

## MODEL DATA MINING UNTUK KAREKTERISTIK DATA TRAVELLER PADA PERUSAHAAN TOUR AND TRAVEL (Studi Kasus: Lombok Ceria Holiday)

Saikin<sup>1</sup>, Kusrini<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Universitas Amikom Yogyakarta

Jln. Basuki Rahmat No.105 Praya Lombok Tengah 83511

Jl. Ringroad Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta

<sup>1</sup> eken.apache@gmail.com, <sup>2</sup> Kusrini@amikom.ac.id

---

### Abstract

*For companies engaged in the service sector, such as tour and travel companies data processing is very important to know the characteristics or interests of tourists in traveling. Difficult to predict the needs or interests of tourists, is an obstacle faced by tour and travel companies so that management must be able to make the right and fast decisions, in order to provide good service and satisfaction to customers. Decisions taken must be well considered based on the data held, especially those closely related to the characteristics of traveler data. Data analysis to look for patterns of data characteristics of a traveler requires good methods and accurate results, so the results of the analysis can find useful information for the company. Data processing method that is often used is data mining. K-Nearest Neighbor (K-NN) is a data mining algorithm that works by applying learning to data. The purpose of this study will be to analyze the data of travelers or tourists on tour and travel companies by using the K-NN algorithm to look for interests or characteristics of tourists in choosing a tourist attraction. In the K-NN method qualification results and testing using the confusion matrix method, the accuracy value is 84%, the precision value is 88%, the recall value is 85 and the f1-score is 85%. The qualification results show that the tourist attraction that tends to be the dominant tourist interest is on the beach tourism object. Based on the results of the analysis using these two methods, the characteristics of tourists tend to choose package number 3 namely the Regular Lombok Package Tour, and the dominant tourist attraction is the tourist attraction, the object beach tourism and followed by dyke attractions.*

**Keywords :** Data Mining, K-NN, Tour and Travel,

### Abstrak

Bagi perusahaan yang bergerak dibidang sektor jasa, seperti perusahaan *tour and travel* pengolahan data sangatlah penting untuk mengetahui karakteristik atau minat wisatawan dalam berwisata. Sulitnya memprediksi kebutuhan atau minat wisatawan, merupakan kendala yang dihadapi perusahaan *tour and travel* sehingga manajemen harus dapat mengambil keputusan yang tepat dan cepat, guna memberikan pelayanan yang baik serta kepuasan kepada *customer*. Keputusan yang diambil harus mempertimbangkan dengan baik berdasarkan data-data yang dimiliki terutama yang berkaitan erat dengan karakteristik data *traveller*. Analisis data untuk mencari pola karakteristik data dari *traveler* memutuhkan metode yang baik dan hasil yang akurat , sehingga hasil dari analisis tersebut dapat menemukan informasi yang berguna bagi perusahaan. Metode pengeolahan data yang sering digunakan ialah data mining. K-Nearest Neighbour (K-NN) merupakan salah satu algoritma data mining yang cara kerjanya menerapkan pembelajaran pada data. Tujuan dari penelitian ini akan melakukan analisis data *travellers* atau wisatawan pada perusahaan *tour and travel* dengan menggunakan algoritma K-NN untuk mencari minat atau karakteristik wisatawan dalam memilih objek wisata. Pada hasil kualifikasi metode K-NN dan pengujian dengan metode *confusion matrix* nilai akurasi yang didapat 84%, nilai presisi 88%, nilai recall 85 dan nilai *f1-score* 85%. Pada hasil kualifikasi menunjukkan bahwa objek wisata yang cenderung dominan menjadi minat wisatawan ada pada objek wisata pantai. Berdasarkan hasil analisis dengan menggunakan kedua metode tersebut bahwa karakteristik wisatawan cenderung memilih paket nomor 3 yakni *Reguler Tour Lombok Package*, dan objek wisata yang dominan minat wisatawan yakni objek wisata pantai dan disusul oleh objek wisata gili.

