

Visualisasi Data Kawasan Pariwisata yang Sering Dikunjungi di Bali dengan Pendekatan Pencocokan String pada Postingan Akun Instagram

Ida Ayu Putu Ari Crisdayanti / 13515067
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13515067@std.stei.itb.ac.id

Abstrak—Makalah ini membahas gagasan mengenai aplikasi algoritma pencocokan string sebagai pendekatan dalam menentukan kawasan pariwisata yang sering dikunjungi di Bali. Pencocokan string dilakukan terhadap postingan-postingan sejumlah akun instagram yang memuat lokasi dari postingan tersebut. Algoritma pencocokan string yang digunakan adalah Boyer Moore. Pola string yang akan dicocokkan adalah string dari nama-nama kabupaten serta kawasan pariwisata di Bali. Melalui metode pencocokan string, akan diperoleh data kuantitatif kemunculan string nama kawasan pariwisata dari sejumlah postingan akun instagram. Data ini kemudian divisualisasikan ke dalam peta kontur berwarna yang menunjukkan intensitas kunjungan wisatawan terhadap destinasi wisata di masing-masing kawasan.

Kata Kunci—Visualisasi Data, Pariwisata, Pencocokan String, Boyer Moore, Postingan Instagram.

I. PENDAHULUAN

Pulau Bali terkenal dengan berbagai destinasi wisata yang menjadi incaran wisatawan baik domestik maupun mancanegara. Dari destinasi-destinasi tersebut terdapat beberapa kawasan yang paling sering dikunjungi oleh wisatawan. Para wisatawan kebanyakan berasal dari luar provinsi Bali sehingga tidak cukup mengetahui kawasan-kawasan yang banyak dituju sebagai destinasi wisata.

Berdasarkan latar belakang tersebut, makalah ini akan membahas ide visualisasi data kawasan pariwisata yang sering dikunjungi di Bali melalui pendekatan pencocokan string pada postingan akun instagram. Hasil visualisasi diperoleh dari data kuantitatif pencocokan string kawasan pariwisata di Bali dengan lokasi sejumlah postingan akun instagram.

Algoritma pencocokan string yang akan digunakan adalah *Boyer Moore*. Algoritma ini dipilih karena efektif dan efisien dalam menangani pencocokan string yang mengandung banyak variasi alfabet. Ide pendekatan menggunakan lokasi postingan akun instagram dipilih karena cukup mewakili informasi destinasi-destinasi wisata yang dikunjungi wisatawan. Perhitungan kuantitatif dilakukan terhadap jumlah kecocokan string.

Data kuantitatif inilah yang dijadikan dasar dari visualisasi data yang akan dibuat. Visualisasi data dibuat dalam bentuk peta berwarna yang merepresentasikan perbedaan intensitas kunjungan wisatawan terhadap destinasi-destinasi wisata yang terdapat di kawasan tersebut. Visualisasi data bertujuan untuk mempermudah pembaca dalam memahami informasi yang ingin disampaikan.

Penulisan makalah ini bertujuan untuk menyampaikan ide visualisasi data kawasan pariwisata di Bali yang sering dikunjungi dengan pendekatan pencocokan string. Diharapkan informasi ini nantinya dapat bermanfaat bagi wisatawan khususnya non lokal serta pihak-pihak yang ingin mengembangkan analisis dari hasil visualisasi data di bidang pariwisata.

II. LANDASAN TEORI

2.1 Boyer Moore

Algoritma *Boyer-Moore* diperkenalkan oleh Bob Boyer dan J.S. Moore pada tahun 1977. Pada algoritma ini pencocokan kata dimulai dari karakter terakhir kata kunci menuju karakter awalnya. Jika terjadi perbedaan antara karakter terakhir kata kunci dengan kata yang dicocokkan maka karakter-karakter dalam potongan kata yang dicocokkan tadi akan diperiksa satu per satu. Hal ini dimaksudkan untuk mendeteksi apakah ada karakter dalam potongan kata tersebut yang sama dengan karakter yang ada pada kata kunci. Apabila terdapat kesamaan, maka kata kunci akan digeser sedemikian rupa sehingga posisi karakter yang sama terletak sejajar, dan kemudian dilakukan kembali pencocokan karakter terakhir dari kata kunci. Sebaliknya jika tidak terdapat kesamaan karakter, maka seluruh karakter kata kunci akan bergeser ke kanan sebanyak m karakter, di mana m adalah panjang karakter dari kata kunci.

