

FTI/IA/SKRIPSI/2768/2023

PENERAPAN MODEL KLASIFIKASI MACHINE LEARNING UNTUK DETEKSI BERITA PALSU

YOGHI MUHAMAD SYAFPUTRA | 56419696 | 2023 | S1 - SKRIPSI | Informatika - Teknologi Industri

Pembimbing: Dr. Asep Juarna | Promotor:

Table Of Contents

3	STRAKSI (BAHASA INDONESIA)	KATA PENGANTAR	DAFTAR ISI	DAFTAR TABEL	DAFTAR GAMBAR	E
---	----------------------------	----------------	------------	--------------	---------------	---

ISIT-

ISIT-

ISIT-

ISIT-

ISIT-

ISIT-

ISIT-

ISIT-

ISIT-

BAB II

LANDASAN TEORI

2.1 Berita Palsu (hoax)

Berita palsu (*hoax*) adalah artikel berita yang sengaja dibuat untuk menyesatkan pembaca, misalnya dengan memposting suatu berita yang direkayasa tentang seseorang atau mengunduh video berisi suatu peristiwa yang dibesar-besarkan, menyebabkan munculnya kebencian dan kemarahan publik. Sindikat kejahatan dunia maya (*ciber crimes*) yang merupakan kejahatan yang ditimbulkan oleh perkembangan dan kemajuan teknologi , dalam beberapa tahun ini menjadi perhatian publik. Ada dua motivasi utama yang menyebabkan beredarnya berita palsu. Pertama adalah uang, artikel berita seakan menjadi virus di media sosial yang dapat menarik pendapatan iklan yang signifikan saat pengguna mengeklik situs aslinya. Motivasi kedua adalah ideologis. Beberapa penyedia berita palsu berusaha untuk memajukan kandidat yang mereka sukai. (Kurniawati, Y. R. 2020)

2.2 Python

Python merupakan sebuah bahasa pemrograman tingkat tinggi yang bisa melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode *Object Oriented Programming* dan juga menggunakan semantik dinamis untuk memberikan tingkat keterbacaan syntax. Sebagai bahasa pemrograman tingkat tinggi, python dapat dipelajari dengan mudah karena telah dilengkapi dengan manajemen memori otomatis python merupakan salah satu Bahasa pemrograman yang cukup populer digunakan dalam pembuatan aplikasi *artificial intelegence*. (Rahmadhika, M. K., & Thantawi, A. M., 2021)

2.3 Artificial Intelligence

Artificial Intelligence atau kecerdasan buatan adalah sistem komputer yang mampu melakukan tugas-tugas yang biasanya membutuhkan kecerdasan manusia. Teknologi ini dapat membuat keputusan dengan cara menganalisis dan menggunakan data yang tersedia di dalam sistem. Proses yang terjadi dalam *Artificial Intelligence* mencakup *learning*, *reasoning*, dan *self-correction*. Proses ini mirip dengan manusia yang melakukan analisis sebelum memberikan keputusan. Menurut John Mc Carthy, 1956, *Artificial Intelligence* adalah untuk mengetahui dan memodelkan proses-proses berpikir manusia dan mendesain mesin agar dapat menirukan perilaku manusia. Cerdas, berarti memiliki pengetahuan ditambah pengalaman, penalaran (bagaimana membuat keputusan dan mengambil tindakan), moral yang baik. (Lubis, M. S. Y. 2021)

2.4 Machine Learning

Machine Learning (ML) cabang dari kecerdasan buatan yang memungkinkan komputer atau sistem untuk belajar secara mandiri dari data dan pengalaman tanpa secara eksplisit diprogram yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Konsep utama di balik *Machine Learning* adalah memberikan kemampuan komputer untuk mengenali pola dalam data dan membuat keputusan atau prediksi berdasarkan pola-pola tersebut. Dalam *Machine Learning*, model atau algoritma diberi data sebagai input dan melalui proses pembelajaran, model tersebut menghasilkan output yang diinginkan. Proses pembelajaran terjadi melalui iterasi dan penyesuaian berulang, di mana model secara otomatis menemukan pola-pola dalam data dan meningkatkan kinerjanya seiring dengan bertambahnya pengalaman dan informasi Terdapat beberapa model pada *Machine Learning* seperti klasifikasi dan prediksi . Ciri khas dari Machine Learning adalah adanya proses pelatihan, pembelajaran, atau training. Oleh karena itu, *Machine Learning* membutuhkan data untuk dipelajari yang disebut sebagai data training. Klasifikasi adalah metode dalam *Machine Learning* yang digunakan oleh mesin untuk memilah atau mengklasifikasikan obyek berdasarkan ciri tertentu sebagaimana manusia mencoba

