



**PENGEMBANGAN DETEKSI BERITA PALSU
PADA MEDIA SOSIAL MENGGUNAKAN
*DEEP LEARNING***

DISERTASI

Rio Yunanto
NPM: 99220708

**PROGRAM DOKTOR TEKNOLOGI INFORMASI
UNIVERSITAS GUNADARMA
2021**

Abstrak

Media sosial telah menjadi bagian tak terpisahkan didalam kehidupan masyarakat modern yang dapat memudahkan untuk saling berinteraksi dan berkomunikasi. Jika sebelumnya seseorang hanya dapat berkomunikasi dengan cara menghubungi orang tersebut melalui panggilan suara dan pesan teks, saat ini setiap aktifitas pengguna dapat diunggah ke media sosial sehingga orang lain dapat mengetahui dan mengikuti aktifitas pengguna tersebut tanpa harus menanyakan langsung melalui panggilan suara atau pesan teks. Media sosial juga memiliki dampak negatif yang membahayakan, salah satunya yaitu kasus perundungan digital dan potensi depresi terhadap anak-anak meningkat pesat, dan penggiringan opini dengan menggunakan berita palsu. Target khusus yang ingin dicapai adalah dapat mengidentifikasi data pada media sosial yang merupakan berita palsu atau propaganda digital, menghasilkan rancangan metode deteksi dini berita palsu atau propaganda digital, dan membuktikan metode tersebut efektif menghambat peredaran berita palsu atau propaganda digital pada media sosial. Penelitian ini akan dilakukan dalam beberapa tahapan yaitu: 1) identifikasi berita palsu pada media sosial, 2) analisis pola dan kriteria berita palsu, 3) merancang metode deteksi dini menggunakan deep learning, 4) mengembangkan deteksi dini berita palsu. Bila dikaitkan perkembangan strategis bidang ilmu komputer dan teknologi informasi, penelitian ini sejalan dengan dengan perkembangan teknologi yang cepat dan sesuai dengan bidang keahlian peneliti.

Kata Pengantar

Bismillahirrahmanirrahim

Assalamu'alaikum Warohmatullohi Wabarokatuh

Bagian ini adalah Kata pengantar beserta ucapan terimakasih. isi sesuai dengan kebutuhan

Bantuan moril dan materil dari berbagai pihak selama tahap penyelesaian disertasi ini sungguh sangat hebatnya. Untuk itu dengan tulus perkenankan penulis mengucapkan terima kasih kepada :

1. Yayasan Pendidikan Gunadarma, yang telah memberikan beasiswa studi di Program Doktor Teknologi Informasi Universitas Gunadarma.
2. Ibu Prof. Dr. E.S. Margianti, S.E., MM., Rektor Universitas Gunadarma dan Bapak Prof. Suryadi Harmanto, SSi., MMSI., Pembantu Rektor II Universitas Gunadarma yang telah memberikan kesempatan dan kepercayaan untuk melanjutkan studi di Program Doktor Universitas Gunadarma. Dengan segala fasilitas yang disediakan, dukungan dan semangat serta kemudahan-kemudahan yang diberikan sehingga proses perkuliahan, penelitian dan penulisan disertasi dapat berjalan dengan baik.
3. Bapak Prof. Dr. Yuhara Sukra, MSc., Koordinator Program Pasca Sarjana Universitas Gunadarma, yang telah mengajarkan kejujuran dan kebenaran di dalam penelitian, serta memberikan motivasi untuk menyelesaikan penelitian dan disertasi ini dengan rasa tanggung jawab.

4. Bapak Prof. Ir. Busono Soerowirdjo, MSc., PhD., Direktur Program Doktor Universitas Gunadarma dan Sekaligus sebagai promotor yang telah membimbing dan mengajarkan kejujuran beserta kebenaran di dalam penelitian, serta memberikan motivasi untuk menyelesaikan penelitian dan disertasi ini dengan rasa tanggung jawab.
5. Bapak Prof. Dr. Sarifuddin Madenda, Ketua Program Doktor Teknologi Informasi yang selalu memberikan semangat dan masukan, memantau proses perkembangan sampai penyelesaian disertasi ini.
6. tambahan

Semoga Allah Subhanahu wa ta'ala membalas semua keikhlasan, perhatian dan bantuan Bapak. Ibu, dan Saudara, dengan balasan kebaikan yang berlipat ganda. Kritik dan saran dari pembaca, penulis harapkan untuk perbaikan di masa yang akan datang. Kesempurnaan hanya milik Allah Subhanahu wa ta'ala, kekurangan serta kekhilafan ada di diri penulis.

Wassalamu'alaikum Warohmatullohi Wabarokatuh

Jakarta, Juli 2021

Penulis

(Rio Yunanto)

Daftar Isi

Abstrak.....	ii
Kata Pengantar.....	iii
Daftar Isi	v
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	5
1.5 Kontribusi Penelitian	5
BAB II.....	7
TELAAH PUSTAKA	7
2.1 Berita Palsu	7
2.2 Kecerdasan Buatan	12
2.3 <i>Data Mining</i>	13
2.3.1 <i>Machine Learning</i>	13
2.3.2 <i>Deep Learning</i>	14
2.4 Penelitian Terkait	15

BAB III	19
METODE PENELITIAN	19
3.1 Gambaran Umum Penelitian.....	19
3.1.1 Pengumpulan Data dan Identifikasi Berita Palsu	21
3.1.2 Analisis Algoritma dan Studi Pustaka	21
3.1.3 Perancangan Model Deteksi Dini Berita Palsu	21
3.1.4 Menguji Akurasi Deteksi Dini Berita Palsu	22
3.2 Skema Klasifikasi Berita Palsu.....	22
Daftar Pustaka.....	24

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial saat ini telah menjadi bagian tak terpisahkan didalam kehidupan masyarakat modern. Media sosial sejatinya adalah teknologi yang dapat memudahkan seseorang atau kelompok untuk saling berinteraksi dan berkomunikasi. Jika sebelumnya seseorang hanya dapat berkomunikasi dengan cara menghubungi orang tersebut melalui panggilan suara dan pesan teks, saat ini setiap aktifitas pengguna dapat diunggah ke media sosial sehingga orang lain dapat mengetahui dan mengikuti aktifitas pengguna tersebut tanpa harus menanyakan langsung melalui panggilan suara atau pesan teks. Penelitian terkait media sosial telah mampu membuat beberapa aplikasi yang mengintegrasikan konten pengguna media sosial dari berbagai wilayah tertentu untuk mengekstraksi informasi penting dalam waktu yang singkat. Salah satu contohnya berdasarkan kasus gempa bumi Jepang pada tahun 2009, dikembangkan algoritma yang menggabungkan pengguna Twitter sebagai sensor sosial untuk deteksi suatu peristiwa secara waktu nyata (Reuter and Kaufhold, 2018).

