

Analisis Kepuasan Pengunjung Menggunakan Metode Random Forest Untuk Wisata Pantai pada Pesawaran

Andika Pratama Yudha¹⁾, Ryan Puji Cahyono^{2*)}

¹Sistem Informasi Akuntansi

²Teknik Komputer

*) ryanpujic@gmail.com

Abstrak

Tingkat Kepuasan Pengunjung Terhadap Pariwisata Pantai Yang Berada Di Pesawaran adalah salah satu pariwisata yang digunakan sebagai sistem komunikasi praktis dan efektif bagi masyarakat yang berkunjung ke pesawaran. Namun demikian, tidak semua orang tau pantai yang berada di pesawaran. Untuk menjembatani komunikasi pada pariwisata tersebut mereka mampu melihat indahnya pantai yang berada di pantai pesawaran. Pada makalah ini dilakukan analisis dan implementasi algoritma klasifikasi Random Forest untuk mengenali Tingkat Kepuasan Pengunjung Terhadap Pariwisata Pantai Yang Berada Di Pesawaran dengan bantuan dengan daftar daftar pariwisata. Data daftar pariwisata tersebut kemudian diolah dengan menggunakan skeleton tracking untuk menghasilkan titik koordinat yang diperlukan. Adapun algoritma klasifikasi Random Forest menghasilkan rata-rata akurasi sebesar 97% mengenali Tingkat Kepuasan Pengunjung Terhadap Pariwisata Pantai Yang Berada Di Pesawaran.

Kata Kunci: Data Mining, Klasifikasi, Algoritma Random Forest.

PENDAHULUAN

Seiring berkembangnya teknologi untuk pengumpulan dan penyimpanan data, kita harus mampu menghasilkan informasi dari data yang ada. Karena ukuran data yang sangat besar, diharapkan dapat diperoleh informasi yang lebih penting.

Salah satu cara untuk mengekstrak informasi dan pola dari kumpulan data yang besar adalah dengan menggunakan teknik data mining (Rahmanto & Hotijah, 2020; Sarasvananda et al., 2021). Data mining mencakup banyak teknik untuk menghasilkan informasi atau pola dari sekumpulan data. Salah satu teknik tersebut adalah klasifikasi. Random Forest adalah algoritma klasifikasi yang sangat akurat (Chen et al., 2019; Rahat et al., 2019). Sebuah hutan acak adalah teknik ensemble yang terdiri dari beberapa pohon keputusan sebagai pengklasifikasi. Kelas yang dihasilkan dari proses klasifikasi ini diambil dari sebagian besar kelas yang dihasilkan oleh pohon keputusan di hutan acak (Astuti, 2017; Dina Sunia, Kurniabudi, 2019; Rahayu & Purnama, 2022). Pemungutan suara pada pohon keputusan yang tersedia meningkatkan akurasi hutan acak.

KAJIAN PUSTAKA

Metode Random Fores

Seiring berkembangnya teknologi untuk akuisisi dan penyimpanan data, kita harus mampu menghasilkan informasi dari data yang ada. Karena ukuran data yang sangat besar, diharapkan dapat diperoleh informasi yang lebih penting (An'ars, 2022a; Hendrastuty, 2021a, 2021b).

Salah satu cara untuk mengekstrak informasi dan pola dari kumpulan data yang besar adalah dengan menggunakan teknik data mining. Ada banyak teknik. Metode Random Forest (RF) merupakan metode yang dapat meningkatkan akurasi hasil karena pembuatan simpul anak dilakukan secara acak untuk setiap simpul (Ahmad et al., 2018; Alim et al., 2020; Isnain et al., 2021).

Prosedur ini digunakan untuk membangun sebuah pohon keputusan yang terdiri dari node root, inner dan leaf dengan cara mengambil atribut dan data secara random sesuai dengan peraturan yang berlaku. Root node adalah node yang berada paling atas dari pohon keputusan, atau biasa disebut dengan root (Alita et al., 2021; Rahmanto, 2021; Styawati et al., 2020).

Node internal adalah node percabangan, yang memiliki setidaknya dua pintu keluar dan hanya satu pintu masuk. Sebuah simpul daun atau simpul terminal adalah simpul terakhir dengan hanya satu pintu masuk dan tidak ada jalan keluar (An'ars, 2022b, 2022c; Hendrastuty, 2021c). Pohon keputusan dimulai dengan menghitung nilai entropi sebagai penentu tingkat polusi atribut dan nilai perolehan informasi. Untuk menghitung nilai entropi, gunakan rumus seperti pada rumus 1, tetapi nilai perolehan informasi menggunakan rumus 2 (An'ars, 2022d; Hendrastuty, 2021d; Nabila, Isnain, & Permata, 2021a). data mining untuk menghasilkan informasi atau pola dari kumpulan data, Salah satu teknik tersebut adalah klasifikasi (Ali et al., 2020; Nabila, Isnain, & Permata, 2021b).