membedakan benda satu dengan yang lain. Sedangkan prediksi atau regresi digunakan oleh mesin untuk menerka keluaran dari suatu data masukan berdasarkan data yang sudah dipelajari dalam training. Metode Machine Learning yang paling populer yaitu Sistem Pengambil Keputusan, *Support Vector Machine* (SVM) dan *Neural Network*. (Wibisono, L. 2023).

2.5 Preprocessing Teks

Preprocessing teks adalah serangkaian langkah atau proses yang dilakukan pada teks mentah sebelum dilakukan analisis lebih lanjut. Tujuannya adalah untuk membersihkan, mengubah, dan mempersiapkan teks agar siap digunakan dalam proses pemrosesan bahasa alami (NLP) dan metode klasifikasi. *Preprocessing* teks bertujuan untuk meningkatkan kualitas dan akurasi analisis teks dengan menghilangkan *noise*, mengatasi variasi penulisan, dan menghasilkan representasi teks yang lebih informatif.

Proses *preprocessing* teks melibatkan berbagai langkah yang mungkin termasuk penghapusan karakter khusus, konversi ke huruf kecil, penghapusan tautan atau URL, penghapusan elemen HTML, tokenisasi, penghapusan *stopwords*, *stemming*, dan *lemmatization*. Setiap langkah memiliki tujuan tertentu dan dapat disesuaikan sesuai dengan karakteristik data dan tujuan analisis.

Hasil dari *preprocessing* teks adalah teks yang telah "bersih" dan siap digunakan dalam tahap selanjutnya, seperti vektorisasi dan penerapan metode klasifikasi. *Preprocessing* teks yang baik dapat mempengaruhi performa model klasifikasi dengan meminimalkan *noise* dan menghasilkan representasi teks yang lebih relevan.

Pada penelitian ini, *preprocessing* teks akan diterapkan pada judul berita sebelum dilakukan penerapan metode klasifikasi untuk deteksi berita palsu. Langkah-langkah *preprocessing* teks tersebut akan membantu meningkatkan akurasi dan kualitas analisis deteksi berita palsu, perubahan yang dilakukan diantaranya :

1. Konversi ke Huruf Kecil (*Lowercasing*): Mengubah semua karakter dalam teks menjadi huruf kecil untuk menghindari variasi dalam penulisan huruf besar dan huruf kecil yang seharusnya memiliki makna yang sama.
2. Penghapusan Karakter Khusus (*Special Character Removal*): Menghapus karakter khusus seperti tanda baca, simbol, dan karakter non-alfanumerik dari teks menggunakan ekspresi reguler.
3. Penghapusan URL dan Tautan (*URL and Link Removal*): Menghapus URL dan tautan dari teks menggunakan ekspresi reguler untuk menghilangkan informasi yang tidak relevan.
4. Penghapusan HTML Tags (*HTML Tag Removal*): Menghapus elemen-elemen HTML dari teks untuk memastikan hanya teks murni yang digunakan dalam analisis.
5. **Tokenisasi**: Memecah teks menjadi kata-kata individual (token) menggunakan spasi sebagai pemisah. Langkah ini membantu dalam analisis kata lebih lanjut.
6. Penghapusan Stopword (*Stopword Removal*): Menghapus stopwords (kata-kata umum yang tidak memberikan kontribusi signifikan) dari teks, seperti "the", "is", "and", dll.
7. Stemming atau Lemmatization: Menerapkan stemming atau lemmatization untuk mengubah kata-kata menjadi bentuk dasar atau akar kata. Ini dilakukan untuk mengurangi variasi kata yang memiliki makna yang sama.

