

LAPORAN
RANCANGAN PROYEK CAPSTONE

***HOTEL RESERVATIONS CLASSIFICATION USING RANDOM FOREST,
KKN & SVM***



Oleh Kelompok 11:

Muhammad Mujahid N.	(Ketua)
Reyhan Pramudita	(Anggota)
Gendhis Bestari Tanjung	(Anggota)
Anisah Fadhilah Putri	(Anggota)
Devita Fatricia	(Anggota)
Mang Agri Sativani	(Anggota)

STUDI INDEPENDEN BERSERTIFIKAT
PT. BISA ARTIFISIAL INDONESIA
KAMPUS MERDEKA

2023

DAFTAR ISI

DAFTAR ISI.....	i
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Ruang Lingkup Proyek.....	1
BAB II TINJAUAN LITERATUR.....	2
2.1 Reservasi Hotel.....	2
2.2 Metode Klasifikasi	2
2.3 <i>Random Forest</i> , KKN, dan SVM	2
BAB III METODOLOGI.....	4
3.1 Pengumpulan Data dan <i>Preprocessing</i>	4
3.2 Ekstraksi Fitur dan <i>Feature Selection</i>	4
3.3 Proses Klasifikasi	5
3.4 Evaluasi	5
BAB IV CARA KERJA PEMROGRAMAN	6
4.1 Instalasi API <i>Kaggle</i>	6
4.2 Eksplorasi Data	6
4.3 <i>Pre-Processing</i>	6
BAB V PENUTUP.....	7
5.1 Kesimpulan.....	7
5.2 Saran.....	7
DAFTAR PUSTAKA.....	8

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hotel adalah tempat peristirahatan sementara yang menyediakan fasilitas akomodasi dan layanan lainnya seperti kamar tidur, kamar mandi, restoran, kolam renang, spa, dan pusat kebugaranyang dikelola oleh perusahaan denganmelakukanpembayaran sesuai ketentuan. Hotel dikelompokkan dalam kategori berdasarkan fasilitas dan harga, seperti hotel bintang lima, empat bintang, tiga bintang. Proses reservasi hotel adalah bagian penting dari operasi hotel, yang memungkinkan tamu untuk memesan kamar sesuai dengan kebutuhan mereka. Metode reservasi yang paling populer saat ini adalah secara *online* (melalui situs web hotel/aplikasi) atau melalui telepon. Metode reservasi secara *online* ini juga mengakibatkan semakin besarnya faktor kemungkinan pembatalan reservasi, sehingga diperlukan adanya identifikasi reservasi hotel.

Kecerdasan Buatan merupakan ilmu yang ditujukan untuk pembuatan software dan hardware yang memiliki fungsi dalam untuk melakukan pemikiran seperti manusia. *Machine learning* merupakan cabang dari kecerdaan buatan yang cara bekerjanya belajar dari pemikiran manusia dengan menggunakan algoritma matematika. Selaras dengan hal tersebut, kami sebagai kelompok 11 menginisiasi sebuah sistem pemrograman untuk mengidentifikasi reservasi hotel dalam metode klasifikasi menggunakan *Random Forest*, *KKN*, dan *SVM* yang akan dibahas lebih lanjut pada laporan rancangan Proyek *Captone* ini.

1.2 Ruang Lingkup Proyek

Hotel reservasi adalah proses pemesanan kamar atau akomodasi di sebuah hotel oleh calon tamu. Hal ini dilakukan melalui berbagai metode, termasuk telepon, email, situs web hotel, agen perjalanan, atau platform pemesanan online.

BAB II

TINJAUAN LITERATUR

2.1 Reservasi Hotel

Secara garis besar, reservasi hotel diartikan sebagai sebuah proses pemesanan kamar atau akomodasi di hotel untuk periode waktu tertentu. Prosedur ini melibatkan tindakan menghubungi hotel, melalui telepon, email atau situs web hotel, dan memberikan informasi tentang tanggal yang diinginkan, jumlah orang, dan jenis kamar yang diinginkan yang memungkinkan pengunjung untuk menjamin ketersediaan kamar di hotel pada tanggal yang diinginkan.

Kemudian, untuk sistematisasi reservasi hotel sendiri dimulai dari pengumpulan informasi di berbagai platform; kemudian, menambah pada tindakan verifikasi untuk memastikan validitas informasi yang diterima; setelah itu, dilanjutkan dengan proses persetujuan antara dua belah pihak, yakni konsumen dan pihak hotel yang dituju dan setelah persetujuan diterima; berlanjut pada pembuatan reservasi dan konfirmasi dalam rangka menghindari pembatalan reservasi.

2.2 Metode Klasifikasi

Pada dasarnya, klasifikasi diartikan sebagai salah satu tugas yang penting data mining. Sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan merupakan pengelompokan fitur ke dalam kelas yang sesuai. Vektor fitur pelatihan tersedia dan telah diketahui kelasnya, kemudian vektor fitur pelatihan tersebut dimanfaatkan untuk merancang pemilah. Pengenalan pola ini disebut *terbimbing*, supervised.