Media sosial sebagai bagian perkembangan teknologi yang tidak dapat dihindari, tentunya akan memiliki pengaruh positif dan negatif bagi penggunanya. Beberapa dampak negatif media sosial, antara lain; 1) privasi pengguna menjadi tidak ada batasan

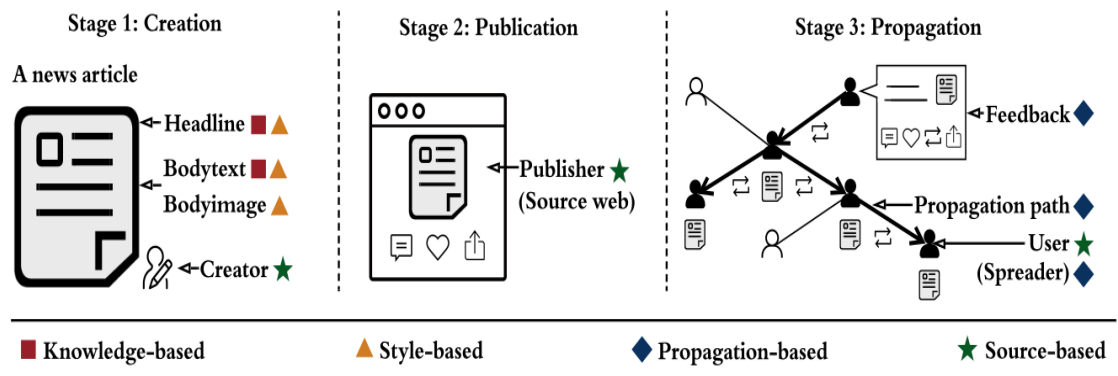
karena semua informasi diri dan aktifitas harian dapat setiap saat diunggah ke media sosial, 2) kejahatan pencurian data dan penipuan semakin meningkat karena pelaku kejahatan dapat berpura-pura menjadi seseorang tertentu menggunakan data pengguna di media sosial, 3) kasus perundungan digital dan potensi depresi terhadap anak-anak meningkat pesat, 4) penggiringan opini dengan menggunakan berita palsu meningkatkan potensi konflik horisontal terkait SARA (suku, ras, dan agama), dan lain-lain (Hidaya *et al.*, 2019). Media sosial secara tidak bijaksana mengakibatkan ketergantungan teknologi dan penggunaan yang berlebihan berdampak negatif pada organisasi. Secara khusus, terdapat tiga dimensi penggunaan media sosial yang berlebihan di tempat kerja (yaitu, sosial yang berlebihan, hedonis, dan kognitif). Dimensi ini terkait dengan konflik penggunaan teknologi dan produktifitas kerja, yang pada gilirannya menurunkan kinerja kerja karyawan (Cao and Yu, 2019).

Berita palsu berdampak sangat buruk bagi penerima yang mudah percaya, bahkan dapat membunuh karakter manusia. Berita palsu dapat manipulasi individu atau kelompok dalam jangka waktu yang lama, tanpa disadari mental masyarakat akan terbentuk kearah pemahaman yang salah. Penelitian sebelumnya melaporkan bahwa media sosial memiliki dampak yang signifikan terhadap penyebaran ketakutan dan kepanikan terkait wabah COVID-19 dan berpotensi munculnya pengaruh negatif terhadap kesehatan mental dan psikologis masyarakat. Korelasi statistik positif yang signifikan antara penggunaan media sosial dan penyebaran kepanikan terkait COVID-19 menunjukkan bahwa mayoritas remaja dan dewasa berusia 18-35 tahun menghadapi kecemasan psikologis (Ahmad and Murad, 2020).

Berbagai upaya terus dilakukan oleh pemerintah dan tokoh-tokoh masyarakat, bahkan komunitas sosial masyarakat juga melakukan edukasi pencegahan beredarnya berita palsu. Sebuah berita palsu yang muncul di media sosial bisa jadi karena disengaja atau tidak disengaja. Berita palsu yang cepat sekali viral dan tersebar luas, patut dicurigai sebagai bagian dari propaganda digital atau penggiringan opini publik. Diperlukan adanya suatu teknologi yang dapat melakukan deteksi dini berita palsu sehingga pihak-pihak terkait dapat melakukan penanggulangan dampak negatif yang terjadi.

Saat internet bukan lagi hal yang asing di Indonesia, internet bahkan sudah hadir di semua wilayah terdepan dan terluar di Indonesia meskipun dengan beberapa keterbatasan. Beberapa keterbatasan di wilayah terdepan dan terluar di Indonesia berupa sinyal selular yang tidak stabil, dan ketersediaan listrik yang belum maksimal, tidak menyurutkan antusias pengguna internet di wilayah tersebut. Pengguna internet di Indonesia tercatat mencapai 171 juta penduduk, dan 95 persen dari pengguna internet tersebut memanfaatkan internet untuk berkegiatan di media sosial (Ramadhani and Pratama, 2020). Aktivitas masyarakat yang cukup tinggi di media sosial dan layanan aplikasi pesan berakibat pada meningkatnya potensi terpapar berita palsu atau hoax. Survei yang dilakukan Mastel menunjukkan bahwa 62,80 persen penyebaran berita palsu atau hoax melalui jejaring media sosial dan aplikasi berkegiatan pesan (Gerintya, 2018).

Secara garis besar, deteksi berita palsu secara manual dapat dibagi menjadi; 1) pemeriksaan fakta berbasis ahli atau pakar, dan 2) pemeriksaan fakta berbasis bersumber data yang banyak. Pemeriksaan fakta manual berbasis ahli yaitu mengandalkan ahli atau pakar di bidang pemeriksaan fakta menghasilkan data yang sangat kredibel, mudah dikelola, dan mengarah pada hasil yang sangat akurat, tetapi mahal. Deteksi berita palsu secara manual tidak sesuai dengan kondisi saat ini dimana volume informasi baru terus menerus muncul dengan cepat, terutama di media sosial. Salah satu solusi untuk mengatasi skalabilitas munculnya berita palsu, teknik pengecekan berita secara otomatis mulai dibutuhkan dan mengandalkan teknik *Natural Language Processing* (NLP), dan *Machine Learning* (ML). Penelitian terdahulu telah dilakukan untuk mengevaluasi strategi yang dapat dilakukan untuk mendeteksi berita palsu dari empat perspektif, yaitu: 1) *knowledge-based*, 2) *style-based*, 3) *propagation-based*, dan 4) *source-based* (Zhou and Zafarani, 2020).



Gambar 1.1. Empat perspektif deteksi berita palsu (Zhou and Zafarani, 2020)

Paparan yang telah diuraikan mengindikasikan bahwa tantangan dalam menangani penyebaran berita palsu atau berita bohong di media sosial yaitu dengan membangun deteksi dini berita palsu yang bekerja secara otomatis dan cepat. Mesin deteksi dini berita palsu tersebut harus mampu mengolah data dalam jumlah yang sangat besar dan diharapkan dapat mendeteksi penyebaran berita palsu secepat mungkin, agar dampak kerusakan sosial di masyarakat dapat dicegah sedini mungkin.

1.2 Rumusan Masalah

Berdasarkan penjelasan latar belakang yang telah diuraikan diatas, maka masalah yang akan dirumuskan berkaitan dengan pengembangan metode deteksi dini berita palsu pada media sosial diantaranya sebagai berikut:

1. Pendekatan apa yang dilakukan untuk mengidentifikasi konten yang merupakan berita palsu di media sosial?
2. Bagaimana merancang model deteksi dini konten berita palsu di media sosial?

1.3 Tujuan Penelitian

Berdasarkan perumusan masalah yang diuraikan pada bab sebelumnya mengenai kedudukan korporasi dalam penyelenggaraan Negara dan Pemerintahan maka tujuan penelitian diantaranya sebagai berikut :

1. Menyusun strategi pendekatan cara mengidentifikasi berita palsu di media sosial.
2. Menghasilkan model deteksi dini berita palsu di media sosial.