**Kata kunci :** Data Mining, K-NN, Tour and Travel,

---

## 1. Pendahuluan

Dalam dunia teknologi dan bisnis peranan data sangatlah penting. Pengolahan data yang baik dan tepat akan menjadi kesempatan yang luas bagi perusahaan untuk bertahan pada persaingan bisnis, karena bagi perusahaan saat ini untuk menjalankan proses bisnis bukan hanya strategi yang baik, tapi bagaimana mengelola data dengan baik dan benar.

Bagi perusahaan yang bergerak dibidang sektor jasa, seperti perusahaan *tour and travel* pengolahan data sangatlah penting untuk mengetahui karakteristik atau minat wisatawan dalam berwisata. Sulitnya memprediksi kebutuhan atau minat wisatawan, merupakan kendala yang dihadapi perusahaan *tour and travel* sehingga manajemen harus dapat mengambil keputusan yang tepat dan cepat, guna memberikan pelayanan yang baik serta kepuasan kepada *customer*. Keputusan yang diambil harus mempertimbangkan dengan baik berdasarkan data-data yang dimiliki terutama yang berkaitan erat dengan karakteristik data *traveller*.

Pengenalan karakteristik data *traveller* dengan menggunakan data-data yang telah ada di masa lalu untuk memprediksi kebutuhan atau minat para *customer*/wisatawan dimasa yang akan datang sangat diperlukan oleh perusahaan. Agar menjadi acuan perusahaan untuk meningkatkan pelayanan.

Analisis data untuk mencari pola karakteristik data dari *traveler* memutuhkan metode yang baik dan hasil yang akurat, sehingga hasil dari analisis tersebut dapat menemukan informasi yang berguna bagi perusahaan. Metode pengolahan data yang sering digunakan ialah data mining. K-Nearest Neighbour (K-NN) merupakan salah satu algoritma data mining yang cara kerjanya menerapkan pembelajaran pada data. K-NN berkerja dengan mencari sejumlah k objek data atau pola (dari semua pola latih yang ada) yang paling dekat dengan pola masukan, kemudian memilih kelas dengan sejumlah pola terbanyak diantara k pola latih, (suyanto, 2017).

Penelitian yang dilakukan oleh Rohman 2015 dengan judul "Model Algoritma *K-Nearest Neighbor* (KNN) untuk memprediksi kelulusan mahasiswa". Tujuan dari penelitian ini untuk mengukur tingkat akurasi KNN dalam memprediksi tingkat kelulusan mahasiswa. Atribut yang dianalisis seperti fakultas, jenis kelamin, umur, IPK dan semester dari 1 sampai

semester 4. Dari 1582 data mahasiswa ada 8 dimensi yang terdiri dari 2 kelas yakni kelas tepat dan terlambat. Pengujian model menggunakan *cross validation* dan pengujian hasil kualifikasi menggunakan *confusion matrix*. Hasil dari kualifikasi menunjukkan 671 mahasiswa lulus tepat waktu sedangkan terlamat berjumlah 911 mahasiswa. Sedangkan nilai hasil validasi hasil kualifikasi menunjukkan nilai *accuracy* 82,25%, dan nilai *area under cover cuper (AUC)* 0,500.

Berdasarkan penelitian sebelumnya pada penelitian ini akan melakukan analisis data *travellers* atau wisatawan pada perusahaan *tour and travel* dengan menggunakan algoritma K-NN untuk mencari minat atau karakteristik wisatawan dalam berwisata.

## 2. Tinjauan Pustaka

Data Mining: Algoritma *K-Means* Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi, (Sari dan Hartama, 2018). Tujuan dari penelitian ini mengelompokan dari data kunjungan wisatawan di hotel berbintang pada setiap provinsi-provinsi selama kurun waktu tahun 2003-2016 menggunakan Algoritma *K-Means*. Hasil dari penelitian ini Hampir 90% provinsi di Indonesia masih memiliki potensi wisata yang rendah. Itu dibuktikan dari jumlah tamu asing yang menginap di hotel berdasarkan laporan dari badan pusat statistik.

Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik *Clustering*, (Adiana dkk, 2018). Tujuan utama penelitian ini adalah segmentasi pelanggan dan mengukur kesetiaan mereka terhadap sebuah produk UKM dengan mengkombinasikan model RFM dan teknik *clustering* menggunakan metode K-Means. Peneliti ini menyimpulkan karakteristiknya masing-masing. *Cluster* pertama dengan 30 pelanggan masuk ke dalam kategori *typical customers*. *Cluster* kedua terdiri dari 8 pelanggan masuk ke dalam kategori *superstar*. Kemudian 89 pelanggan termasuk dalam kategori *dormant customer*. Kemudian pengujian dilakukan menggunakan *global silhouette* dan *variance*. Hasilnya menunjukkan bahwa data di setiap *cluster* berada pada *cluster* yang tepat dan memiliki kemiripan yang dekat. Aplikasi model RFM dan teknik *cluster* khususnya algoritma K-Means, dapat membantu mengidentifikasi pelanggan dengan nilai tinggi dan berpotensi memberi keuntungan serta dapat membantu dalam merancang strategi pemasaran yang tepat untuk ketiga *cluster* yang terbentuk.

Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode *Random Forest*, (Nugroho dan Emiliyawati, 2017). Penelitian ini dilakukan dalam rangka untuk mengklasifikasi faktor-faktor yang mempengaruhi tingkat penerimaan konsumen terhadap mobil menggunakan metode *Random Forest* (RF). Hasil dari penelitian ini Sistem Klasifikasi variabel yang mempengaruhi tingkat penerimaan konsumen terhadap mobil telah selesai dibangun sehingga produsen dapat menjadikan hasil Klasifikasi dari sistem untuk proses produksi mobil. Klasifikasi variabel yang mempengaruhi tingkat penerimaan konsumen terhadap mobil dapat dilakukan dengan metode *Random Forest* dengan algoritma *gainratio*. Pohon keputusan yang dihasilkan dapat berbeda-beda akibat penentuan atribut serta pengambilan data dilakukan secara acak. Variabel yang menempati sebagai *root node* dalam setiap pohon keputusan yang dihasilkan memiliki pengaruh yang paling signifikan pada tingkat penerimaan konsumen terhadap mobil.

Penerapan Data Mining Dalam Pengelompokan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. DKI Jakarta Dengan K-Means, (Maulida, 2018). Tujuan dari penelitian ini Mengelompokkan data wisatawan dengan metode K-Mean dan mengetahui objek wisata unggulan di Prov. DKI Jakarta yang memiliki potensi yang paling rendah dalam kunjungan wisatawan ke Indonesia khususnya Prov. DKI Jakarta. Kesimpulan dari penelitian ini diperoleh dari pengelompokan dengan metode K-Mean bahwa C1 terdiri dari 1 objek wisata unggulan yakni Taman Impian Jaya Ancol, C2 terdiri dari 2 objek wisata unggulan yakni Taman Mini Indonesia Indah Dan Kebon Binatang Ragunan, C3 terdiri dari 5 objek wisata unggulan yakni Monumen Nasional, Museum Nasional, Museum Satria Mandala, Museum Sejarah Jakarta Dan Pelabuhan Sunda Kelapa. Dan Pelabuhan Sunda Kelapa. Dan dari penelitian menunjukkan bahwa 5 objek wisata unggulan yang berada di *cluster* paling rendah menjadi catatan bagi pemerintah Prov. DKI Jakarta. Dengan adanya perbaikan sarana dan prasarana objek wisata unggulan dapat meningkatkan jumlah kunjungan wisatawan yang berdampak pada pengenalan objek wisata dan peningkatan devisa negara. Pemilihan paket wisata menggunakan metode Analytical Hierarchy Process (AHP), (Suryani dkk, 2017). Tujuan dari penelitian ini merancang suatu system pendukung keputusan pemilihan