Boyer-Moore merupakan salah satu Algoritma *Pattern Matching* yang cukup terkenal. Algoritma ini menggunakan beberapa kasus pengecekan teks (input karakter yang akan dibaca) dengan *pattern* (pola yang akan disaring). Algoritma *Boyer-Moore* adalah algoritma pencarian string yang mencari dengan cara membandingkan sebuah huruf dengan huruf yang ada di *pattern* yang dicari, dan menggeser *pattern* tersebut hingga posisinya sama dengan teks yang dicari dan membandingkan kata tersebut. Cara ini disebut *character*

jump. Algoritma *string matching* Boyer-Moore ini berbasis pada 2 metode yaitu :

1. *The Looking-Glass Technique*

The Looking-Glass Technique melakukan perbandingan suatu karakter akhir pada kata *w* dengan suatu karakter pada teks *s*. Jika karakter tersebut sama maka jendela karakter akan berjalan mundur pada kedua string dan mengecek kembali kedua karakter. Mencari Suatu kecocokan string pada teks dengan pola yang akan dicari dengan cara memindahkan atau menggesernya sampai Teks string selesai.

2. *The Character-Jump Technique*

Character-jump Technique melakukan suatu aksi ketika perbandingan antara dua karakter yang berbeda. Ada dua aksi yang tergantung pada teks *s* dan kata *w* yang dimiliki; jika *p* yaitu karakter pada *s* yang sedang diproses yang tidak cocok maka ada dua kemungkinan aksi. Mencari karakter yang sesuai dan cara penggeseran sebuah karakter perbandingan terakhir.

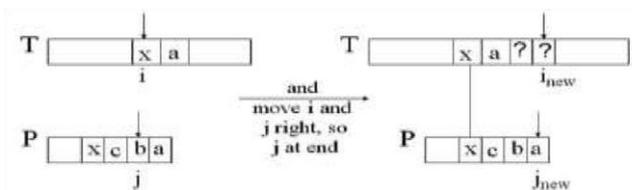
Beberapa kasus yang ada pada algoritma ini antara lain:

- Jika suatu karakter Pola (*P*) mengandung karakter *x* dimana *x* adalah anggota dari Teks yang telah dibandingkan maka perbandingan karakter selanjutnya dimulai karakter *P* yang sama.

Misal :

T = ..xa..??

P = xcba



Gambar 2.1.1 Algoritma *Boyer Moore*^[3]

Pada Gambar 2.1.1 dapat dijelaskan bahwa *T* yang akan dicocokkan dengan *pattern P* adalah kata “*xa*”. Dengan menggunakan algoritma Boyer Moore, pencocokan akan dimulai dari kanan ke kiri sesuai dengan *T* yang diminta yaitu “*xa*” pada *pattern P* “*xcba*”. Jika *T* mengalami ketidakcocokan maka *T* akan melompat sejauh *n* karakter *T* pada *Pattern P*. jika terjadi kecocokan maka proses akan berhenti.

- Jika Perbandingan karakter terakhir pada suatu pola sama dengan teks adalah sama. Maka pergeseran karakter selanjutnya bergeser satu kali.

Misal :

T = ..wax..??

P = cwax

P	C	w	A	x
T	W	a	X	
		w	A	x

Tabel 2.1.1 Kinerja Algoritma *Boyer Moore*^[3]

Pada tabel 2.1.1 dapat dijelaskan bahwa *T* yang akan dicocokkan dengan *pattern P* adalah kata “*cwax*”. Dengan menggunakan algoritma Boyer Moore, pencocokan akan dimulai dari kanan ke kiri sesuai dengan *T* yang diminta yaitu “*wax*” pada *pattern P* “*cwax*”. Jika *T* mengalami ketidakcocokan maka *T* akan melompat sejauh *n* karakter *T* pada *Pattern P* untuk selanjutnya memulai pencocokan. jika terjadi kecocokan (match) maka proses akan berhenti.