1.4 Batasan Masalah

Berdasarkan tujuan penelitian tersebut, untuk menghindari meluasnya permasalahan yang ada dalam domain penelitian, penulis membuat batasan agar fokus dan penyelesaian masalah penelitian dapat tercapai. Batasan masalah dalam penelitian yang dilakukan sebagai berikut:

1. Sumber data yang digunakan diperoleh dari media sosial Twitter berbahasa Indonesia.
2. Strategi mengidentifikasi berita palsu menggunakan pendekatan berbasis konten dan berbasis konteks.

1.5 Kontribusi Penelitian

Kontribusi penelitian dapat dibagi menjadi tiga bidang yaitu, kontribusi keilmuan, kontribusi teknologi, dan kontribusi penelitian bagi masyarakat. Adapun manfaat yang diperoleh dalam penelitian pada topik pengembangan model pendeteksian berita palsu pada media sosial adalah:

1. Kontribusi Keilmuan: Menambah khazanah keilmuan dalam bidang *machine learning* serta *artificial intelligence*.

2. Kontribusi Teknologi: Menghasilkan model deteksi dini untuk memudahkan masyarakat mengenali berita palsu dan mencegah peredarannya yang dapat merugikan banyak pihak.
3. Kontribusi Masyarakat: Pengembangan dataset berita palsu berbahasa Indonesia yang dapat dimanfaatkan oleh peneliti lainnya.

BAB II

TELAAH PUSTAKA

2.1 Berita Palsu

Salah satu sisi gelap dari media sosial saat ini yaitu menjadi sarana penyebaran informasi salah dan berita palsu. Informasi yang salah dapat memengaruhi opini dan persepsi pembaca. Berita bohong bahkan dapat memanipulasi sentimen masyarakat untuk alasan politik atau ekonomi. Pengertian "berita palsu" atau "informasi salah" dapat dipahami sebagai informasi yang tidak akurat bahkan tidak jarang dibuat salah secara sengaja yang kemudian didistribusikan dengan maksud untuk menipu atau menggiring opini mereka yang membacanya.

Pada awal tahun 2020, sejumlah mitos mulai menjadi berita utama terkait dugaan pengobatan COVID-19, yang biasa disebut dengan virus corona. Banyak klaim palsu disajikan sebagai fakta, termasuk gagasan bahwa makan lebih banyak daging merah atau minyak kelapa dapat mengobati virus. Contoh misinformasi tersebut dapat membahayakan wacana publik dengan memengaruhi individu untuk membuat keputusan yang salah terkait politik atau kesejahteraan pribadi. Secara sederhana misinformasi berarti salah informasi, yaitu konten informasi didalamnya memang salah, tetapi oleh pihak yang menyebarkan informasi tersebut percaya bahwa informasi itu

benar. Penyebar informasi tak memiliki tendensi untuk membahayakan orang lain. Selain misinformasi ada juga disinformasi, yaitu pihak penyebar informasi mengetahui konten informasi didalamnya salah. Namun pihak penyebar tetap dengan sengaja menyebarkan informasi salah tersebut untuk menipu, mengancam, bahkan membahayakan pihak lain. Ada juga malinformasi yaitu konten informasi yang didalamnya benar. Tetapi justru informasi tersebut digunakan untuk mengancam keberadaan pihak lain atau pihak dengan identitas tertentu. Malinformasi dapat bisa dikategorikan ke dalam hasutan kebencian (Asyik, 2019).

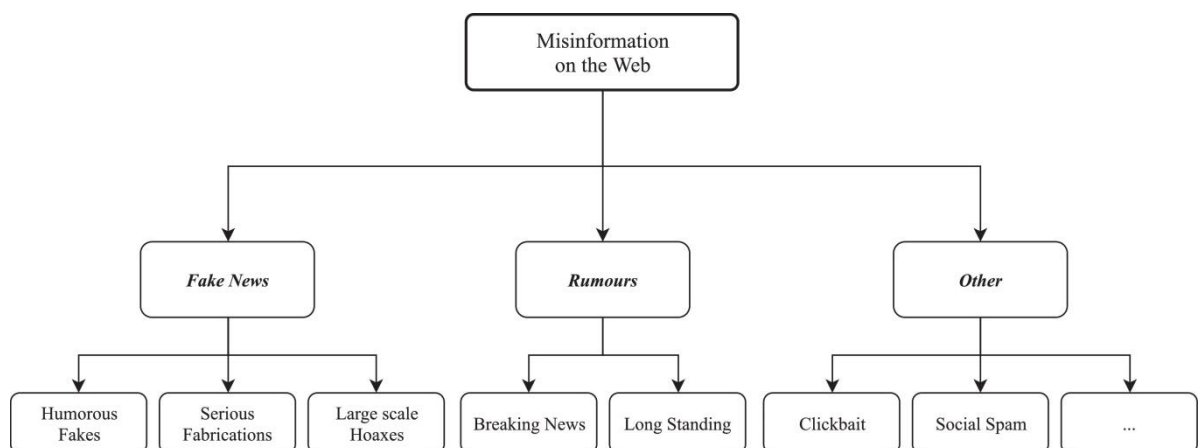
Berita palsu boleh jadi dapat dipahami sebagai berita bohong atau hoax, yang mana di dalam KBBI disebutkan bahwa hoax merupakan informasi yang direkayasa untuk menutupi informasi sebenarnya. Hoax juga dapat dipahami sebagai upaya pemutarbalikan fakta menggunakan informasi yang seolah-olah meyakinkan tetapi tidak dapat diverifikasi kebenarannya. Saat ini hoax tumbuh subur sebagai eksese negatif dari jejaring media sosial dan aplikasi pengiriman pesan. Salah satu contoh pemberitaan palsu yang paling umum adalah mengklaim sesuatu kejadian dengan suatu sebutan yang berbeda dengan kejadian yang sebenarnya untuk tujuan menghasut atau pembunuhan karakter pihak lain (Lararenjana, 2020).

Secara formal masih belum disepakati definisi universal untuk berita palsu, bahkan di dalam dunia jurnalisme. Pada dasarnya definisi yang jelas dan akurat dapat membantu untuk membedakan dan menganalisis berbagai jenis berita palsu, sehingga diperlukan beberapa konsep untuk menyajikan definisi yang luas dan sempit pada beberapa istilah berita palsu tersebut. Berdasarkan bagaimana istilah dan konsep berita palsu tersebut didefinisikan, maka dapat dibedakan satu dari yang lain berdasarkan tiga karakteristik utama, yaitu: 1) keaslian (mengandung pernyataan non-faktual atau tidak), 2) niat (bertujuan untuk menyesatkan atau menghibur publik), dan 3) konten informasi tersebut merupakan berita atau bukan. Rangkuman konsep-konsep berita palsu tersebut dapat dilihat pada Tabel 1., contohnya, disinformasi adalah informasi palsu (berita atau bukan berita) dengan niat jahat untuk menyesatkan publik (Zhou and Zafarani, 2020).