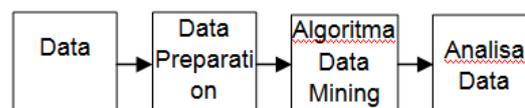
paket wisata berbasis web selain dapat untuk membantu dalam hal mempromosikan agen perjalanan ini, juga dapat mempermudah wisatawan untuk melakukan proses pemilihan paket wisata dan pemesanan paket wisata. Hasil dari penelitian ini dibangun aplikasi pemilihan paket wisata dengan menerapkan metode *Analytical Hierarchy Process* (AHP). Pengguna (*User*) dari aplikasi ini dapat melakukan pemilihan paket wisata yang sesuai dengan kriteria yang ditentukan. Pengguna (*User*) dari aplikasi ini dapat melakukan pemilihan paket wisata yang sesuai dengan kriteria yang ditentukan.

### 3. Metodologi Penelitian

Pada penelitian ini menggunakan jenis atau tipe penelitian tindakan yang berfokus pada tindakan sosial. Penelitian tindakan (*action research*) adalah penelitian baik kualitatif maupun kuantitatif. Penelitian tindakan adalah cara melakukan masalah pada saat yang bersamaan. Penelitian tindakan ini merupakan metode yang didasarkan pada tindakan masyarakat yang seringkali diselenggarakan pada suatu latar yang luas, seperti di rumah sakit, pabrik, sekolah, dan lain sebagainya, (Hasibuan, 2007).

#### 3.1 Metode

*Data Mining* (Witten, 2011) didefinisikan sebagai proses penemuan pola dalam data. Berdasarkan tugasnya, *data mining* dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, *clustering* dan asosiasi (Larose, 2005). Proses dalam tahap *data mining* (Gambar 1.) terdiri dari tiga langkah utama (Sumathi, 2006), yaitu :



Gambar 1 Langkah-langkah dalam Proses *Data Mining* (Maimon & Rokach, 2010)

#### 1. Data Preparation

*Data preprocessed* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

#### 2. Algoritma Data Mining

Teknik yang dibuat didalam data mining yaitu salah satunya bagaimana menelusuri data yang ada untuk membentuk sebuah model, kemudian menggunakan model tersebut bisa

dapat mengenali pola data yang lain yang tidak berada dalam basis data yang sudah tersimpan. Teknik seperti ini juga dapat dimanfaatkan untuk kebutuhan prediksi. Dalam data mining, pengategorisasian data juga bisa dilakukan. Bertujuan agar kita dapat mengetahui pola universal data-data yang ada, (Eko, 2012). Penggunaan algoritma *data mining* dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai, (Leidiyana, 2013).

### 3.2 Klasifikasi

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, 2006).

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi (Han, 2006). Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011).

### 3.3 K-nearest neighbor

*k-nearest neighbor* (kNN) termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. kNN dilakukan dengan mencari kelompok *k* objek dalam data *training* yang paling dekat (mirip) dengan objek pada data baru atau data *testing* (Wu, 2009).

Algoritma KNN adalah salah satu metode yang digunakan untuk analisis klasifikasi, namun beberapa dekade terakhir metode KNN juga digunakan untuk prediksi, (Alkhatib, 2013). Mencari jarak terdekat antara data yang akan dievaluasi dengan *k* tetangga (*neighbor*) terdekatnya dalam data pelatihan. Ruang ini dibagi menjadi kelompok-kelompok berdasarkan klasifikasi data training. Sebuah titik pada tempat ini ditandai dengan kelas *c*, dimana jika kelas *c* merupakan klasifikasi yang sangat banyak ditemui pada *k* buah tetangga terdekat titik tersebut. Untuk sistem kerja KNN data training di proyeksikan ke tempat berdimensi banyak, yang mana dari masing-masing dimensi mempersentasikan fitur dari data, (Whidhiasih, 2013). Model persamaan algoritma KNN dibawah ini:

