

FTI/IA/SKRIPSI/2671/2023

PEMBUATAN SISTEM PREDIKSI SKOR REVIEW FILM BERBASIS
MACHINE LEARNING

HABIL MEDIO BHAKTI | 52419667 | 2023 | S1 - SKRIPSI | Informatika - Teknologi Industri

Pembimbing: Dr. Yuhilza Hanum, SSi., MEng. | Promotor:

Table Of Contents

ABSTRAKSI (BAHASA INGGRIS)	KATA PENGANTAR	DAFTAR ISI	DAFTAR TABEL	DAFTAR GAMBAR	E
----------------------------	----------------	------------	--------------	---------------	---

BAB 2

TELAAH PUSTAKA

2.1 Machine Learning

Machine Learning menjelaskan kapasitas sistem untuk belajar dari data pelatihan untuk menyelesaikan masalah tertentu dan untuk mengotomatisasi proses pembuatan model analitis dan menyelesaikan tugas terkait (Janiesch , Zschech , & Heinrich, 2021).

Machine Learning juga dapat mempelajari data yang ada dan data yang diperoleh sehingga bisa melakukan tugas tertentu. Dalam hal ini machine learning akan digunakan untuk memproses data berupa teks dan angka sesuai dengan kategori dari setiap fitur data. Model machine learning yang dibuat akan mencakup subbagian dari machine learning yang disebut deep learning.

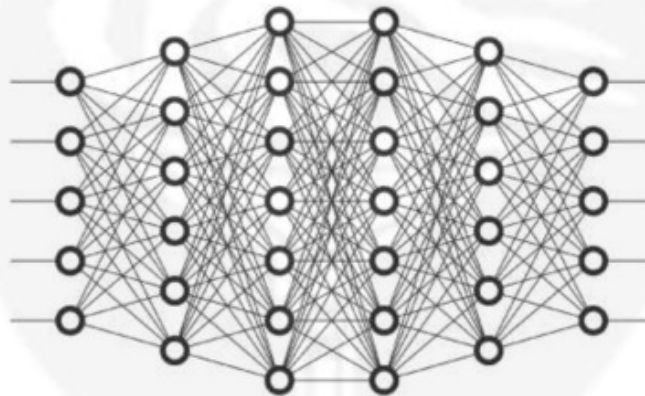
2.1.1 Deep Learning

Deep learning adalah konsep pembelajaran mesin berdasarkan jaringan saraf tiruan(Artificial Neural Networks). Dalam berbagai implementasi, model deep learning mengungguli model machine learning sederhana dan pendekatan analisis data tradisional(Janiesch , Zschech , & Heinrich, 2021).

Deep learning melibatkan pembuatan jaringan syaraf tiruan yang terinspirasi oleh struktur dan fungsi otak manusia. Jaringan ini terdiri dari lapisan-lapisan node yang saling berhubungan, yang disebut neuron, di mana setiap neuron memproses data masukan dan meneruskan keluarannya ke lapisan berikutnya. Aspek kunci dari machine learning adalah adanya beberapa lapisan tersembunyi(hidden layer) antara lapisan input dan output(input layer & output layer), yang memungkinkan jaringan untuk mempelajari representasi data secara hierarkis

2.1.1.1 Neural Network

Artificial Neural Network (ANN), atau biasa disebut Neural Network (NN), adalah model matematika atau model komputasi yang terinspirasi oleh struktur dan/atau aspek fungsional jaringan saraf biologis. Jaringan saraf terdiri dari kelompok neuron buatan yang saling berhubungan, dan memproses informasi menggunakan pendekatan koneksionis untuk komputasi. Dalam kebanyakan kasus, Neural Network adalah sistem adaptif yang mengubah strukturnya berdasarkan informasi eksternal atau internal yang mengalir melalui jaringan selama fase pembelajaran. Neural Network modern adalah alat pemodelan data statistik nonlinier yang biasa digunakan untuk memodelkan hubungan yang kompleks antara input dan output atau untuk menemukan pola dalam data (D'Addona, 2019).