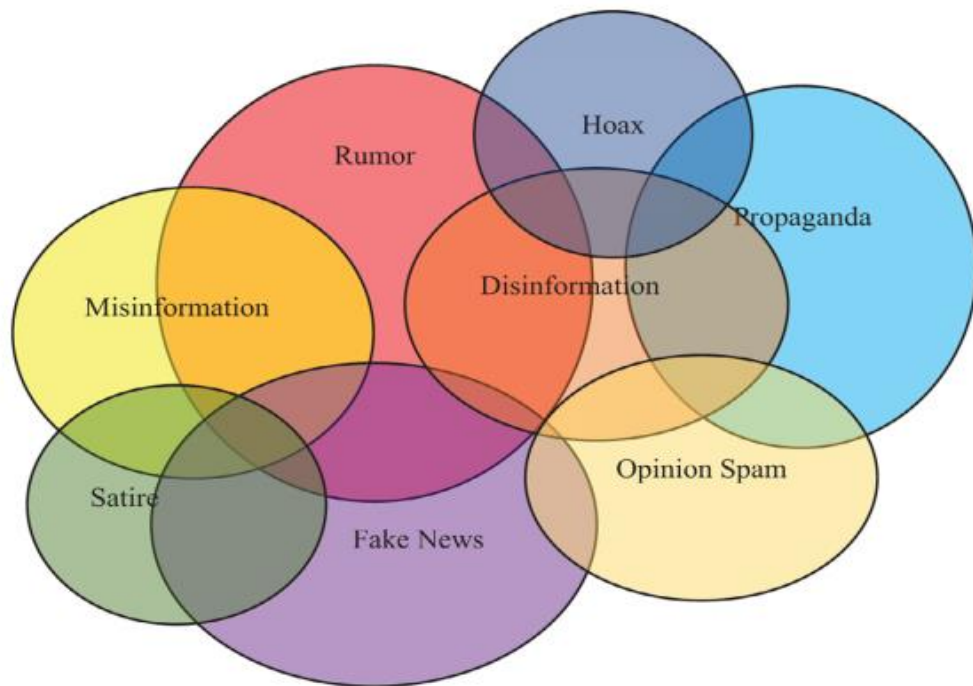
Tabel 2.1. Perbandingan antara konsep terkait pada berita palsu

Konsep	Keaslian	Maksud/Niat	Berita?
<i>Deceptive news</i>	Non-faktual	Menyesatkan	Ya
<i>False news</i>	Non-faktual	Tidak terdefinisi	Ya
<i>Satire news</i>	Tidak Utuh	Menghibur	Ya
<i>Disinformation</i>	Non-faktual	Menyesatkan	Tidak terdefinisi
<i>Misinformation</i>	Non-faktual	Tidak terdefinisi	Tidak terdefinisi
<i>Cherry-picking</i>	Biasanya faktual	Menyesatkan	Tidak terdefinisi
<i>Clickbait</i>	Tidak terdefinisi	Menyesatkan	Tidak terdefinisi
<i>Rumor</i>	Tidak terdefinisi	Tidak terdefinisi	Tidak terdefinisi

Para peneliti sebelumnya juga memperkenalkan beberapa definisi dari aspek-aspek utama informasi palsu yang beredar di internet. Meskipun istilah yang paling umum digunakan oleh kantor media resmi adalah berita palsu dan rumor, tetapi para peneliti juga menganalisis aspek lain yang terkait dengan informasi yang salah. Beberapa kategorisasi sederhana dari berbagai jenis informasi yang salah terlihat pada Gambar 2., dan pada gambar tersebut hanya berfokus pada berita palsu dan rumor (Bondielli and Marcelloni, 2019).



Gambar 2.1. Kategorisasi jenis misinformasi (Bondielli and Marcelloni, 2019)



Gambar 2.2. Diagram berita salah (Meel and Vishwakarma, 2020)

Tabel 2.2. Kategori berita salah di media sosial (Meel and Vishwakarma, 2020)

No	Kategori	Definisi	Dampak
1	<i>Rumor</i>	Konten informasi yang belum diverifikasi tetapi belum tentu salah, boleh jadi bisa benar juga.	Ketidakpastian dan kebingungan tentang fakta
2	<i>Fake News</i>	Informasi palsu yang disebarkan dengan menyamar sebagai berita otentik, biasanya disebarkan melalui portal yang mirip media berita, atau blog, dengan maksud untuk mendapatkan keuntungan secara politik, finansial, atau menambah jumlah pembaca	Dapat merusak citra suatu lembaga, orang, atau masyarakat tertentu, untuk mendapatkan keuntungan finansial atau politik
3	<i>Disinformation</i>	Informasi yang sengaja disebarkan untuk menipu dengan niat tertentu yang telah	Dapat menodai citra pihak lawan, atau untuk

		direncanakan sebelumnya	mempromosikan suatu ide, demi keuntungan finansial
4	<i>Hoax</i>	Konten cerita palsu, terutama melalui lelucon, prank, atau muslihat, yang digunakan untuk menyamarkan kebenaran	Konten kepalsuan yang disebar dapat dianggap sebagai kebenaran atau kenyataan
5	<i>Propaganda</i>	Konten prasangka yang tidak adil dan menipu yang disebar pada komunitas tertentu sesuai dengan strategi yang telah ditentukan sebelumnya, atau untuk mempromosikan agenda politik tertentu	Mendapatkan keuntungan politik, atau keuangan
6	<i>Opinion Spam</i>	Ulasan palsu, atau komentar palsu yang sengaja dibuat bias tentang suatu produk, atau layanan tertentu	Dapat menyesatkan opini pelanggan yang tidak benar
7	<i>Conspiracy theories</i>	Penjelasan tentang suatu peristiwa yang memicu konspirasi oleh aktor-aktor jahat dan berkuasa, seringkali bermotif politik dan berdasarkan prasangka dan bukti yang tidak cukup	Sangat berbahaya, karena dapat menyesatkan persepsi orang dan masyarakat
8	<i>Clickbait</i>	Penggunaan judul berita yang menyesatkan secara sengaja untuk mendorong pengunjung mengklik halaman web tertentu	Untuk memperoleh pendapatan iklan, untuk memicu serangan phishing
9	<i>Misinformation</i>	Konten informasi yang menjadi salah secara tidak sengaja, akibat dari kecerobohan pembuat konten, atau bias kognitif	Penafsiran fakta yang tidak terlalu berbahaya tetapi salah
10	<i>Satire/parody</i>	Konten yang berisi humor dan ironi, tidak ada niat untuk merugikan tetapi berpotensi untuk menipu	Motifnya untuk menghibur tetapi terkadang berpotensi menimbulkan efek buruk

2.2 Kecerdasan Buatan

Kecerdasan buatan (*artificial intelligence*) merupakan salah satu bagian dari ilmu komputer yang mempelajari cara menanamkan suatu kecerdasan kepada suatu mesin atau sistem yang bisa diatur. Kecerdasan buatan dalam konteks ilmiah bisa disebut juga intelegensi artifisial yang dapat membuat mesin dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia. Suatu kecerdasan dapat diartikan sebagai kemampuan dalam memiliki pengetahuan dan pengalaman, atau kemampuan dalam melakukan penalaran, mengambil keputusan, atau melakukan suatu tindakan.