Random Forest adalah algoritma klasifikasi yang sangat akurat. Sebuah hutan acak adalah teknik ensemble yang terdiri dari beberapa pohon keputusan sebagai pengklasifikasi (Alita & Isnain, 2020; Devika et al., 2019). Kelas yang dihasilkan dari proses klasifikasi ini diambil dari sebagian besar kelas yang dihasilkan oleh pohon keputusan di hutan acak. Pemungutan suara pada pohon keputusan yang tersedia meningkatkan akurasi hutan

acak(Bakri, 2017; Nabila, Isnain, Permata, et al., 2021; Nabila, Rahman Isnain, et al., 2021).

$$Entropy(Y) = - \sum p(c|Y) \log_2 p(c|Y)$$

Dimana Y merupakan kasus himpunan dan $p(c|Y)$ merupakan proporsi nilai Y terhadap kelas c.

$$Information\ Gain(Y, a)$$

$$= Entropy(Y) - \sum$$

$$|Yv|$$

$$|Ya|$$

di mana Nilai(a) adalah semua nilai yang mungkin untuk himpunan a. Yv adalah subclass dari Y dan kelas v sesuai dengan kelas a. Ya, semua nilai yang sesuai dengan a.

RapidMiner

RapidMiner merupakan sebuah solusi untuk melakukan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik(Budiman & Sidiq, n.d.; Darwis et al., 2017; Mardinata & Khair, 2017).

RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operas(Ichsan et al., 2020; Susanto & Puspaningrum, 2019; Wongkar & Angdresey, 2019)i.

METODE

Data yang digunakan dalam penelitian ini diperoleh dengan memanfaatkan perangkat sensor kinect serta perangkat lunak Weka dimana sensor kinect digunakan sebagai media masukan untuk menangkap citra atau gambar serta kedalaman. Selanjutnya, data tersebut diolah hingga menghasilkan koordinat skeleton dan sudut diantara dua vektor yang terdapat pada tubuh menggunakan skeleton tracking(Edhy Sunanta, 2013; Ferdiana, 2020). Data yang diperoleh selanjutnya diuji menggunakan sebuah paket tools machine learning praktis Weka untuk mendapatkan nilai akurasi dari algoritma klasifikasi Random Forest(Damuri et al., 2021; Dr. T. Senthil Kumar, 2020; Marlina & Bakri, 2021).

HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang digunakan adalah data survey yang diperoleh dari kepuasan pengunjung terhadap wisata pantai di Pesawaran dan dapat dilihat pada gambar di bawah ini:

Gambar 1 Pengumpulan Data

Pemrosesan Data

Pra-pemrosesan data dilakukan dengan menghapus nama stempel waktu, nama pengguna, nama lengkap, alamat, jenis kelamin, usia, pekerjaan, tempat wisata pantai yang dikunjungi dan deskripsi singkat mengapa Anda tertarik untuk jalan-jalan ON • The Beach telah memutuskan bahwa Anda akan berkunjung ke pantai lagi kedepannya, bagaimana sikap petugas dalam memberikan pelayanan.

	Bagaimana waktu Dalam Melakukan Pelayanan	Bagaimana Fasilitas Di pantai Tersebut	Bagaimana Keamanan Di Pantai Tersebut	Bagaimana Biaya Di pantai Tersebut	Bagaimana Pemadangan Di Pantai Tersebut	Bagaimana Kebersihan Dipantai Tersebut
0	1	0	1	0	1	1
1	1	0	0	2	1	4
2	1	0	1	0	0	1
3	1	0	1	0	1	1
4	1	1	1	2	2	0
...
113	1	1	1	0	2	1
114	1	0	1	2	1	1
115	1	1	1	2	0	0
116	0	1	0	2	0	2
117	1	0	0	2	1	1

118 rows x 6 columns

```

0 0
1 0
2 0
3 0
4 0
...
113 0
114 0
115 0
116 0
117 0
Name: Apakah Anda Puas Berada Di Pantai Tersebut, length: 118, dtype: int64