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Keterangan:

$x$  = sampel data

$y$  = data uji

$D$  = jarak

$$fknn(x') = \frac{1}{K} \sum_{i \in N_k(x')} y_i$$

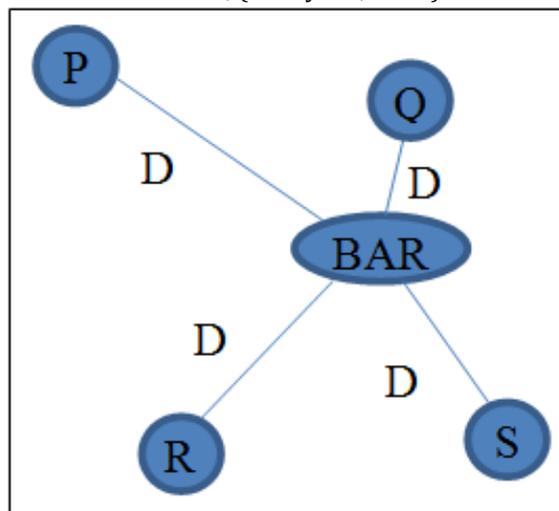
$x'$  = Perkiraan atau estimasi

$K$  = Jumlah tetangga terdekat

$N_k(x')$  = Tetangga terdekat

$y_i$  = Output tetangga terdekat

Contoh kasus, misal diinginkan untuk mencari solusi terhadap masalah seorang pasien baru dengan menggunakan solusi dari pasien lama. Untuk mencari solusi dari pasien baru tersebut digunakan kedekatan dengan kasus pasien lama, solusi dari kasus lama yang memiliki kedekatan dengan kasus baru digunakan sebagai solusinya. Terdapat pasien baru dan 4 pasien lama, yaitu P, Q, R, dan S (Gambar 2). Ketika ada pasien baru maka yang diambil solusi adalah solusi dari kasus pasien lama yang memiliki kedekatan terbesar, (Leidiyana, 2013).



Gambar 2. ilustrasi kasus algoritma kNN, (Leidiyana, 2013).

### 3.4 Confusion matrix

Metode ini menggunakan tabel matriks seperti pada Tabel 1 jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007).

Tabel 1 Model *Confusion Matrix* (Bramer, 2007)

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	true positives	false negatives
-	false positives	true negatives

*True positives* adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity* (*recall*), *specificity*, *precision* dan *accuracy*.

*Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif. Untuk menghitung digunakan persamaan di bawah ini, (Han, 2006) :

$$sensitivity = \frac{TP}{P}$$

$$specificity = \frac{TN}{N}$$

$$precision = \frac{TP}{TP + FP}$$

$$accuracy = sensitivity \frac{P}{(P + N)} + specificity \frac{N}{(P + N)}$$

Keterangan:

TP = jumlah *true positives*

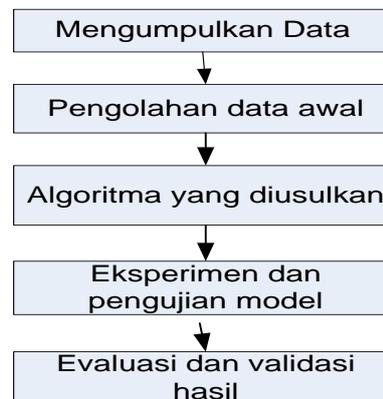
TN = jumlah *true negatives*

P = jumlah *record* positif

N = jumlah *tupel* negatif

FP = jumlah *false positives*

### 3.5 Alur Penelitian



Gambar 1 Alur Penelitian

#### 1. Pengumpulan Data

Pengumpulan data tidak hanya sekedar bagaimana mendapatkan data dan dikumpulkan akan tetapi bagaimana mendeskripsikan data tersebut sehingga dapat diolah dengan algoritma yang telah ditentukan. Berikut adalah tahapan dalam pengumpulan data.