Tabel 2.3. Kategori definisi kecerdasan (buatan) (Bondielli and Marcelloni, 2019)

Thinking Humanly	Thinking Rationally
Acting Humanly	Acting Rationally

Pemodelan kecerdasan buatan saat ini bergantung pada algoritma untuk mencapai hasil yang diinginkan, sehingga kecerdasan buatan dapat dikategorikan sebagai berikut (Bondielli and Marcelloni, 2019);

1. *Thinking humanly*, yaitu bagaimana suatu mesin memiliki kemampuan untuk dapat berpikir seperti manusia.
2. *Acting humanly*, yaitu bagaimana suatu mesin dapat bertindak seperti manusia, memiliki kemampuan memproses bahasa alami, dan merepresentasikan pengetahuan penalaran otomatis.
3. *Thinking rationally*, yaitu bagaimana suatu mesin memiliki kemampuan berpikir rasional, berdasarkan pengetahuan yang ditanamkan, dan bagaimana berinteraksi dengan lingkungan berdasarkan data yang ada.
4. *Acting Rationally*, yaitu bagaimana suatu mesin dapat bertindak secara rasional dengan fokus pada efektivitas dan efisiensi tindakannya.

Perkembangan teknologi komputer dan internet yang semakin pesat turut menyumbang banyak data di internet. Data-data tersebut bahkan sudah banyak yang berlabel sehingga memungkinkan untuk dianalisis. Data dalam bentuk numerik, akan lebih mudah untuk digali dibandingkan dengan data teks. Pendekatan matematis dan statistik sangat memungkinkan untuk menggali data tersebut. Saat ini, pendekatan yang dilakukan para peneliti bukan hanya pendekatan matematis dan statistik, namun sudah menggunakan pendekatan kecerdasan buatan dengan model pembelajaran mesin.

2.3 Data Mining

Data mining merupakan salah satu cara yang digunakan untuk mendapatkan pengetahuan baru dengan memanfaatkan jumlah data yang sangat besar. Beberapa teknik telah dikembangkan dan diimplementasikan untuk mengekstrak pengetahuan dan informasi untuk menemukan pola pengetahuan yang mungkin berguna untuk pengambilan keputusan. Teknik-teknik yang digunakan untuk mengekstrak pengetahuan dalam data mining adalah pengenalan pola, clustering, asosiasi, prediksi dan klasifikasi. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. *Clustering* digunakan untuk pengelompokan data berdasarkan kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap kluster yang lain. Mengintegrasikan metode clustering dengan klasifikasi didapat hasil model yang didapat memiliki akurasi dan *robustness* yang lebih baik jika hanya dilakukan dengan metode klasifikasi saja (Defiyanti and Jajuli, 2015).

2.3.1 Machine Learning

Machine learning adalah cabang dari bidang ilmu kecerdasan buatan dan menjadi salah satu cara yang dapat digunakan dalam kecerdasan buatan untuk mencapai hasil yang maksimal. *Machine learning* mengandalkan bekerja dengan kumpulan data

kecil hingga besar dengan memeriksa dan membandingkan data tersebut untuk menemukan pola umum dan mengeksplorasi perbedaannya. Salah satu penerapan yang umum dari *machine learning* adalah prediksi hasil berdasarkan data yang ada. *Machine learning* mampu mempelajari pola dari kumpulan data yang ada, dan kemudian menerapkannya ke kumpulan data yang tidak diketahui untuk memprediksi hasilnya. Teknik klasifikasi adalah teknik yang sering digunakan dalam *machine learning* untuk melakukan suatu proses prediksi hasil (Latha and Jeeva, 2019).

2.3.2 Deep Learning

Deep learning adalah bagian dari kecerdasan buatan dan *machine learning*, yang merupakan pengembangan dari *neural network multiple layer* untuk memberikan ketepatan tugas seperti deteksi objek, pengenalan suara, terjemahan bahasa dan lain-lain. *Deep learning* berbeda dari teknik *machine learning* yang tradisional, karena *deep learning* secara otomatis melakukan representasi dari data seperti gambar, video atau text tanpa memperkenalkan aturan kode atau pengetahuan domain manusia. *Deep learning* dikembangkan pada tahun 1950 namun baru tahun 1990 dapat di aplikasikan dengan sukses. Learning algoritma yang digunakan sekarang pada task yang kompleks hampir sama seperti learning algoritma yang digunakan untuk menyelesaikan masalah permainan pada tahun 1980, meskipun model algoritma yang digunakan berubah menjadi training yang sederhana dari arsitektur *deep learning*. Hal yang penting pada pengembangan model yang sekarang adalah kita dapat mendukung dengan sumber daya yang dibutuhkan agar menjadi sukses (Yanuar, 2018).

Deep learning merupakan teknik modern terkini untuk pemrosesan gambar dan analisis data, dengan hasil yang menjanjikan dan potensinya sangat besar. Telah dilakukan survei terhadap 40 penelitian yang menggunakan teknik *deep learning*. Selain itu, telah dilakukan juga perbandingan *deep learning* dengan teknik populer lainnya, sehubungan dengan perbedaan kinerja teknik klasifikasi atau analisis regresi. Hasilnya menunjukkan bahwa *deep learning* memberikan akurasi tinggi, mengungguli teknik pemrosesan gambar yang umum digunakan (Kamilaris and Prenafeta-Boldú, 2018).

2.4 Penelitian Terkait

Melalui telaah penelitian terkait yang sudah dilakukan sebelumnya, peneliti dapat mengembangkan sebuah metode dan pendekatan baru yang bisa melengkapi apa yang belum dikerjakan oleh para peneliti pendahulu.

Tabel 2.4. Telaah penelitian terkait

No	Peneliti, Tahun	Judul Penelitian	Metode	Hasil Penelitian	Keterbatasan
1	Nayoga, B. P., Adipradana, R., Suryadi, R., & Suhartono, D., 2021	Hoax Analyzer for Indonesian News Using Deep Learning Models	Membandingkan 7 model klasifikasi, yaitu: LSTM, BI-LSTM, GRU, BI-GRU, Naïve Bayes, SVM, 1D-CNN untuk mendeteksi berita palsu berbahasa Indonesia. Sumber data sebanyak 1.000 set data, yaitu: 600 data dari https://data.mendeley.com/datasets/p3hfgr5j3m/1 dalam bahasa Indonesian, ditambah 128 berita valid dan 223 berita palsu dari https://github.com/pierobeat/Hoax-News-Classification , ditambah lagi secara manual 49 berita palsu dari kompas.com .	1D-CNN mencapai hasil terbaik, yaitu akurasi 97.9%, presisi 97.9%, dan recall 97.8%. Model Deep Neural Network (DNN) secara umum lebih baik dibandingkan model klasifikasi konvensional.	GRU dan BI-GRU tidak kompatibel dengan Dropout, serta belum teratasi masalah kehilangan NaN di GRU.
2	Nasir, J. A., Khan, O. S., & Varlamis, I. (2021)	Fake news detection: A hybrid CNN-RNN based deep learning approach	Membangun model kombinasi CNN-RNN ISOT Fake News Dataset, 45.000 artikel. Dataset: https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php FA-KES Dataset, 804 artikel. Dataset: https://zenodo.org/record/2607278#.YQPD6I4zbKt	Model hybrid CNN-RNN dapat meningkatkan performa deteksi berita palsu. Hasil eksperimen : akurasi 100% pada kumpulan data ISOT. akurasi 60% pada data FA-KES.	Akurasi masih belum maksimal yaitu: 60% pada kumpulan data FA-KES, 804 artikel.