```

Gambar 2 Pemrosesan Data

Validasi dan Pengujian

Pengecekan dilakukan dengan validasi silang. Salah satu jenis validasi silang adalah validasi silang 10 kali lipat. Berikut adalah matriks kebingungan dan hasil validasi silang 10 kali lipat menggunakan Python:

```
[ ] predictions = rfore.predict(x_test)
```

```
▶ from sklearn.metrics import confusion_matrix
  print(confusion_matrix(y_test, predictions))
```

```
↳ [[34  0]
    [ 1  1]]
```

```
[ ] from sklearn.metrics import classification_report
  print(classification_report(y_test, predictions))
```

	precision	recall	f1-score	support
0	0.97	1.00	0.99	34
1	1.00	0.50	0.67	2
accuracy			0.97	36
macro avg	0.99	0.75	0.83	36
weighted avg	0.97	0.97	0.97	36

Gambar 3 Hasil Dan Validasi

Hasil dari pengujian menggunakan metode Ten-Fold Cross Validation diperoleh nilai akurasi 97%

KESIMPULAN

Dari penelitian yang dilakukan untuk menganalisis kepuasan pengunjung menggunakan metode random forest pada wisata pantai di Pesawalan, dapat diambil kesimpulan sebagai berikut.

1. Random Forest adalah algoritma yang sangat efektif untuk klasifikasi. Hal ini dapat dilihat dari plot persentase kesalahan dibandingkan dengan algoritma klasifikasi lainnya. Hutan Acak menunjukkan persentase kesalahan yang bersaing dengan algoritme lain, dengan beberapa data menunjukkan persentase kesalahan terendah
2. Random forest memiliki kelemahan dalam hal kestabilan akurasi hasil. Mengingat input dan parameter data yang sama, beberapa proses berturut-turut akan memiliki akurasi yang berbeda. Hal ini terkait dengan fungsi acak yang dihasilkan untuk mengacak pemilihan baris data dan pemilihan pemecah atribut kandidat.

REFERENSI

- Ahmad, I., Sulistiani, H., & Saputra, H. (2018). The Application Of Fuzzy K-Nearest Neighbour Methods For A Student Graduation Rate. *Indonesian Journal of Artificial Intelligence and Data Mining*, 1(1), 47–52.
- Ali, Z. M., Hassoon, N. H., Ahmed, W. S., & Abed, H. N. (2020). The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification.

- International Journal of Psychosocial Rehabilitation*, 24(03), 2143–2151.
<https://doi.org/10.37200/ijpr/v24i3/pr200962>
- Alim, S., Lestari, P. P., & Rusliyawati, R. (2020). Sistem Pakar Diagnosa Penyakit Tanaman Kakao Menggunakan Metode Certainty Factor Pada Kelompok Tani Pt Olam Indonesia (Cocoa) Cabang Lampung. *Jurnal Data Mining Dan Sistem Informasi*, 1(1), 26–31.
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50–58.
- Alita, D., Sari, I., Isnain, A. R., & Styawati, S. (2021). Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 17–23.
- An'ars, M. G. (2022a). Sistem Informasi Manajemen Berbasis Key Performance Indicator (KPI) dalam Mengukur Kinerja Guru. *Jurnal Data Mining Dan Sistem Informasi*, 3(1), 8–18.
- An'ars, M. G. (2022b). Sistem Informasi Manajemen Berbasis Key Performance Indicator (KPI) dalam Mengukur Kinerja Guru. *Jurnal Data Mining Dan Sistem Informasi*, 3(1), 8–18.
- An'ars, M. G. (2022c). Sistem Informasi Manajemen Berbasis Key Performance Indicator (KPI) dalam Mengukur Kinerja Guru. *Jurnal Data Mining Dan Sistem Informasi*, 3(1), 8–18.
- An'ars, M. G. (2022d). Sistem Informasi Manajemen Berbasis Key Performance Indicator (KPI) dalam Mengukur Kinerja Guru. *Jurnal Data Mining Dan Sistem Informasi*, 3(1), 8–18.
- Astuti, F. (2017). Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma Hard C-Means. *Data Manajemen Dan Teknologi Informasi*, 18(1), 64–69.
- Bakri, M. (2017). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Vol*, 11, 1–4.
- Budiman, F., & Sidiq, M. (n.d.). *RANCANG BANGUN APLIKASI SISTEM INFORMASI APLIKASI DATA PETAMBAK*.
- Chen, W., Yan, X., Zhao, Z., Hong, H., Bui, D. T., & Pradhan, B. (2019). Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China). *Bulletin of Engineering Geology and the Environment*, 78(1), 247–266.
- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *Jurnal Riset Komputer*, 8(6), 219–225. <https://doi.org/10.30865/jurikom.v8i6.3655>
- Darwis, D., Wamiliana, W., & Junaidi, A. (2017). Proses Pengamanan Data Menggunakan Kombinasi Metode Kriptografi Data Encryption Standard dan Steganografi End Of File. *Prosiding Seminar Nasional METODE KUANTITATIF 2017*, 1(1), 228–240.
- Devika, R., Avilala, S. V., & Subramaniaswamy, V. (2019). Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 679–684.
- Dina Sunia, Kurniabudi, P. A. J. (2019). Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means. *Jurnal Ilmiah Mahasiswa Teknik Informatika, Vol 1 No 2(2016)*, 121–134.
- Dr. T. Senthil Kumar. (2020). Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm. *Journal of Artificial Intelligence and Capsule Networks*, 2(3), 185–193. <https://doi.org/10.36548//jaicn.2020.3.006>
- Edhy Sunanta. (2013). *Model Integrasi Database Penduduk Indonesia dengan Berbagai Sistem Informasi Berbasis Komputer*. Akprind.
- Ferdiana, R. (2020). A Systematic Literature Review of Intrusion Detection System for Network Security: Research Trends, Datasets and Methods. *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 1–6.