- Wawancara (*interview*) dilakukan bersama Zulyadaen manajer Lombok Ceria Holiday Travel, untuk mendapat data dan informasi pelanggan, paket travel dan informasi tentang transaksi pemesanan paket travel oleh konsumen.
- Bagian in Data yang digunakan pada penelitian ini adalah data sekunder yang diambil dari proses transaksi pembelian paket wisata ditravel Lombok ceria holiday selama tiga tahun terakhir dan jumlah data yang diperoleh sebanyak 184 data wisatawan.

#### 2. Pengolahan Data Awal

umlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 184 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*).

- Pembersihan data (*data cleaning*) merupakan tahapan untuk membersihkan data dengan menghilangkan data-data yang salah, memperbaiki kecacauan data dan memeriksa data yang tidak konsisten. Terjadinya *missing value* dikarenakan terjadi kekosongan data atau objek data yang hilang. Dan memperivikasi data *outlier* atau pencilan data.

- Tahapan seleksi adalah tahapan memilih data dari setruktur basis data yang dimiliki untuk dijadikan data yang akan digunakan pada data mining. Dari 184 data yang diperoleh akan diseleksi menjadi 91 data tanpa menghilangkan kualitas dari data.
- *Transformation* melakukan pengelompokan beberapa atribut yang diambil dari berbagai tabel diletakan dalam satu tabel.

Tabel 1 transformasi atribut

No	Atribut	Keterangan
1	Peserta	Jumlah peserta wisatawan dalam rombongan paket
2	Hari	Jumlah hari perjalanan wisata yang diminta wisatawan
3	Paket	Paket wisata yang dipilih atau yang diambil wisatawan
4	destinasy	Objek wisata pilihan wisatawan

- Inisialisasi data bertujuan untuk mengubah data yang bertipe data karakter menjadi tipe data number untuk memudahkan dalam melakukan penklasifikasian data. Inisialisasi dilakukan pada paket wisata dan jenis objek wisata.

Tabel 2 inisialisasi paket wisata

Paket	Inisial
HONEYMOON LOMBOK PACKAGE	1
FREE AND EASY LOMBOK PACKAGE	2
REGULER TOUR LOMBOK PACKAGE	3
LOMBOK PACKAGE TOUR 5D-4N	4
LOMBOK PACKAGE TOUR 4D-3N	5

- Pada objek wisata berdasarkan data yang didapat dikategorikan dalam 5 kategori. Kateri tersebut terbagi dalam berbagai jenis paket wisata berikut kateregori jenis objek wisata

Tabel 3 jenis objek wisata

No	Kategori objek wisata
1	Pantai
2	Gili
3	Desa
4	Air terjun
5	Gunung

### 3. Model Yang Diusulkan

Algoritma yang digunakan pada penelitian ini ialah dengan menggunakan algoritma K-NN Penentuan nilai K label kelas pada penelitian ini berdasarkan kategori objek wisata, disini penentuan kategori objek wisata ditentukan menjadi 5 kategori

Tabel 4 Jenis K Labels

Objek Wisata	K Labels
Pantai	1
Gili	2
Desa	3
Air Terjun	4
Gunung	5

### 4. Menguji Model

Pada bagian ini merupakan tahapan untuk menguji hasil dan validasi dari klasifikasi data. Metode yang digunakan ialah metode *confusion matrix*. Matrix ini akan menemukan akurasi dari data, serta akan mencari nilai *recall* dan presisi dari kualifikasi data.

### 4. Hasil dan Pembahasan

Penelitian ini menggunakan 97 record dari 184 dataset yang didapatkan pada tahap pengumpulan data. Dari fitur data set digunakan objek wisata merupakan Label tujuan sedangkan label paket wisata, jumlah peserta dan jumlah hari merupakan data latih.