Gambar 2. 1 Visualisasi Artificial Neural Network

Sumber: freecodecamp.com

2.1.2 Analisis Regresi

Analisis regresi adalah teknik statistik untuk memperkirakan hubungan antar variabel yang memiliki hubungan sebab dan akibat. Penelitian ini akan menggunakan metode bernama Regresi Linier Berganda. Regresi Linier Berganda adalah model regresi yang menggunakan satu variabel dependen dan lebih dari satu variabel independen (Güler, N. & Uyanık, G. K.).

Bentuk umum persamaan regresi linier berganda adalah $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p + \varepsilon$, dimana y adalah variabel dependen, b_0 adalah nilai intercept, x_1, x_2, \dots, x_p adalah variabel independen, b_1, b_2, \dots, b_p adalah koefisien yang berhubungan dengan variabel independen x_1, x_2, \dots, x_p , dan ε mewakili nilai galat.

Proses regresi linear berganda melibatkan pencarian koefisien b_0, b_1, \dots, b_p yang paling sesuai dengan data yang diamati. Hal ini umumnya dilakukan dengan menggunakan metode "ordinary least squares" (OLS), yang bertujuan untuk meminimalkan jumlah perbedaan kuadrat antara nilai yang diamati dan nilai yang diprediksi oleh model.

2.1.2.1 Mean Absolute Error

Mean Absolute Error adalah salah satu metric yang bertujuan untuk mengukur besaran rata-rata dari perbedaan absolut antara N vektor yang diprediksi $S = \{x_1, x_2, \dots, x_N\}$ dan $S^* = \{y_1, y_2, \dots, y_N\}$, yang sesuai fungsi kerugian didefinisikan sebagai:

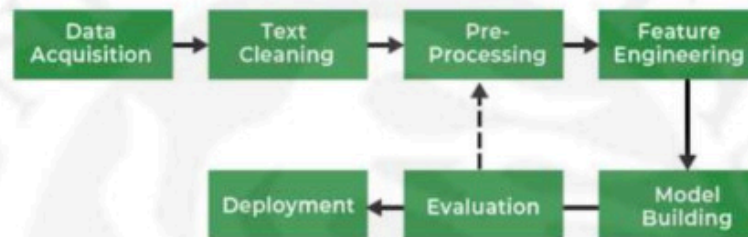
$$\mathcal{L}_{MAE}(S, S^*) = \frac{1}{N} \sum_{i=1}^N \|x_i - y_i\|_1,$$

Gambar 2. 2 Fungsi Mean Absolute Error

Dimana $\|\cdot\|_1$ menunjukkan norma L1 (Qi, Jun et al, 2020).

2.1.3 Natural Language Processing

Natural Language Processing (NLP) adalah bidang kecerdasan buatan yang berkaitan dengan pemahaman otomatis bahasa manusia oleh komputer. NLP telah melihat banyak keberhasilan dalam beberapa tahun terakhir karena peningkatan daya komputasi dan munculnya pembelajaran yang mendalam, dan keberhasilan ini telah meluas ke berbagai bidang seperti biomedis dan klinis (Vasilakes, Zhou, & Zhang, 2020).



Gambar 2. 3 Pipeline Natural Language Processing

Sumber: geeksforgeeks.org

Pipeline Natural Language Processing mengacu pada urutan proses yang terlibat dalam menganalisis dan memahami bahasa manusia. Tahap pertama yang dilakukan adalah akuisisi data. Data dapat diperoleh dari beberapa sumber seperti melalui Public Dataset, Web Scrapping, Image to Text, atau Data augmentation.

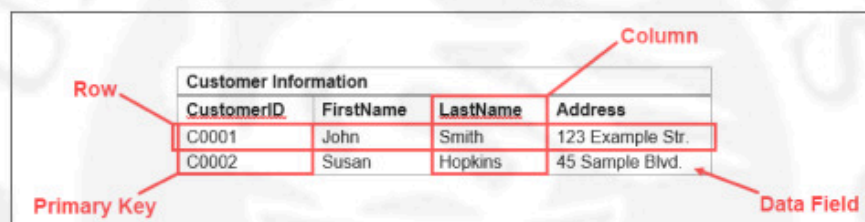
Tahap selanjutnya yang dilakukan adalah Text cleaning, hal ini bertujuan untuk mengoreksi dan membersihkan hal seperti pengejaan yang salah, karakter special, atau tag html. Selanjutnya akan dilakukan beberapa metode preprocessing seperti Tokenization, Lowercasing, Stop word removal sebelum melakukan feature engineering. Tujuan Feature Engineering adalah untuk merepresentasikan teks dalam vektor numerik sedemikian rupa sehingga algoritma machine learning dapat