2.3 Random Forest, KKN, dan SVM

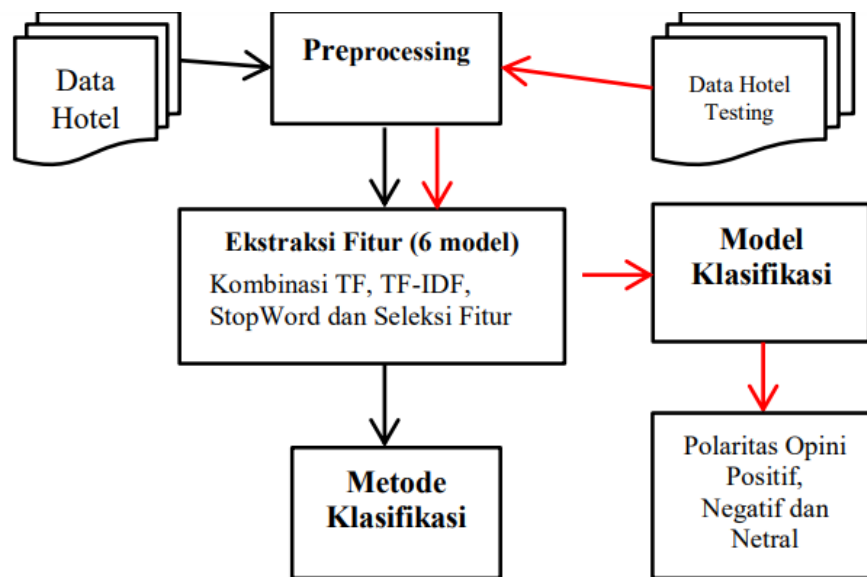
- A. *Random Forest* adalah algoritma dalam *machine learning* yang digunakan untuk pengklasifikasian data set dalam jumlah besar. Karena fungsinya bisa digunakan untuk banyak dimensi dengan berbagai skala dan performa yang tinggi. Klasifikasi ini dilakukan melalui penggabungan tree dalam *decision tree* dengan cara training dataset yang dimiliki (Anonim, 2022).
- B. Algoritma *K-Nearest Neighbor* (KNN) merupakan algoritma *machine learning* sederhana dan mudah diterapkan yang dapat digunakan untuk menyelesaikan

masalah klasifikasi dan regresi. Serta, dapat diartikan sebagai algoritma machine learning yang bersifat *non-parametric* dan *lazy learning*. Tepatnya, bersifat *non-parametric* memiliki makna bahwa metode tersebut tidak membuat asumsi apa pun tentang distribusi data yang mendasarinya. Dengan kata lain, tidak ada jumlah parameter atau estimasi parameter yang tetap dalam model, terlepas data tersebut berukuran kecil ataupun besar; dan bersifat *lazy learning*, yang artinya tidak menggunakan titik data training untuk membuat model. Singkatnya pada algoritma KNN tidak ada fase training, walaupun ada juga sangat minim.

- C. SVM adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. *Support Vector Machine* (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. Sebagai salah satu metode *pattern recognition*, usia SVM terbilang masih relatif muda. Walaupun demikian, evaluasi kemampuannya dalam berbagai aplikasinya menempatkannya sebagai *state of the art* dalam *pattern recognition* (Wibawa, Aji Prasetya., dkk, 2018).

BAB III METODOLOGI

Berikut ini merupakan gambaran usulan metode beserta pembahasannya secara terperinci yang digunakan dalam perancangan Proyek *Capstone* ini:



3.1 Pengumpulan Data dan *Preprocessing*

Data set yang digunakan untuk uji coba sistem reservasi hotel ini diambil dari website www.kaggle.com. Kemudian, *Preprocessing* dilakukan melakukan *case folding*, *filtering* dan *tokenizing*. *Case folding* adalah proses untuk mengubah huruf besar menjadi huruf kecil. *Tokenizing* digunakan untuk memecah dokumen menjadi token-token kata yang siap untuk proses selanjutnya. *Filtering* digunakan untuk mengambil kata-kata penting dari hasil token dengan membuang *stopword*. Dalam uji coba data set, penggunaan *stopword* digunakan sebagai parameter ekstraksi fitur.

3.2 Ekstraksi Fitur dan *Feature Selection*

Selanjutnya, dari data pelatihan dilakukan ekstraksi fitur, pada umumnya menggunakan pembobotan kata dalam dokumen, sehingga dokumen dapat dinyatakan dalam bentuk vektor. Pembobotan kata dapat menggunakan metode Binary, TF (Term Frequency), TFIDF (Term Frequency-Inverse Document Frequency). TF menggunakan kemunculan kata dalam sebuah dokumen, sedangkan Idf berfungsi mengurangi bobot suatu kata jika kemunculannya banyak tersebar di seluruh dokumen. Serta, untuk *Feature*

Selection diartikan sebagai *preprocessing* data untuk memilih fitur yang berpengaruh dan membuang fitur yang tidak berpengaruh dalam analisa data. Proses ini untuk mengurangi dimensi fitur yang besar dengan *Gain Ratio* dan *cfsSubsetEval*.