No	Peneliti, Tahun	Judul Penelitian	Metode	Hasil Penelitian	Keterbatasan
3	Utami, M. P., Nurhayati, O. D., & Warsito, B., 2020	Hoax Information Detection System Using Apriori Algorithm and Random Forest Algorithm in Twitter	Mengimplementasikan algoritma Apriori dan Random Forest. Sumber data sebanyak 5.000 pada periode 2016-2020 dari Turnbackhoax.id yang telah disusun klasifikasi kelas hoax dan non-hoax, dengan mengajukan permintaan.	Tingkat akurasi, presisi, dan recall, dari sistem yang dibangun mencapai 100% dengan nilai support minimal 23.	Seharusnya terlebih dahulu melakukan uji coba yang diperlukan dengan pendekatan semi-supervised learning.
4	Deepak, S., & Chitturi, B., 2020	Deep neural approach to Fake-News identification	Secara umum menggunakan model FNN dan LSTM yang ditambahkan mined features pada dataset. Sumber data dari dataset berita palsu George McIntires yang mencakup berita palsu dan berita asli dalam rasio 1:1, sebanyak 10.558 data dengan 4 kolom.	Model FNN dan LSTM yang dibangun mampu memberikan peningkatan yang signifikan dalam presisi, recall, akurasi dan skor F1 ketika LSTM, dalam kombinasi dengan representasi word2vec.	Perlu diperluas lebih lanjut dengan membuat model ensemble (LSTM/SVM) atau (LSTM/FNN).
5	Konkobo, P. M., Zhang, R., Huang, S., Minoungou, T. T., Ouedraogo, J. A., & Li, L., 2020	A Deep Learning Model for Early Detection of Fake News on Social Media	Model yang diusulkan terdiri dari 4 blok utama, yaitu: blok ekstraksi opini, blok penilaian kredibilitas pengguna, blok konstruksi jaringan pengguna dan pengklasifikasi. Membandingkan klasifikasi Naive Bayes, Decision Tree, SVM, PPC_RNN+CNN, dan SSLNews. Sumber data dari Politifact dan Gossipcop (400 berita palsu, dan 400 berita asli). Dataset: https://github.com/KaiDMML/FakeNewsNet/tree/master/dataset	Model dengan 4 blok mencapai akurasi 71,10% di Politifact dan 68,07% di Gossipcop.	Perlu untuk membangun dataset lintas bahasa dan menganalisis dampak bahasa pada klasifikasi berita palsu.

No	Peneliti, Tahun	Judul Penelitian	Metode	Hasil Penelitian	Keterbatasan
6	Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. 2019	Supervised Learning for Fake News Detection	<p>Membandingkan 5 model, yaitu: k-Nearest Neighbors (KNN), Naive Bayes (NB), Random Forests (RF), Support Vector Machine (SVM) dengan kernel RBF, dan XGBoost (XGB).</p> <p>Sumber data sebanyak 2.282 artikel berita dari BuzzFeed terkait dengan pemilihan presiden AS 2016 yang telah diberi label oleh jurnalis dan diperkaya dengan komentar pengguna Facebook. Dataset: https://github.com/gsantia/BuzzFace</p>	<p>AUC sangat relevan untuk deteksi berita palsu karena ambang keputusan dapat digunakan untuk mengontrol tradeoff antara hasil benar dengan rasio false positive.</p> <p>Hasil AUC terbaik diperoleh RF dan XGB, secara statistik, RF=0,85($\pm 0,007$) dan XGB=0,86($\pm 0,006$).</p>	Salah mengklasifikasi masih sekitar 40%, karena beberapa berita benar diposting oleh media yang dihosting di alamat IP yang sama dengan sumber berita palsu, yang masuk daftar hitam.
7	Benamira, A., Devillers, B., Lesot, E., Ray, A. K., Saadi, M., & Malliaros, F. D., 2019	Semi-Supervised Learning and Graph Neural Networks for Fake News Detection	<p>Secara umum terdapat 2 langkah, yaitu:</p> <ol style="list-style-type: none"> 1) Konstruksi Graph: mencari kesamaan antar artikel, menggunakan KNN berdasarkan jarak Euclidean. 2) Klasifikasi: melakukan klasifikasi atas kesamaan graph menggunakan 2 metode, yaitu: Graph Convolutional Networks (GCN) dan Attention Graph Neural Network (AGNN). <p>Sumber data dari Horne 2017 Fake News Data yang terdiri dari 150 artikel berlabel, 75 palsu, dan 75 asli. Dataset: https://github.com/rpitrust/fake-newsdata1</p>	Nearest-Neighbour Graph sederhana di antara artikel disertai dengan Graph Neural Networks memberikan hasil yang baik, untuk metode deteksi berbasis konten semi-supervised.	Perlu lebih banyak pengujian kinerja pada kumpulan data yang lebih besar serta multi-label.
8	Bharadwaj, P., & Shao, Z., 2019	Fake News Detection with Semantic Features and Text Mining	Membandingkan 3 model, yaitu: Naive Bayes, Random Forest, dan RNN, masing-masing dengan 6 kelompok fitur.	<p>Random forest mencapai akurasi 95,66%.</p> <p>Ini menyiratkan</p>	Perlu dilakukan kombinasi dengan isyarat linguistik dan meta data

No	Peneliti, Tahun	Judul Penelitian	Metode	Hasil Penelitian	Keterbatasan
			Sumber data dari 6.256 artikel, 50% artikel berlabel palsu dan sisanya asli dari kaggle.com. Dataset: https://www.kaggle.com/c/fake-news/data	bahwa fitur semantik berguna untuk deteksi berita palsu. Sebagai langkah selanjutnya	lainnya untuk meningkatkan kinerja deteksi.