memahami atribut teks. Dalam NLP proses rekayasa fitur ini dikenal sebagai Representasi Teks atau Vektorisasi Teks. Lalu data siap untuk diimplementasi ke dalam model.

2.1.4 Data Terstruktur (Structured data)

Data terstruktur adalah ketika data dalam format standar, memiliki struktur yang terdefinisi dengan baik, sesuai dengan model data, memiliki suatu urutan, dan mudah diakses oleh manusia dan program. Data ini umumnya disimpan dalam database (<https://www.tibco.com/reference-center/what-is-structured-data>).

Dataset yang digunakan untuk penelitian ini termasuk jenis data terstruktur yang didapat dari situs kaggle. Data akan berbentuk csv yang memiliki format tabular.



Customer Information			
CustomerID	FirstName	LastName	Address
C0001	John	Smith	123 Example Str.
C0002	Susan	Hopkins	45 Sample Blvd.

Gambar 2. 4 Contoh Data Terstruktur

Sumber: mongodb.com

2.1.5 Python

Python adalah bahasa pemrograman tingkat tinggi, berorientasi objek, dengan semantik dinamis. Struktur data bawaan tingkat tinggi, dikombinasikan dengan pengetikan dinamis dan pengikatan dinamis, membuat python sangat menarik untuk Rapid Application Development, serta untuk digunakan sebagai bahasa scripting atau glue language untuk menghubungkan komponen yang ada. Sintaks Python yang sederhana dan mudah dipelajari, dan menekankan keterbacaan dapat mengurangi biaya pemeliharaan program. Python mendukung modul dan paket, yang mendorong modularitas program dan penggunaan ulang kode. Interpreter Python dan perpustakaan standar yang luas tersedia dalam bentuk source atau biner tanpa biaya

untuk semua platform utama, dan dapat didistribusikan secara bebas(www.python.org).

2.1.5.1 Pemrograman Berbasis Objek

Python adalah Bahasa Pemrograman Berorientasi Objek (Object-Oriented Programming). Pemrograman Berorientasi Objek adalah paradigma pemrograman berdasarkan konsep 'objek', yang dapat berisi data dan kode yang memanipulasi data tersebut. Dalam OOP, penekanannya adalah pada pemodelan entitas dunia nyata sebagai objek perangkat lunak, dan mendefinisikan properti (atribut) dan perilaku (metode) mereka untuk berinteraksi satu sama lain. Pendekatan ini mempromosikan desain modular, penggunaan kembali kode, dan representasi yang lebih intuitif dari sistem yang kompleks(Balagurusamy , 2008). OOP dalam python memiliki beberapa konsep penting, yaitu enkapsulasi, pewarisan, dan polimorfisme.

Enkapsulasi adalah bagaimana membungkus data dan method yang menyusun sebuah class hingga class dipandang sebagai suatu modul, Inheritance (pewarisan) adalah proses pewarisan data dan method dari suatu class kepada class yang lain. Pewarisan ini bersifat menyeluruh sehingga semua data dan method yang dimiliki oleh class asalnya akan diturunkan kepada class baru, sedangkan polimorfisme adalah sesuatu yang memiliki banyak bentuk, diartikan sebagai modul yang memiliki kesamaan nama, namun behaviour (tingkah laku yang berbeda) sehingga listing kode implementasinya berbeda (Hermawan, 2004).

2.1.5.2 Tensorflow

TensorFlow, dalam istilah yang paling umum, adalah framework perangkat lunak untuk perhitungan numerik berdasarkan grafik dataflow. Tensorflow dirancang terutama sebagai antarmuka untuk mengekspresikan dan mengimplementasikan algoritma machine learning, yang utamanya adalah deep neural network(Hope, Resheff, & Lieder, 2017).