#### 4.1 K-NN

Kualifikasi dengan algoritma K-NN bertujuan untuk memprediksi objek wisata yang akan dipilih oleh wisatawan yang berdasarkan paket wisata, jumlah peserta dalam paket dan jumlah hari dalam paket yang dipesan. Penentuan K label kelas sebanyak 5 berdasarkan kategori objek wisata yakni pantai, gili, desa, air tujun dan gunung.

Tabel 5. Data klasifikasi K-NN

paket	Obwisata	Peserta	Hari
1	2	2	4
1	2	2	3
1	2	2	5
1	2	2	4
3	4	5	3
3	4	6	3
3	4	4	3
3	4	7	3
3	4	3	3
3	4	7	3
4	5	2	3
4	5	2	3
4	5	2	3
5	3	3	3
5	3	9	3
Dst	..	..	..

#### 4.2 Pengujian Model

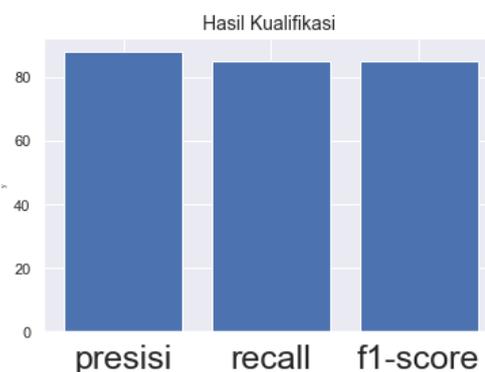
Hasil uji yang didapat dengan metode *confusion matrix*, hasil akurasi yang didapat 84%, presisi berjumlah 88% dan *recall* 85 dan *f1-score* 85.

	precision	recall	f1-score	support
1	1.00	0.75	0.86	4
2	1.00	1.00	1.00	1
3	1.00	1.00	1.00	2
4	0.80	0.80	0.80	5
5	0.50	1.00	0.67	1
micro avg	0.85	0.85	0.85	13
macro avg	0.86	0.91	0.86	13
weighted avg	0.88	0.85	0.85	13

The accuracy KNN:  
 = 0.8461538461538461

Gambar 3 hasil klasifikasi

hasil kualifikasi juga digambarkan dalam bentuk bar chart dimana menampilkan hasil perhitungan *confusion matrix* pada sisi presisi, *recall* dan *f1-score*.



Gambar 3 Grafik Pengukuran *Confusion Matrix*

## 5. Kesimpulan dan saran

### Kesimpulan

Pada hasil kualifikasi metode K-NN dan pengujian dengan metode *confusion matrix* nilai akurasi yang didapat 84%, nilai presisi 88%, nilai *recall* 85 dan nilai *f1-score* 85%. Pada hasil kualifikasi menunjukkan bahwa objek wisata yang cenderung dominan menjadi minat wisatawan ada pada objek wisata pantai. Berdasarkan hasil analisis dengan menggunakan kedua metode tersebut bahwa karakteristik wisatawan cenderung memilih paket nomor 3 yakni *Reguler Tour Lombok Packagage*, dan objek wisata yang dominan minat wisatawan yakni objek wisata pantai dan disusul oleh objek wisata gili.

### Saran

Pada penelitian ini masih terdapat kekurangan dan perlunya tahap-tahap penelitian yang perlu diperbaiki, terutama pada tahapan pengolahan data awal. Penelitian ini juga hanya difokuskan pada objek wisata yang ada dalam satu wilayah, untuk penelitian selanjutnya disarankan untuk melakukan penelitian untuk scope wilayah yang lebih luas dengan menggunakan metode yang lain

### Daftar Pustaka:

- [1] Adiana dkk, 2018, Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering, JUTEL, ISSN 2579-3675, e-ISSN 2579-5538, Vol.2 No.1 April 2018, Yogyakarta.
- [2] Alkhatib K, Najadat H, Hmeidi I, Shatnawi MKA. 2013. Stock Price Prediction Using *k-Nearest Neighbor (kNN)* Algorithm. *International Journal of Business, Humanities and Technology*. 3(3):32-44.