3.3 Proses Klasifikasi

Setelah itu, proses klasifikasi menggunakan *supervised learning*, terdapat beberapa metode klasifikasi yaitu Random Forest, KKN, dan SVM. *Supervised learning*, terdapat dua tahap yaitu tahap pembelajaran dan tahap klasifikasi. Tahap pembelajaran untuk membangun model klasifikasi dari data pelatihan yang telah dilabeli berdasarkan class positif, negatif atau netral. Tahap klasifikasi untuk melakukan klasifikasi untuk menentukan class dari data uji coba. Algoritma Random Forest adalah sekumpulan *classifier* yang terdiri dari banyak pohon keputusan dan melakukan klasifikasi berdasarkan keluaran dari hasil klasifikasi setiap pohon keputusan anggota. Algoritma KNN mengklasifikasikan data baru ke class tertentu, berdasarkan jarak terdekat dari data tetangganya. Serta, *Support vector machines* (SVM) adalah suatu metode dengan pendekatan persamaan *Lagrangian* yang merupakan bentuk dual dari SVM melalui *quadratic programming*.

3.4 Evaluasi

Evaluasi pemilahan data pelatihan dan data testing menggunakan teknik *crossvalidation* dan *percentage split*. *Cross validation* menggunakan *10fold* dan *percentage split* 66%. Evaluasi kinerja metode klasifikasi menggunakan *True Positive(TP)*, *False Positive(FP)*, *False Negative(FN)*, *True Negative(TN)*. Nilai TN merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan FP merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, TP merupakan data positif yang terdeteksi benar. FN merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif.

BAB IV

CARA KERJA PEMROGRAMAN

4.1 Instalasi API *Kaggle*

Kaggle adalah situs dan platform untuk berlomba membuat model terbaik untuk menganalisa dan memprediksi suatu dataset. Dataset ini diberikan oleh suatu perusahaan, dengan suatu deskripsi masalah tertentu. Cara instalasi buka di website melalui [kaggle.com](https://www.kaggle.com)

4.2 Eksplorasi Data

Eksplorasi data langkah untuk memahami data sebelum dilakukan praproses. Tujuanya untuk menyeleksi teknik pemrosesan dan analisis dataset yang di miliki.

4.3 *Pre-Processing*

Data preprocessing proses mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini diperlukan untuk memperbaiki kesalahan pada data mentah yang seringkali tidak lengkap dan memiliki format yang tidak teratur.

Cara kerja :

A. Data Cleaning

Dalam langkah *data cleaning*, data mentah akan dibersihkan melalui beberapa proses seperti mengisi nilai yang hilang, menghaluskan *noisy* data, dan menyelesaikan inkonsestensi yang ditemukan.

B. Data Integration

Memastikan data memiliki format dan atribut yang sama, Menghapus atribut yang tidak dibutuhkan dari semua sumber data, Mendeteksi nilai data yang konflik.

C. Data Transformation

Pada tahap ini, data akan dinormalisasi dan digeneralisasi. Normalisasi data dilakukan untuk memastikan bahwa tidak ada data yang berlebihan, sementara generalisasi data dilakukan untuk menyeragamkan data.

BAB V

PENUTUP

5.1 Kesimpulan

Hotel adalah tempat peristirahatan sementara yang menyediakan fasilitas akomodasi dan layanan lainnya seperti kamar tidur, kamar mandi, restoran, kolam renang, spa, dan pusat kebugaranyang dikelola oleh perusahaan dengan melakukan pembayaran sesuai ketentuan. reservasi hotel diartikan sebagai sebuah proses pemesanan kamar atau akomodasi di hotel untuk periode waktu tertentu. Prosedur ini melibatkan tindakan menghubungi hotel, melalui telepon, email atau situs web hotel, dan memberikan informasi tentang tanggal yang diinginkan, jumlah orang, dan jenis kamar yang diinginkan yang memungkinkan pengunjung untuk menjamin ketersediaan kamar di hotel pada tanggal yang diinginkan.

5.2 Saran

Project ini merupakan tugas akhir sebagai penilaian dari kegiatan msib, Maka saran dari penulis agar pembimbing bisa memberikan arahan dengan memberitahu kesalahan dalam project ini, kemudian hal hal apa yang tidak harus di gunakan dalam membuat project ini, khususnya project hotel reservation ini.

DAFTAR PUSTAKA

<https://accurate.id/teknologi/data-preprocessing/>

<https://indoml.com/2017/08/22/panduan-menggunakan-kaggle>