- Hendrastuty, N. (2021a). Rancang Bangun Aplikasi Monitoring Santri Berbasis Android (Studi Kasus: Pesantren Nurul Ikhwan Maros). *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 21–34.
- Hendrastuty, N. (2021b). Rancang Bangun Aplikasi Monitoring Santri Berbasis Android (Studi Kasus: Pesantren Nurul Ikhwan Maros). *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 21–34.
- Hendrastuty, N. (2021c). Rancang Bangun Aplikasi Monitoring Santri Berbasis Android (Studi Kasus: Pesantren Nurul Ikhwan Maros). *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 21–34.
- Hendrastuty, N. (2021d). Rancang Bangun Aplikasi Monitoring Santri Berbasis Android (Studi Kasus: Pesantren Nurul Ikhwan Maros). *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 21–34.
- Ichsan, A., Najib, M., & Ulum, F. (2020). Rancang Bangun Rekomendasi Penerima Bantuan Sosial Berdasarkan Data Kesejahteraan Rakyat. *Jurnal Teknologi Dan Sistem Informasi*, 1(2), 71–79.
- Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 31–37.
- Mardinata, E., & Khair, S. (2017). *Membangun Sistem Informasi Pengelolaan Data Nasabah*. 17(1), 27–35.
- Marlina, D., & Bakri, M. (2021). PENERAPAN DATA MINING UNTUK MEMPREDIKSI TRANSAKSI NASABAH DENGAN ALGORITMA C4. 5. *Jurnal Teknologi Dan Sistem Informasi*, 2(1), 23–28.
- Nabila, Z., Isnain, A. R., & Permata, P. (2021a). Mining Data Analysis for Clustering of Covid-19 Case in Lampung Province Using K-Means Algorithm. *The 1st International Conference on Advanced Information Technology and Communication (IC-AITC)*.
- Nabila, Z., Isnain, A. R., & Permata, P. (2021b). Mining Data Analysis for Clustering of Covid-19 Case in Lampung Province Using K-Means Algorithm. *The 1st International Conference on Advanced Information Technology and Communication (IC-AITC)*.
- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108.
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTSI>
- Rahat, A. M., Kahir, A., & Masum, A. K. M. (2019). Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset. *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 266–270.
- Rahayu, S., & Purnama, J. J. (2022). KLASIFIKASI KONSUMSI ENERGI INDUSTRI BAJA MENGGUNAKAN TEKNIK DATA MINING. *Jurnal Teknoinfo*, 16(2), 395–407.
- Rahmanto, Y. (2021). RANCANG BANGUN SISTEM INFORMASI MANAJEMEN KOPERASI MENGGUNAKAN METODE WEB ENGINEERING (Studi Kasus: Primkop Kartika Gatam). *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 24–30.
- Rahmanto, Y., & Hotijah, S. (2020). PERANCANGAN SISTEM INFORMASI GEOGRAFIS KEBUDAYAAN LAMPUNG BERBASIS MOBILE. *Jurnal Data Mining Dan Sistem Informasi*, 1(1), 19–25.
- Sarasvananda, I. B. G., Anwar, C., Pasha, D., & Styawati, S. (2021). ANALISIS SURVEI KEPUASAN MASYARAKAT MENGGUNAKAN PENDEKATAN E-CRM (Studi Kasus: BP3TKI Lampung). *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 1–9.
- Styawati, S., Yulita, W., & Sarasvananda, S. (2020). SURVEY UKURAN KESAMAAN SEMANTIC ANTAR KATA. *Jurnal Data Mining Dan Sistem Informasi*, 1(1), 32–37.
- Susanto, E. R., & Puspaningrum, A. S. (2019). *Rancang Bangun Rekomendasi Penerima Bantuan Sosial Berdasarkan Data Kesejahteraan Rakyat*. 15(1), 1–12.

Wongkar, M., & Angdresey, A. (2019). Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter. *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 1–5.