Algoritma *Boyer-Moore* ini adalah perbandingan karakter dalam sebuah string yang dilakukan dari belakang ke depan atau dari kanan ke kiri karakter. Jika algoritma *Boyer Moore* membandingkan teks “*MAKALAH*” misalnya, algoritma ini melakukan pengecekan apakah karakter ke tujuh dari teks yang dibandingkan adalah karakter ‘*H*’. Jika karakter ketujuh adalah ‘*H*’ maka ia akan melakukan pengecekan apakah karakter sebelumnya (ke-6) adalah ‘*A*’. Demikian seterusnya hingga menemukan bahwa karakter pertama adalah ‘*M*’.

Alasan kenapa *Boyer-Moore* melakukan pengecekan dari belakang akan lebih jelas jika kita mengamati apa yang terjadi jika pengecekan menghasilkan nilai yang tidak sama. Misalnya, algoritma mendapati karakter ke-8 dengan karakter ‘*X*’ bukannya ‘*H*’. ‘*H*’ tidak muncul sama sekali pada “*MAKALAH*”, dan ini berarti tidak ada kesamaan sama sekali karakter ‘*X*’ dengan semua karakter dalam “*MAKALAH*”. Sehingga, dengan hanya melakukan sekali pengecekan pada karakter ke-8 kita dapat mengabaikan pengecekan karakter ke-1 sampai ke-7 dan langsung melanjutkan pengecekan karakter dimulai dari karakter ke-9, tepat setelah ‘*X*’. Algoritma ini pada awalnya melakukan perhitungan sebuah tabel untuk menentukan banyaknya ‘loncatan’ karakter yang akan dilakukan setelah mendapati sebuah perbandingan yang tidak cocok.

Karakteristik utama dari algoritma *Boyer-Moore* adalah algoritma ini melakukan pencocokan string mulai dari kanan (belakang). dengan karakteristik tersebut, ketidakcocokan saat terjadi perbandingan string akan membuat pergerakan *pattern* melompat lebih jauh untuk menghindari perbandingan karakter pada string yang diperkirakan gagal. Sehingga proses pencarian string akan lebih optimal.

Contoh cara kerja algoritma *Boyer Moore* ini adalah sebagai berikut :



Gambar 2.1.2 Contoh Penggunaan Algoritma *Boyer Moore*^[1]

2.2 Visualisasi Data

2.2.1 Visualisasi, Representasi, dan Seleksi

Visualisasi adalah konversi data ke dalam format visual atau tabel sehingga karakteristik dari data dan relasi diantara item data atau atribut dapat dianalisis atau dilaporkan^[2]. Visualisasi data adalah salah satu teknik yang baik dan menarik untuk eksplorasi data. Manusia memiliki kemampuan membangun yang baik untuk menganalisis sejumlah besar informasi yang dipresentasi secara visual. Ia dapat mendeteksi pola umum atau tren dan pola yang tidak umum.

Representasi adalah pemetaan informasi menjadi format visual. Objek data, atribut-atributnya, dan relasi diantara objek-objek data diterjemahkan ke dalam elemen grafis seperti titik, garis, bentuk-bentuk tertentu, dan warna. Contohnya objek-objek sering direpresentasikan sebagai titik. Nilai atribut-atributnya dapat direpresentasikan sebagai posisi dari titik-titik atau karakteristik dari titik sebagai contoh warna, ukuran dan bentuk. Jika posisi digunakan, maka relasi pada titik, apakah terbentuk dalam group atau sebuah titik dapat dengan mudah dilihat.

Seleksi adalah eliminasi atau penekanan kembali dari beberapa objek tertentu dan atribut-atribut. Seleksi juga merupakan pemilihan subset dari atribut. Reduksi secara dimensi sering digunakan untuk mengurangi jumlah dimensi menjadi 2 atau 3 dimensi. Alternatifnya, sepasang atribut dapat dipertimbangkan. Seleksi dapat juga merupakan pemilihan subset dari objek. Sebagian wilayah dari layar hanya dapat menunjukkan banyak titik.