- [3] Bramer, Max. 2007. *Principles of Data Mining*. London : Springer
- [4] Eko Prasetyo, 2014, *Data Mining mengolah Data Menjadi Informasi Menggunakan Matlab*, Penerbit Andi, Yogyakarta.
- [5] Fitria dan Hartono, 2017, *Peramalan Jumlah Penumpang Pada Siluet Tour And Travel Kota Malang Menggunakan Metode Triple Exponential Smoothing*, penerbit, jetika., Malang.
- [6] Gorunescu, Florin. 2011. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin.
- [7] Hutabarat, 2018, *Penerapan Datamining untuk memprediksi Permintaan Produk Kartu Perdana Internet Menggunakan Algoritma C5.0 (Studi Kasus: Vidha Ponsel)*, *Jurnal Pelita Informatika*, ISSN 2301-9425 (Media Cetak), Vol April 2018, Medan.
- [8] Han, J., & Kamber, M. 2006. *Data Mining Concept and Tehniques*. San Fransisco.
- [9] Linda Maulida, 2018, *Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggul DI Prov. DKI Jakarta Dengan K-Means*, JISka, Banten.
- [10] Nanda Agung Putra dkk, 2017, *Klasifikasi Sepeda Motor Berdasarkan Karakteristik Konsumen Dengan Metode K-Nearest neighbour Pada Big data Menggunakan Hadoop Single Node Cluster*, JTIK, Malang.
- [11] Nugroho dan Emiliyawati, 2017, *Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest*, *Jurnal Teknik Eliktro*, Semarang.
- [12] Fadli, S., & Imtihan, K. (2018). *ANALISIS DAN PERANCANGAN SISTEM ADMINISTRASI DAN TRANSAKSI BERBASIS CLIENT SERVER*. *Jurnal Informatika dan Rekayasa Elektronik*, 1(2), 7-14.
- [13] Lombok, K. I. S. (2015). *Perencanaan Strategi Sistem Informasi Pendidikan Pada Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Lombok*. *Bianglala Informatika*, 3(2).
- [14] Putra dan Wadisman, 2018, *Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means*, *Intecom*, e-ISSN : 2614-1574, Vol Maret 2018, Sumatra Selatan.
- [15] R. N. Whidhiasih, N. A. Wahanani dan Supriyanto. 2013. *Klasifikasi buah belimbing berdasarkan citra red-greenblue menggunakan knn dan lda*, *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic* 1(1) : 29-35.
- [16] Sari dan Hartama, 2018, *Data Mining: Algoritma K-Means Pada Pengelompokkan Wisata Asing ke Indonesia Menurut Provinsi*, Sensasi, Yogyakarta.
- [17] Suryani dkk, 2017, *Pemilihan Paket Wisata Menggunakan Metode Analytical Hierarchy Process (AHP)*, *Jurnal Informatika Mulawarman*, Samarinda.
- [18] Suyanto, 2017, *Data Mining untuk Klasifikasi dan Klasterisasi Data*, *Penerbit Informatika Bandung*, Bandung.
- [19] P rasetyo Eko. 2012. *Data mining konsep dan aplikasi menggunakan matlab*, *ANDI*. Yogyakarta.
- [20] Pitchayadejanant and Nakpathom, 2018, *Data mining approach for arranging and clustering the agro tourism activities in orchard*, *Kasetsart Journal of Social Sciences*, (2018) 407e413, 2018. Thailand
- [21] Lukman, M., Bagye, W., Fahmi, H., & Imtihan, K. (2019). *PEMANFAATAN TEKNOLOGI GOOGLE MAPS API UNTUK APLIKASI PENDETAKSIAN LOKASI RAWAN KRIMINALITAS BERBASIS ANDROID Studi kasus: Desa Ganti dan Desa Mujur, Kecamatan Praya Timur, Kabupaten Lombok Tengah*. *Jurnal Informatika dan Rekayasa Elektronik*, 2(1), 52-59.