2.1.5.3 Keras

Keras adalah API jaringan saraf tingkat tinggi yang dikembangkan dengan fokus untuk memungkinkan eksperimen cepat. Kemampuan untuk beralih dari ide ke hasil dengan penundaan sesedikit mungkin adalah kunci untuk melakukan penelitian yang baik. Keras memiliki fitur utama berikut: Memungkinkan kode yang sama untuk berjalan pada CPU atau GPU, dengan mulus. API yang mudah digunakan yang memudahkan pembuatan prototipe model pembelajaran mendalam dengan cepat. Dukungan bawaan untuk convolutional networks (untuk computer vision), recurrent networks (untuk sequence processing), dan kombinasi keduanya. Mendukung arsitektur network arbitrer: model multi-input atau multi-output, layer-sharing, model-sharing, dan lain-lain. Keras pada dasarnya sesuai untuk membangun model machine learning apa pun, dari memory network hingga mesin neural Turing machine. Keras juga mampu berjalan di atas beberapa back-end termasuk TensorFlow, CNTK, atau Theano(Allaire & Chollet, 2021)

Selain standard neural network, Keras memiliki dukungan untuk convolutional neural network dan recurrent neural networks. Ini mendukung layer utilitas lainnya seperti dropout, normalisasi batch, dan pooling. Keras memungkinkan pengguna untuk menghasilkan model di smartphone (iOS dan Android), di web, atau di Java Virtual Machine.

2.2 API (Application Programming Interface)

Application Programming Interface menyediakan abstraksi untuk masalah dan menentukan bagaimana klien harus berinteraksi dengan komponen perangkat lunak yang mengimplementasikan solusi untuk masalah itu. Komponen itu sendiri biasanya didistribusikan sebagai pustaka perangkat lunak, memungkinkannya untuk digunakan dalam beberapa aplikasi. Intinya, API mendefinisikan blok bangunan yang dapat digunakan kembali yang memungkinkan potongan fungsionalitas modular untuk dimasukkan ke dalam aplikasi(Reddy, 2011).



Gambar 2. 5 Cara Kerja API

Sumber: gamelab.id

Secara keseluruhan, API melayani beragam tujuan, seperti mengakses layanan web, mengintegrasikan sistem perangkat lunak, mendukung pengembangan pihak ketiga, mengambil data, dan mengotomatiskan tugas. Antarmuka ini hadir dalam berbagai bentuk, seperti API web yang memanfaatkan HTTP, API pustaka dengan fungsi/kelas, dan API sistem operasi untuk akses perangkat keras dan sumber daya sistem. Dokumentasi API sangat penting, menawarkan detail titik akhir, parameter, metode autentikasi, dan format respons, memandu pengembang untuk menggunakan API secara efektif dalam aplikasi mereka.

2.3 Web Server

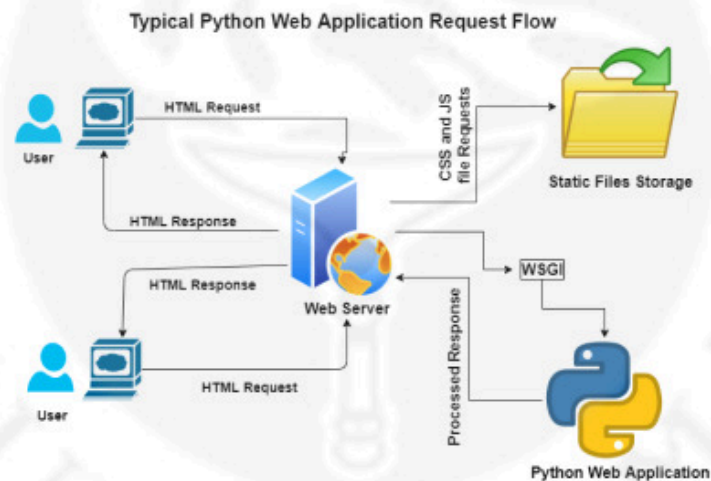
Web Server adalah komputer yang menjalankan situs web. Menggunakan protokol HTTP, server Web mengirimkan halaman Web ke browser serta file data lainnya ke aplikasi berbasis Web. Server Web mencakup perangkat keras, sistem operasi, perangkat lunak server Web, protokol TCP/IP dan konten situs (halaman Web, gambar, dan file lainnya) (pcmag.com).