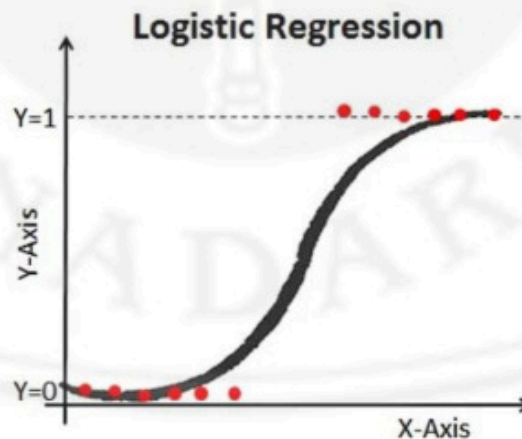
2.6 Metode Klasifikasi

Metode klasifikasi adalah pendekatan komputasional yang digunakan untuk mengelompokkan atau mengkategorikan data ke dalam kelas-kelas yang telah ditentukan. Dalam konteks deteksi berita palsu (*fake news*), metode klasifikasi dapat digunakan untuk memprediksi apakah sebuah judul berita termasuk ke dalam kategori berita palsu atau bukan. Berikut merupakan metode klasifikasi yang digunakan dalam penelitian ini adalah sebagai berikut :

2.6.1 Logistic Regression

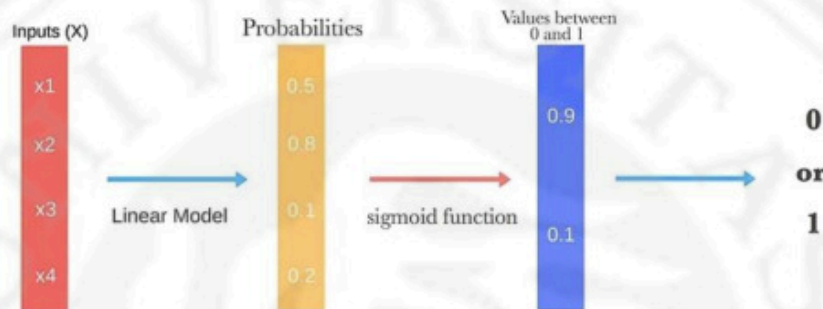
Logistic regression adalah teknik analisis data dalam statistika yang dirancang untuk mengetahui hubungan antar variabel, dimana variabel respon bersifat kategorikal, baik nominal maupun ordinal, dan variabel penjelas bersifat kategoris atau kontinu. Logistic Regression sebenarnya digunakan untuk masalah klasifikasi. Metode ini cocok untuk kasus di mana variabel target adalah biner (dua kelas). *Logistic Regression* menghasilkan model berbasis probabilitas untuk memprediksi probabilitas suatu sampel masuk ke dalam kelas tertentu. Model ini menggunakan fungsi logistik (sigmoid) untuk menghubungkan variabel input dengan variabel target. (Novantika, A., & Sugiman, S. 2022)

Logistic Regression merupakan sebuah model yang digunakan untuk melakukan prediksi apakah sesuatu bernilai benar atau salah (0 atau 1). Apabila pada *linear regression* garis yang terbentuk adalah garis lurus, tetapi pada *logistic regression* garis yang dibentuk mirip dengan huruf “S” antara titik 0 sampai 1. pada gambar 2.1 dapat dilihat bahwa garis yang dibentuk oleh logistic regression memberitahu kita suatu peluang dari sesuatu dengan presentasi 0 (tidak terjadi/tidak tepat) — 0.5 (titik antara) dan 1 (terjadi/tepat).



Gambar 2. 1 Logistic Regression

Logistic regression akan mengukur hubungan antara variabel target (yang ingin diprediksi) dan variabel input (fitur yang digunakan) dengan fungsi logistik. Probabilitas akan dihitung menggunakan fungsi sigmoid untuk mengubah nilai-nilai tadi menjadi 0 atau 1. Dapat dilihat pada gambar 2.2 .(Giodo, M. 2020).