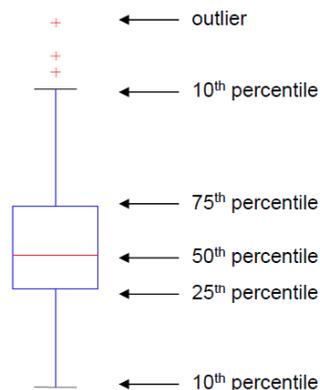
2.2.2 Teknik Visualisasi

(1) Histogram

Histogram biasa digunakan untuk menunjukkan distribusi dari nilai dari variabel tunggal. Histogram membagi nilai ke dalam bin dan menunjukkan batang plot dari sejumlah objek dalam setiap bin. Tinggi dari setiap batang menunjukkan jumlah dari objek. Bentuk histogram tergantung dari jumlah bin.

(2) Box Plot

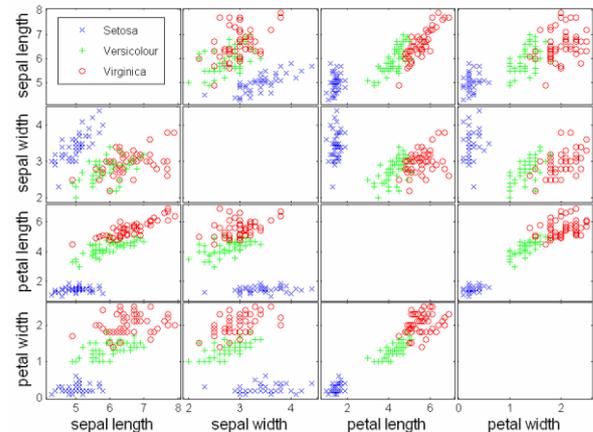
Box plot ditemukan oleh J. Tukey. *Box plot* merupakan cara alternatif untuk menggambarkan distribusi data. Gambar berikut ini menunjukkan bagian dasar dari *box plot*.



Gambar 2.2.1 Dasar box plot^[2]

(3) Scatter Plot

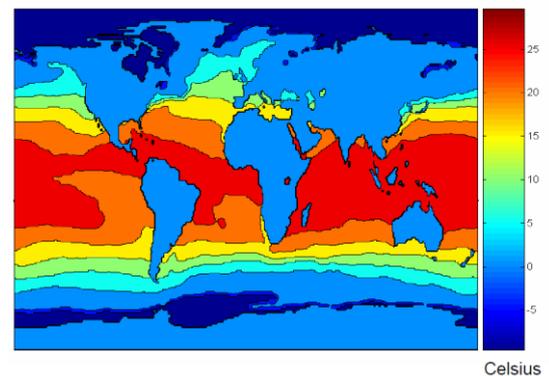
Scatter plot merupakan nilai-nilai atribut yang menentukan posisi. *Scatter plot* dua dimensi adalah bentuk yang paling umum, tapi dapat juga dibuat dalam *scatter plot* tiga dimensi. Kadangkala atribut tambahan dapat digambarkan dengan menggunakan ukuran, bentuk dan warna dari tanda yang mewakili objek. Akan sangat berguna jika kita memiliki array dari *scatter plot* yang secara terpadu merangkum hubungan dari beberapa pasang atribut.