BAB III

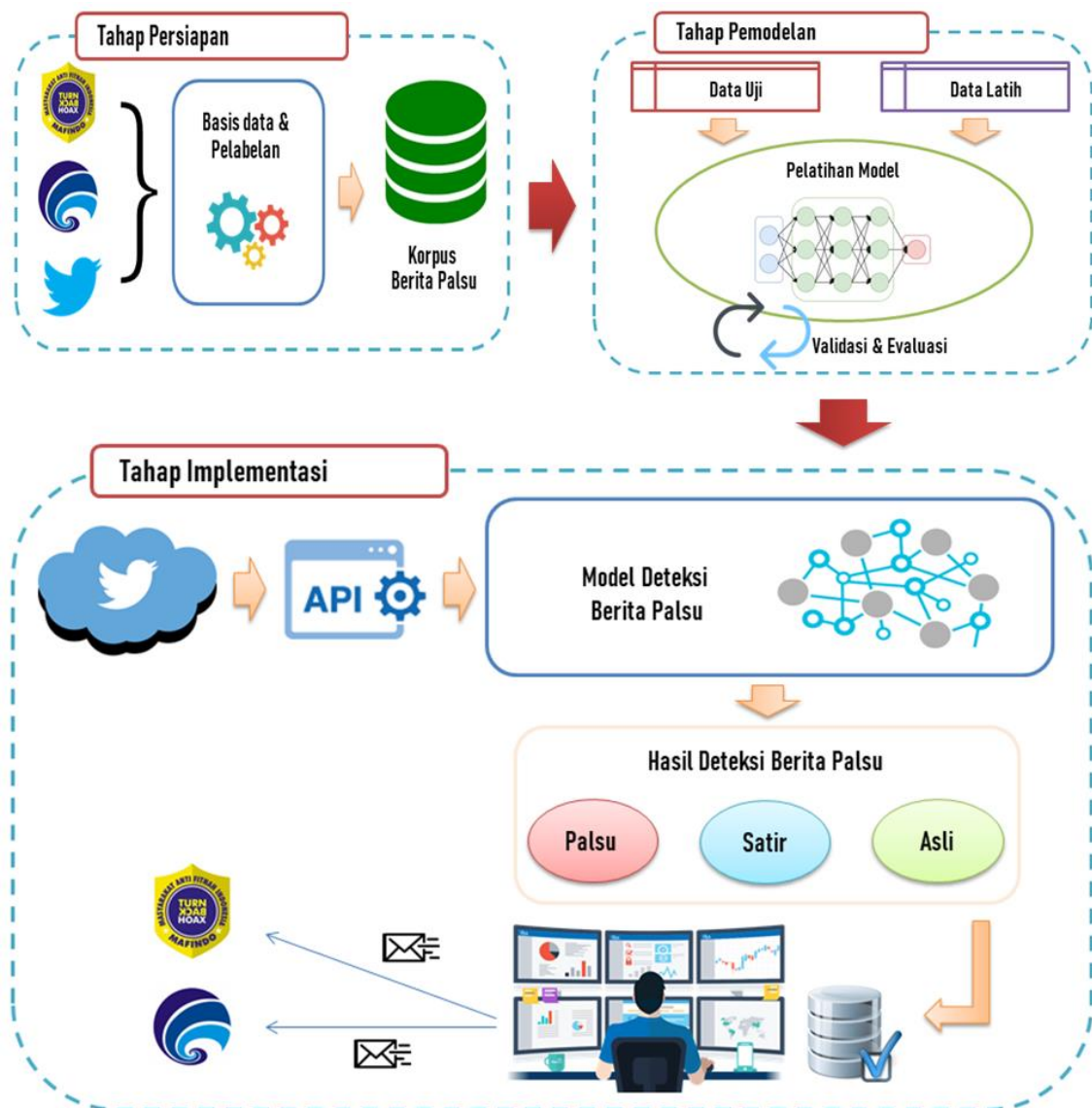
METODE PENELITIAN

3.1 Gambaran Umum Penelitian

Penelitian deteksi dini berita palsu ini menggunakan jenis konsep penelitian dan pengembangan (*research and development*). Metode penelitian dan pengembangan adalah metode penelitian yang digunakan untuk menghasilkan produk tertentu, dan menguji keefektifan produk tersebut. Penelitian jenis ini tujuannya adalah mengembangkan produk berdasarkan uji coba untuk kemudian direvisi sampai menghasilkan produk yang layak pakai.

Langkah-langkah penelitian yang dilakukan secara umum terdiri dari tiga langkah besar, yaitu 1) tahap persiapan, 2) tahap pemodelan, 3) tahap implementasi. Tahap persiapan dimulai dengan proses mengumpulkan data, proses pelabelan data, dan pembuatan korpus berita palsu. Tahap pemodelan dimulai dengan proses ekstraksi fitur, penyusunan data latih dan data uji, pengkodean algoritma, dan pembentukan model setelah lolos validasi dan evaluasi. Tahapan implementasi dimulai dengan memonitor media sosial secara berkala dan melakukan deteksi dini pada percakapan yang berpotensi kuat. Model deteksi dini berita palsu kemudian akan menghasilkan prediksi tiga kelas pengelompokkan, yaitu palsu, satir, atau asli. Hasil prediksi tersebut

kemudian disimpan didalam basis data dan berita yang terdeteksi palsu diteruskan kepada pihak-pihak yang berkepentingan, seperti kemkominfo dan komunitas Mafindo.



Gambar 3.1. Skema Umum Penelitian

3.1.1 Pengumpulan Data dan Identifikasi Berita Palsu

Langkah pertama dalam proses riset dan pengembangan adalah mengumpulkan data dari media sosial yang merupakan sumber data yang sangat besar dan memiliki perubahan yang kontinu dan sangat singkat rentang waktu perubahannya juga memiliki relevansi yang cukup besar untuk berbagai bidang. Dalam 60 detik terdapat jutaan perubahan yang terjadi dalam media sosial, sehingga sangat banyak hal yang menjadi *issue* dari berbagai bidang, dari mulai ekonomi, politik, agama hingga olahraga.

3.1.2 Analisis Algoritma dan Studi Pustaka

Setelah identifikasi berita palsu pada media sosial, langkah selanjutnya adalah melakukan analisis pola dan kriteria berita palsu di media sosial. Beberapa di antaranya adalah dengan melakukan unggah ulang foto atau vidio yang sudah ada sebelumnya, lalu dihubung-hubungkan dengan kejadian atau peristiwa baru.

3.1.3 Perancangan Medel Deteksi Dini Berita Palsu

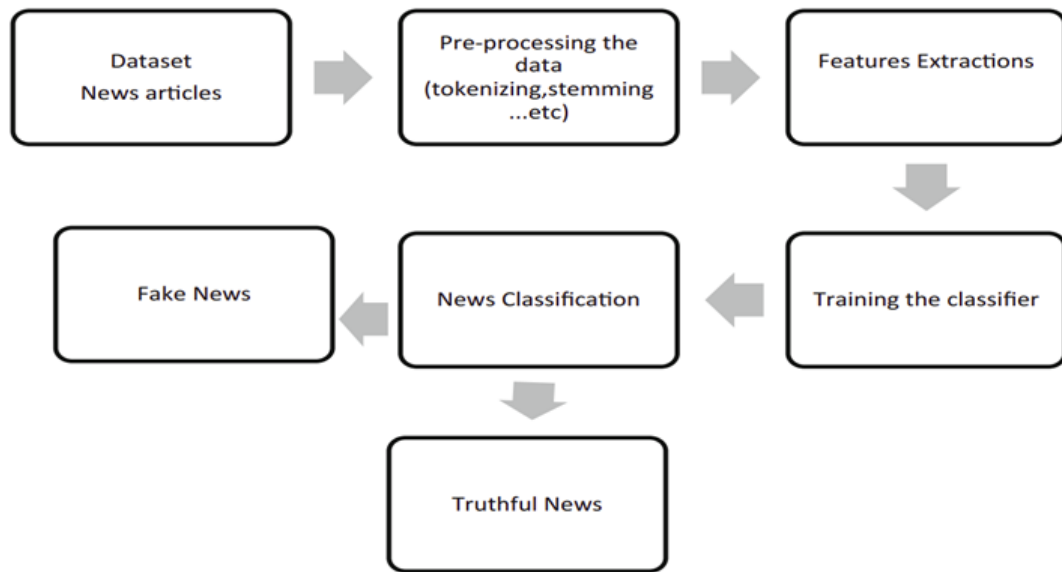
Langkah selanjutnya adalah metode deteksi dini menggunakan deep learning. Ada banyak metode untuk mengetahui apakah suatu konten pada media sosial terindikasi berita palsu atau bukan. Metode-metode tersebut dapat dikelompokkan menjadi 2 bagian yaitu metode manual (non-komputer) dan metode berbasis komputer. Metode yang tergolong manual adalah metode berpikir kritis, sedangkan yang termasuk metode berbasis komputer adalah metode *machine learning*. Algoritma *deep learning* menerima input berupa kata dan frasa yang terdapat pada media sosial dan dapat melakukan proses klasifikasi secara *supervised learning*. Metode klasifikasi *supervised learning* selain memiliki tahapan proses lebih singkat, juga semakin mendekati 100 persen otomatis, sehingga peran manusia dalam proses klasifikasi secara keseluruhan berkurang.

