	4	4	1	4	0	0
313.	4	4	1	4	0	0
314.	4	4	1	4	0	0
315.	4	4	1	4	0	0
316.	4	4	1	4	0	0
317.	4	4	1	4	0	0
318.	4	4	1	4	0	0
319.	4	4	1	4	0	0
320.	4	4	1	4	0	0
321.	4	4	1	4	0	0
322.	4	4	1	4	0	0
323.	4	4	1	4	0	0
324.	4	4	1	4	0	0
325.	4	4	1	4	0	0
326.	4	4	1	4	0	0
327.	4	4	1	4	0	0

328.	4	4	1	4	0	0
329.	4	4	1	4	0	0
330.	4	3	0	3	1	1
331.	4	4	1	4	0	0
332.	4	4	1	4	0	0
333.	4	4	1	4	0	0
334.	4	4	1	4	0	0
335.	4	4	1	4	0	0
336.	4	4	1	4	0	0