Gambar 2.2.2 Array scatter plot dari atribut iris^[2]

(4) Contour Plot

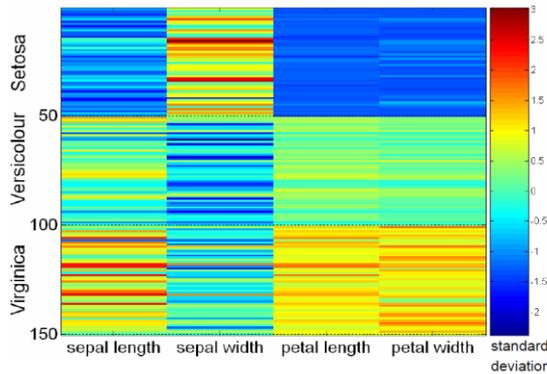
Contour plot berguna jika atribut kontinu diukur dengan garis. *Contour plot* membagi ke dalam daerah-daerah yang nilainya sama. Garis *contour* yang membentuk batas dari daerah/wilayah menghubungkan titik-titik dengan nilai yang sama. Contoh yang paling umum adalah *contour maps of elevation*. *Contour plot* juga menggambarkan suhu, curah hujan, tekanan udara dsb.



Gambar 2.2.3 Contour Plot: Sea Surface Temperature December 1998^[2]

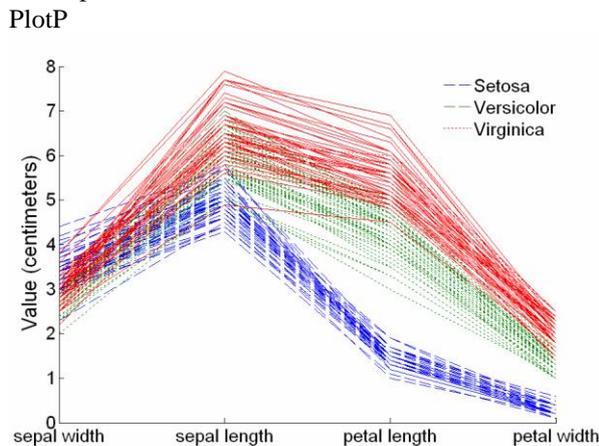
(5) Matrix Plot

Matriks plot dapat menempatkan/plot data matriks. Matriks plot ini dapat berguna pada saat objek diurut menurut kelas. Biasanya atribut-atribut dinormalisasikan untuk mencegah satu atribut dari dominasi plot. Plot dari kesamaan atau jarak matriks dapat juga menjadi berguna untuk menggambarkan hubungan diantara objek-objek.



Gambar 2.2.4 Visualisasi dari matriks data iris^[2]

- (6) Koordinat Paralel
 Koordinat paralel berguna untuk menempatkan nilai atribut dari data dimensi tinggi. Daripada menggunakan *perpendicular axes*, maka digunakan *parallel axes*. Nilai atribut untuk setiap objek di-plot sebagai titik pada setiap sumbu koordinat yang berhubungan dan titik-titik dihubungkan dengan garis. Kemudian, setiap objek direpresentasikan sebagai garis. Seringkali, garis merepresentasikan kelas yang berbeda dari group objek bersama, paling tidak untuk beberapa atribut. Pengurutan atribut adalah penting mengingat kondisi pengelompokan (*group*) seperti di atas.

