



IDENTIFIKASI POLA DAN JENIS HOTEL MENGGUNAKAN ALGORITMA DBSCAN

S1 Data Sains

Dr. Warih Maharani, S.T., M.T.

Ikhwan Wahyudin (1305210065), Helmi Muzakki K (1305213007), Hijrah Wira P (1305210058)

LATAR BELAKANG

Industri perhotelan di Indonesia tumbuh pesat seiring pariwisata dan ekonomi. Keanekaragaman hotel yang meningkat menimbulkan tantangan bagi wisatawan dalam memilih akomodasi yang tepat. Banyaknya pilihan dan terbatasnya informasi sering mempersulit keputusan, sehingga pengalaman menginap bisa kurang memuaskan dan mempengaruhi keseluruhan perjalanan.

TUJUAN

Mengembangkan sebuah sistem yang mampu mengidentifikasi pola dan jenis hotel secara efisien menggunakan algoritma DBSCAN. Dengan menggunakan pendekatan klasterisasi yang adaptif dan fleksibel, untuk menyediakan solusi yang dapat membantu pengguna atau wisatawan dalam menemukan pilihan hotel yang sesuai dengan preferensi dan kebutuhan mereka.

METODE

DBSCAN

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

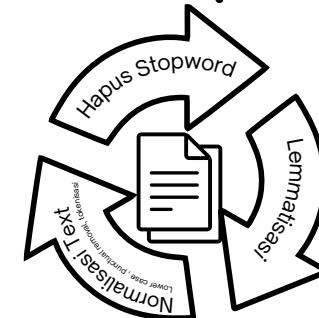
Silhouette Index

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

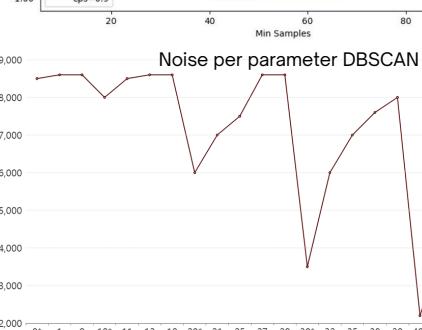
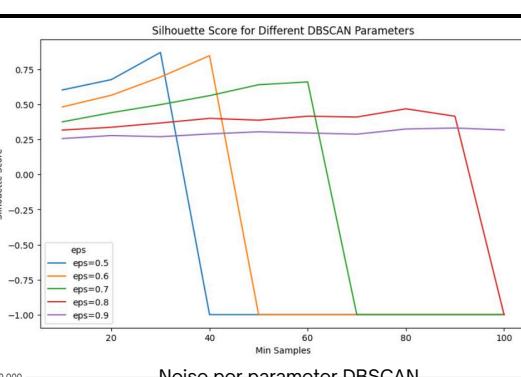
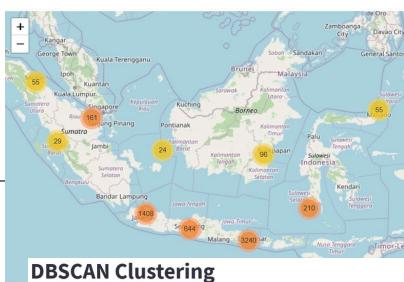
TF-IDF

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

Text Preprocessing



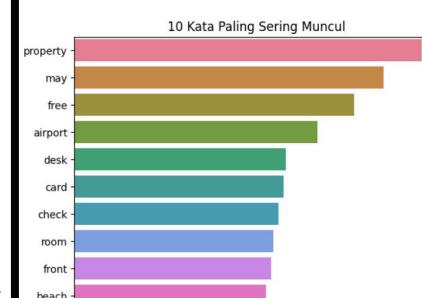
OUTPUT



DATASET

8767 Baris X 16 Kolom Data

Dataset yang digunakan Hotels_Dataset.csv dari situs kaggle.



DISKUSI

- Interpretasi Hasil : Beberapa hasil klaster berhasil diidentifikasi.
- Nilai Silhouette Score : Beberapa percobaan kombinasi parameter ada yang menunjukkan nilai lebih rendah.
- Tantangan : Menentukan parameter eps dan min_sample yang optimal menjadi tantangan, diperlukan kombinasi parameter agar mendapatkan klaster terbaik.
- Potensi Pengembangan : Untuk lebih lanjut bisa di integrasikan dengan sistem rekomendasi dan teknik klaster lain agar lebih bagus.

KESIMPULAN

Algoritma DBSCAN berhasil mengelompokkan hotel berdasarkan fitur teks dan numerik, parameter optimal dengan Silhouette Index terbaik melalui kombinasi eps dan min_samples. Hasil ini memudahkan wisatawan menemukan hotel lain sesuai preferensi mereka.

Referensi :

Nurhaliza, Nana, and Mustakim Mustakim. "Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma DBSCAN: Clustering of Data Covid-19 Cases in the World Using DBSCAN Algorithms." Indonesian Journal of Informatic Research and Software Engineering (IJIRSE) 1.1 (2021): 1-8.