Gambar 2. 2 proses logistic regression

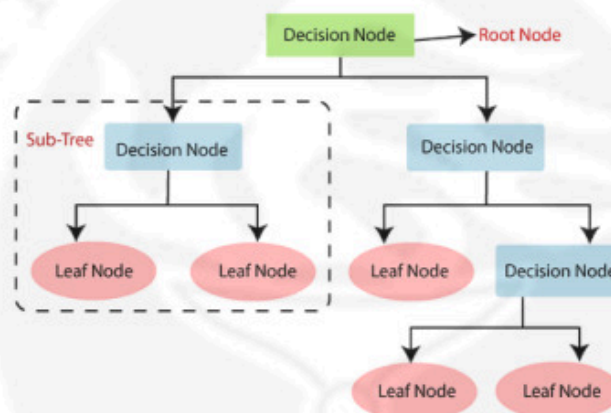
2.6.2 Decision Tree Classification

Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep/kelas data sehingga dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dimana model ini berupa aturan "jikamaka", berupa decision tree, formula matematis atau neural network. Proses *classification* dibagi menjadi 2 (dua) fase yaitu fase learning dan test, pada fase learning sebagian data yang telah diketahui kelas datanya dipergunakan untuk membentuk model perkiraan, yang selanjutnya masuk ke dalam fase test dimana model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Dan apabila akurasinya mencukupi, model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

metode klasifikasi yang membangun struktur pohon keputusan untuk mengambil keputusan berdasarkan serangkaian aturan yang didefinisikan. Pada setiap simpul (node) dalam pohon keputusan, suatu keputusan diambil berdasarkan

fitur-fitur input. Keputusan-keputusan ini membawa kita melalui cabang-cabang pohon hingga mencapai daun (leaf) yang mewakili kelas prediksi. Decision Tree dapat dengan mudah diinterpretasi dan visualisasi, sehingga sangat berguna untuk analisis.

Tujuan penggunaan *Decision tree* adalah untuk membuat training model yang dapat digunakan untuk memprediksi kelas atau nilai variabel target dengan mempelajari aturan keputusan sederhana yang disimpulkan dari data sebelumnya (data training).



Gambar 2. 3 Diagram Decision tree

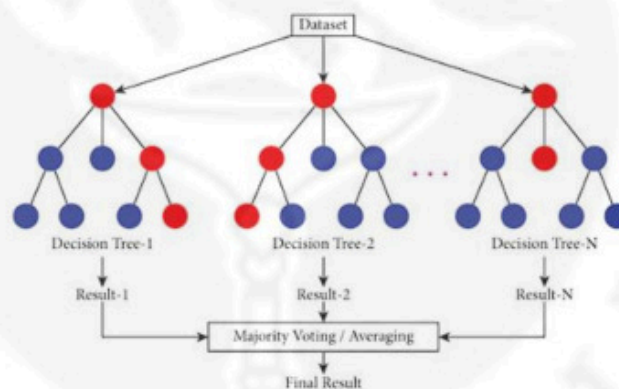
2.6.3 Gradient Boosting Classifier

Gradient Boosting adalah teknik ensemble di mana beberapa model lemah (biasanya pohon keputusan yang disebut "*weak learners*") digabungkan untuk menghasilkan model yang lebih kuat. Dalam setiap iterasi, model mencoba memperbaiki kesalahan model sebelumnya. *Gradient Boosting Classifier* mengkombinasikan prediksi dari beberapa pohon keputusan untuk membuat prediksi yang akurat.

2.6.4 Random Forest Classifier

Random decision forest atau pohon keputusan adalah model yang memperhitungkan pentingnya setiap variabel dalam dataset untuk membagi jalur keputusan menjadi dua dengan mengelompokkan satu set pohon keputusan yang dilatih dengan data yang sama tetapi dengan strategi yang berbeda memberikan serangkaian label yang dihasilkan menghindari kecenderungan pohon keputusan overfitting. Strategi paling umum untuk dilakukan adalah pilihan acak dari serangkaian fitur dan pilih salah satu yang menyediakan informasi lebih lengkap.

Random Forest bekerja dalam dua fase. Fase pertama yaitu menggabungkan sejumlah N decision tree untuk membuat *Random Forest*. Kemudian fase kedua adalah membuat prediksi untuk setiap tree yang dibuat pada fase pertama terlihat pada gambar 2.4. (Trisuvi 2022)



Gambar 2. 4 random forest