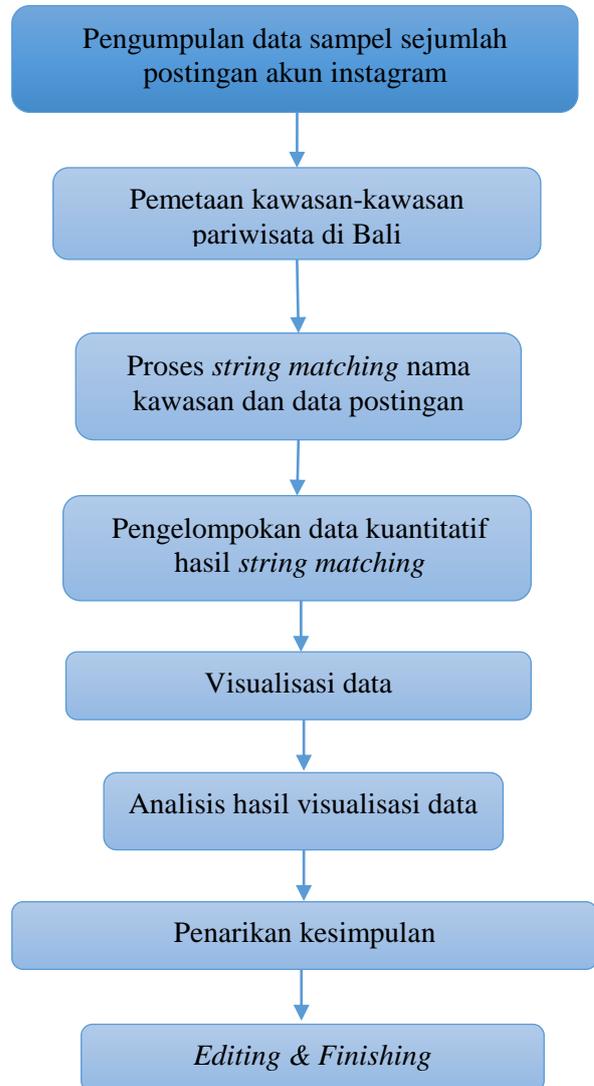


Gambar 2.2.5 Plot koordinat paralel untuk data iris^[2]

- (7) *Star Plot*
 Teknik visualisasi dengan *star plot* pendekatannya sama dengan koordinat paralel, namun sumbu menyebar dari titik tengah. Garis yang menghubungkan nilai dari objek disebut dengan poligon.
- (8) Chernoff Faces
 Pendekatan dengan *Chernoff faces* diciptakan oleh Herman Chernoff. Pendekatan ini berasosiasi dengan setiap atribut dengan karakteristik dari setiap wajah. Nilai dari setiap atribut menentukan penampakan dari hubungan karakteristik wajah. Setiap objek menjadi wajah yang terpisah. *Chernoff faces* ini tergantung pada kemampuan manusia dalam membedakan beberapa wajah.

III. METODOLOGI

Dalam pengerjaan makalah yang berjudul “Visualisasi Data Kawasan Pariwisata yang Sering Dikunjungi di Bali dengan Pendekatan Pencocokan String pada Postingan Akun Instagram”, terdapat beberapa tahapan seperti pada diagram berikut.



Gambar 3.1 Diagram pengerjaan makalah

Pengerjaan makalah diawali dengan pengumpulan sampel data sejumlah postingan akun instagram. Jenis postingan yang diutamakan adalah yang mengandung lokasi foto pada postingan. Data dapat diperoleh dengan pengumpulan secara manual atau melalui *data scraping* yaitu salah satu bidang ilmu dalam basis data. Setelah itu dilakukan pemetaan kawasan pariwisata di Bali yang akan dijadikan dasar pencocokan string dan pengelompokan data. Dari hasil pencocokan string antara nama kawasan dan data lokasi postingan akun instagram akan diperoleh data kuantitatif kemunculan string setiap kawasan. Data ini lalu divisualisasikan ke dalam peta kontur untuk kemudian dianalisis.

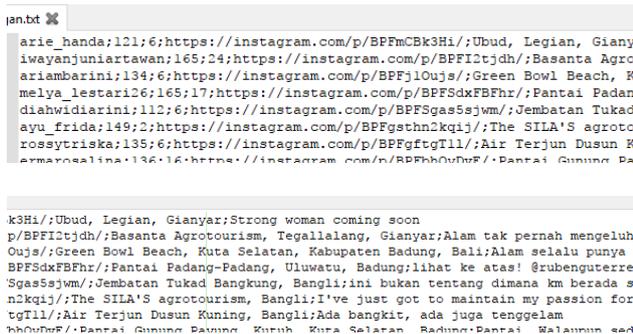
IV. PROSES PENCOCOKAN STRING

Pada proses pencocokan string, terlebih dahulu ditentukan pola string yang akan dicocokkan. String yang dipilih adalah nama-nama kabupaten di provinsi Bali untuk menyederhanakan pemetaan kawasan pariwisata. Berikut adalah nama-nama kabupaten dan ibu kota provinsi Bali.