3.1.4 Menguji Akurasi Deteksi Dini Berita Palsu

Langkah selanjutnya adalah melakukan pengujian perangkat lunak dengan metode *white box* dan *black box*. Pengujian *white box* adalah pengujian yang didasarkan pada pengecekan terhadap detail perancangan, menggunakan struktur kontrol dari desain program secara prosedural untuk membagi pengujian ke dalam beberapa kasus pengujian. Pengujian *black box* adalah pengujian yang dilakukan hanya mengamati hasil eksekusi melalui data uji dan memeriksa fungsional dari perangkat lunak.

3.2 Skema Klasifikasi Berita Palsu

Klasifikasi berita palsu di media sosial terhadap topik tertentu, secara konseptual menggunakan metode *Natural Language Processing* (NLP), analisis teks, dan beberapa bagian komputasi untuk melihat pola suatu kalimat untuk dapat menentukan berita palsu atau berita asli. Terdapat 5 tahapan dalam proses klasifikasi berita palsu di sosial media pada penelitian ini, yaitu: tahap pertama, setelah memiliki kunci Twitter API, data cuitan Twitter dapat diambil dan dikumpulkan di dalam satu dataset. Tahapan kedua disebut proses *tokenization* yaitu metode memecah teks menjadi token-token yang berurutan. Proses *tokenization* sederhana biasanya hanya memecah teks dengan *whitespace* sebagai pembagi, lalu mengubahnya menjadi huruf kecil supaya seragam. Kemudian tahap ketiga *stop-words removing*, yaitu proses menghilangkan sejumlah kelas kata penghubung ataupun yang jumlahnya banyak namun tidak mempengaruhi konten dokumen secara keseluruhan, sebagai bagian dari *pre-processing*. Dilanjutkan tahap keempat yang disebut *stemming*, yaitu metode memetakan kata ke bentuk dasarnya, namun bentuk dasar tersebut tidak berarti sama dengan root word. Tahap kelima, untuk dapat menghasilkan kelas berita palsu atau bukan menggunakan *supervised learning*.



Gambar 3.2. Proses klasifikasi berita palsu

Daftar Pustaka

- Ahmad, A. R. and Murad, H. R. (2020) ‘The Impact of Social Media on Panic During The COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study’, *Journal of Medical Internet Research*, 22(5).
- Asyik, R. (2019) *Inilah Beda Misinformasi, Disinformasi, dan Malinformasi*, *ayobandung.com*. Available at: <https://ayobandung.com/read/2019/02/01/44283/inilah-beda-misinformasi-disinformasi-dan-malinformasi>.
- Benamira, A. *et al.* (2019) ‘Semi-supervised learning and graph neural networks for fake news detection’, in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019*. doi: 10.1145/3341161.3342958.
- Bharadwaj, P. and Shao, Z. (2019) ‘Fake News Detection with Semantic Features and Text Mining’, *International Journal on Natural Language Computing*, 8(3). doi: 10.5121/ijnlc.2019.8302.
- Bondielli, A. and Marcelloni, F. (2019) ‘A survey on fake news and rumour detection techniques’, *Information Sciences*, 497. doi: 10.1016/j.ins.2019.05.035.
- Cao, X. and Yu, L. (2019) ‘Exploring The Influence Of Excessive Social Media Use At Work: A Three-Dimension Usage Perspective’, *International Journal of Information Management*, 46(1), pp. 83–92.
- Deepak, S. and Chitturi, B. (2020) ‘Deep neural approach to Fake-News identification’, in *Procedia Computer Science*. doi: 10.1016/j.procs.2020.03.276.
- Defiyanti, S. and Jajuli, M. (2015) ‘Integrasi Metode Klasifikasi Dan Clustering Dalam

-
- Data Mining’, *Konferensi Nasional Informatika (KNIF)*, 10(15), pp. 39–44.
- Gerintya, S. (2018) *Hoaks dan Bahaya Rendahnya Kepercayaan Terhadap Media*. Available at: <https://tirto.id/hoaks-dan-bahaya-rendahnya-kepercayaan-terhadap-media-cKAX> (Accessed: 28 July 2021).
- Hidaya, N. *et al.* (2019) *Pengaruh Media Sosial Terhadap Penyebaran Hoax Oleh Digital Native*. Makassar.
- Kamilaris, A. and Prenafeta-Boldú, F. X. (2018) ‘Deep Learning In Agriculture: A Survey’, *Computers and Electronics in Agriculture*. doi: 10.1016/j.compag.2018.02.016.
- Konkobo, P. M. *et al.* (2020) ‘A Deep Learning Model for Early Detection of Fake News on Social Media*’, in *Proceedings of 2020 7th IEEE International Conference on Behavioural and Social Computing, BESC 2020*. doi: 10.1109/BESC51023.2020.9348311.
- Lararenjana, E. (2020) *Mengenal Arti Hoax Atau Berita Bohong, Ketahui Jenis dan Ciri-Cirinya*. Available at: <https://www.merdeka.com/jatim/mengenal-arti-hoax-atau-berita-bohong-dan-cara-tepat-menyikapinya-kln.html?page> (Accessed: 28 July 2021).
- Latha, C. B. C. and Jeeva, S. C. (2019) ‘Improving The Accuracy Of Prediction Of Heart Disease Risk Based On Ensemble Classification Techniques’, *Informatics in Medicine Unlocked*, 16. doi: 10.1016/j.imu.2019.100203.
- Meel, P. and Vishwakarma, D. K. (2020) ‘Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities’, *Expert Systems with Applications*, 153, p. 112986. doi: 10.1016/J.ESWA.2019.112986.
- Nasir, J. A., Khan, O. S. and Varlamis, I. (2021) ‘Fake news detection: A hybrid CNN-RNN based deep learning approach’, *International Journal of Information Management Data Insights*, 1(1), p. 100007. doi: 10.1016/j.jjime.2020.100007.
- Nayoga, B. P. *et al.* (2021) ‘Hoax Analyzer for Indonesian News Using Deep Learning Models’, *Procedia Computer Science*, 179, pp. 704–712. doi: 10.1016/J.PROCS.2021.01.059.

- Ramadhani, M. R. and Pratama, A. R. I. (2020) ‘Analisis Kesadaran Cyber Security Pada Pengguna Media Sosial Di Indonesia’, *AUTOMATA*, 1(2). Available at: <https://journal.uui.ac.id/AUTOMATA/article/download/15426/10219>.
- Reis, J. C. S. *et al.* (2019) ‘Supervised Learning for Fake News Detection’, *IEEE Intelligent Systems*, 34(2), pp. 76–81. doi: 10.1109/MIS.2019.2899143.
- Reuter, C. and Kaufhold, M. A. (2018) ‘Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics’, *Journal of Contingencies and Crisis Management*, 26(1). doi: 10.1111/1468-5973.12196.
- Utami, M. P., Nurhayati, O. D. and Warsito, B. (2020) ‘Hoax Information Detection System Using Apriori Algorithm and Random Forest Algorithm in Twitter’, in *6th International Conference on Interactive Digital Media, ICIDM 2020*. doi: 10.1109/ICIDM51048.2020.9339648.
- Yanuar, A. (2018) *Pengenalan Deep Learning*, *ugm.ac.id*. Available at: <https://machinelearning.mipa.ugm.ac.id/2018/06/10/pengenalan-deep-learning> (Accessed: 28 July 2021).
- Zhou, X. and Zafarani, R. (2020) ‘A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities’, *ACM Computing Surveys*, 53(5), pp. 1–40. doi: 10.1145/3395046.