Dalam penelitian ini penulis akan menggunakan web server localhost untuk mengimplementasikan model ke dalam bentuk web. Model machine learning akan di deploy menggunakan tools bernama Flask.

2.3.1 Flask

Python Flask adalah sebuah web framework, Flask adalah modul Python yang memungkinkan pengguna mengembangkan aplikasi web dengan mudah. Flask memiliki core yang sederhana dan mudah dikembangkan: Flask adalah microframework yang tidak menyertakan ORM (Object Relational Manager) atau fitur semacam itu. Framework ini memiliki banyak fitur seperti url routing dan template engine. Flask adalah framework aplikasi web WSGI(Web Server Gateway Interface)(pythonbasics.org).

Dalam penelitian ini, flask akan digunakan untuk melakukan deployment model ke dalam web server agar model dapat melakukan prediksi sesuai dengan input yang diberikan user.



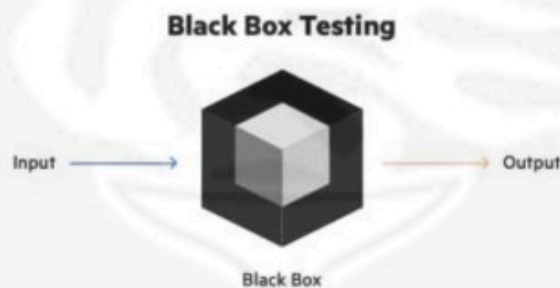
Gambar 2. 6 Cara Kerja Flask Web Application

Sumber: dev.to

2.4 Black Box Testing

Black Box Testing adalah Pengujian fungsional software tanpa mengetahui struktur internal program (Permatasari, 2020). Black Box Testing akan digunakan untuk melakukan tes terhadap fungsi yang ada di dalam aplikasi web yang akan dibuat.

Penguji memperlakukan perangkat lunak sebagai "Black Box", memeriksa input dan outputnya untuk menilai apakah perangkat bekerja seperti yang diharapkan. Metode ini sangat berguna untuk memvalidasi perilaku eksternal sistem dan mengidentifikasi cacat yang mungkin timbul dari interaksi yang tidak benar atau tidak terduga.



Gambar 2. 7 Black Box Testing

Sumber: imperva.com

2.5 Review Film

Menurut KBBI, review atau dalam Bahasa Indonesia disebut ulasan, adalah kupasan, tafsiran, atau komentar. Sedangkan film adalah lakon (cerita) gambar hidup.

Lalu, dapat disimpulkan bahwa ulasan atau review film adalah hasil dari kegiatan mengulas suatu film yang berisi komentar, kritik atau saran terhadap film tersebut yang biasanya tertulis ringkasan isi film tersebut sebelum mengulasnya.

Dengan adanya review film, kita dapat mengetahui kelebihan, kekurangan dan kualitas sebuah film. Menulis review film diperlukan pengetahuan yang luas tentang film serta wawasan tentang tata cara menulis review film yang baik, unsur-unsur film dan kriteria dalam menilai film. Kegiatan mengkritik merupakan kegiatan di dalam mereview film. Semakin dalam dan semakin lengkap informasi yang disampaikan maka semakin pantas review tersebut disebut dengan kritik(Irwansyah, Ade).

Terdapat dua jenis sentimen dalam suatu ulasan atau review, yaitu :

1. Pendapat Reguler(Regular Opinion)

Pendapat reguler berfokus terhadap satu objek, dimanfaatkan untuk menemukan poin positif atau negatif tentang tertentu produk(Qazi et al, 2016).

2. Pendapat Komparatif(Comparative Opinion)

Pendapat Komparatif membahas lebih dari satu objek, biasanya digunakan untuk intelijen kompetitif(Qazi et al, 2016).