1. Kabupaten Badung
2. Kabupaten Bangli
3. Kabupaten Buleleng
4. Kabupaten Gianyar
5. Kabupaten Jembrana
6. Kabupaten Karangasem
7. Kabupaten Klungkung
8. Kabupaten Tabanan
9. Kota Denpasar

Selanjutnya dilakukan pencocokan string dengan memanfaatkan algoritma Boyer Moore. Algoritma ini dipilih karena efektif dan efisien dalam menangani pencocokan string dengan banyak variasi karakter. Namun akibatnya, pencocokan string disini bersifat *exact matching*. Berbeda dengan metode *regular expression* yang mampu menangani berbagai kemungkinan kemunculan string yang dianggap sama dengan pola string yang tepat. Namun hal tersebut masih bisa ditangani dengan mengikutsertakan berbagai string yang ekuivalen dengan nama-nama kawasan pariwisata yang dipilih.

String yang akan dicocokkan dengan *string pattern* disimpan dalam sebuah file eksternal seperti pada gambar berikut.



Gambar 4.1 File eksternal sampel postingan

File eksternal berisi informasi postingan sejumlah akun instagram. Gambar di atas adalah *prototype* data postingan akun yang memuat lokasi diambilnya foto atau video. Untuk memperoleh jumlah data yang besar dapat digunakan metode *data scraping*. *Data scraping* adalah suatu teknik penggalian informasi dari situs web dengan cara mengekstrak data dari situs tertentu. Hasil dari *data scraping* dapat berupa file json atau csv. Berikut adalah contoh hasil *data scraping*.

username	likes	comments	link	
vapesale	39	1	https://instagram.com/p/5Z9mCBk3HL/	Refreshing @_vichitvongs901
van_city_reviews	13	0	https://instagram.com/p/5Z9j1QoeI_/	#throwbackto the fat kief bowl I sr
naughtynnic1	1	0	https://instagram.com/p/5Z9j1OLuqG/	Your one stop shop. Come stop in at so make sure to check those out. An prices y'all! #dontmissout #yous #igititfrommymama
royal_vaporz	27	4	https://instagram.com/p/5Z9dwtOTGs/	
naughtynnic1	2	0	https://instagram.com/p/5Z9dXrrup6/	Your one stop shop. Come stop in at so make sure to check those out. An prices y'all! #dontmissout #yous #igititfrommymama
wevaperehere_com	15	2	https://instagram.com/p/5Z9bhQyDvE/	Want to win this custom #B1 Sigelei 2) Follow @theschwartzliquid 3) H yo
naughtynnic1	2	0	https://instagram.com/p/5Z9MopLupW/	Your one stop shop. Come stop in at so make sure to check those out. An prices y'all! #dontmissout #yous #igititfrommymama
gforce_vapor	38	0	https://instagram.com/p/5Z88JICRgY/	Loving this edit of the shot @theinx #vapefam #vapeon #vapestz #notblowingsmoke #theEARTHrr Mirral KS - Di 1BF SII VFR Wirin

Gambar 4.2 Contoh data scraping^[3]

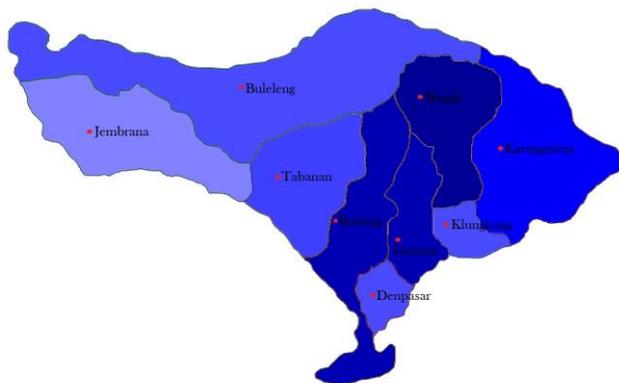
Setelah informasi yang dibutuhkan disimpan dalam file eksternal, file tersebut kemudian dibaca melalui sebuah program. Program tersebut sekaligus melakukan pencocokan string antara informasi pada file eksternal dan *string pattern* yang dipilih. Dari hasil pencocokan string akan diperoleh data kuantitatif kemunculan setiap string nama kawasan. Berikut adalah contoh hasil pencocokan string dari beberapa sampel postingan akun instagram yang digunakan.



Gambar 4.3 Hasil Pencocokan String Sampel Postingan

V. PROSES VISUALISASI DATA

Setelah memperoleh data kuantitatif kemunculan nama-nama string kawasan melalui pencocokan string, dilakukan visualisasi data. Model visualisasi yang dipilih adalah kontur plot. Kontur plot akan memetakan pulau Bali ke dalam beberapa kawasan pariwisata yang dipilih. Visualisasi kontur plot dilakukan dengan memetakan peta pulau Bali ke dalam beberapa warna berbeda yang menggambarkan perbedaan intensitas kunjungan wisatawan di kawasan tersebut. Berikut ini adalah gambaran kontur plot dari hasil visualisasi data kuantitatif.



Gambar 5.1 Peta Kontur Intensitas Kunjungan Wisatawan

Dari hasil visualisasi di atas dapat terlihat perbedaan intensitas kunjungan wisatawan di tiap kawasan. Warna yang lebih gelap menggambarkan intensitas kunjungan yang lebih tinggi. Tentunya visualisasi data ini hanya sebuah *tool* untuk menganalisis perbedaan intensitas kunjungan wisatawan di tiap kawasan. Perbedaan intensitas ini dapat disebabkan oleh beberapa faktor seperti jarak kawasan pariwisata dari ibu kota provinsi, akses menuju destinasi-destinasi wisata yang ada di kawasan tersebut, promosi destinasi wisata yang ada di kawasan tersebut, dan lain-lain.

Melalui visualisasi data intensitas kunjungan wisatawan di kawasan pariwisata Bali, diharapkan dapat memberikan informasi atau pengetahuan baru bagi wisatawan yang belum banyak mengenal destinasi-destinasi wisata yang ramai dikunjungi. Sementara bagi pemerintah daerah, diharapkan produk dari gagasan ini dapat menjadi tolok ukur dalam mengembangkan potensi-potensi pariwisata di Bali. Dari hasil visualisasi data dapat dianalisis lebih lanjut penyebab rendahnya intensitas kunjungan wisatawan di kawasan tertentu.

Jika disebabkan oleh akses yang kurang memadai, maka pemerintah perlu membenahi akses menuju destinasi-destinasi wisata di kawasan tersebut. Jika disebabkan oleh kurangnya pengetahuan wisatawan tentang kawasan tertentu, maka perlu dilakukan promosi yang lebih gencar mengenai destinasi-destinasi wisata di kawasan tersebut dan lain-lain.

VI. KESIMPULAN

Visualisasi data kawasan pariwisata yang sering dikunjungi di Bali dapat dilakukan dengan pendekatan pencocokan string pada postingan akun instagram. Algoritma pencocokan string Boyer Moore cukup efektif dan efisien untuk digunakan karena variasi alfabet yang besar. Proses pencocokan string dapat dimaksimalkan dengan menggunakan *regular expression*. Dari hasil pencocokan string akan diperoleh data kuantitatif yang divisualisasikan menjadi peta kontur serta mampu menggambarkan perbedaan intensitas kunjungan masing-masing kawasan.

VII. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada berbagai pihak yang telah memberikan dukungan serta bantuan dalam menyelesaikan makalah yang berjudul “Visualisasi Data Kawasan Pariwisata yang Sering Dikunjungi di Bali dengan Pendekatan Pencocokan String pada Postingan Akun Instagram”. Penulis mengucapkan terima kasih kepada Bu Masayu, Bu Ulfa, dan Pak Rinaldi Munir selaku pembimbing mata kuliah Strategi Algoritma.

REFERENSI

- [1] Munir, Rinaldi. 2006. *Strategi Algoritma*. Bandung: Penerbit Informatika.
- [2] wsilfi.staff.gunadarma.ac.id/Downloads/files/4410/visualisasi, diakses pada tanggal 9 Mei 2017 pukul 13.30 WIB.
- [3] nero.trunojoyo.ac.id, diakses pada tanggal 16 Mei 2017 pukul 09.30 WIB.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 16 Mei 2016

Ida Ayu Putu Ari Crisdayanti
13515067