

Analisis Crowdsourcing Dari Media Sosial Untuk Sumber Data Alternatif Kajian Bentang Lahan

Budi Fajar Supriyanto^{1*}, Fatwa Ramdani², Afif Supianto³

¹Program Studi Teknologi Informasi, Fakultas Vokasi, Universitas Brawijaya, Malang

²Department of International Public Policy, Faculty of Humanities and Social Sciences, University of Tsukuba

³Pusat Riset Sains Data dan Informasi, Badan Riset dan Inovasi Nasional, Bandung

¹Jln. Veteran No.12-14, Kota Malang, 65133, Indonesia

²1-1-1 Tennodai, Tsukuba, Ibaraki 305-8574, Japan

³Jl. Sangkuriang/Cisitu No.21/154D Komplek BRIN Gedung 20 Lantai 3, Kota Bandung, 40135, Indonesia

email: ¹fajar.deroom@gmail.com, ²fatwa.ramdani.gw@u.tsukuba.ac.jp, ³ahma063@brin.go.id

Abstract – Increasing the number of social media users from time to time provides new potential in crowdsourcing data acquisition. The data acquisition process no longer requires a lot of cost and time, because crowdsourcing can be extracted easily and even without cost. This study raises the issue of whether crowdsourcing data from social media can be used as alternative data in geo-informatics studies. The data acquisition process is carried out from user uploads on social media. The upload includes the location point and text on the caption. The data obtained is then processed to find out the point of the landscape and the tendency to use language. The use of language was analyzed by the RQDA method and the results obtained were 5.37% speaking about the landscape. While the location point of social media compared to the DEMNAS data has an accuracy score of 437.8 which is validated with RMSE method and isn't represented near to 1.0. Suggested that social media data is still far away to be able as an alternative source for landscape data.

Abstrak – Peningkatan jumlah pengguna media sosial dari waktu ke waktu memberikan potensi baru dalam akuisisi data crowdsourcing. Proses akuisisi data tidak lagi membutuhkan banyak biaya dan waktu, karena crowdsourcing dapat digali dengan mudah bahkan tanpa biaya. Kajian ini mengangkat permasalahan apakah data crowdsourcing dari media sosial dapat dijadikan data alternatif dalam kajian geo-informatika. Proses akuisisi data dilakukan dari unggahan pengguna di media sosial. Unggahan menyertakan titik lokasi dan teks pada keterangan. Data yang diperoleh kemudian diolah untuk mengetahui titik lanskap dan kecenderungan penggunaan bahasa. Penggunaan bahasa dianalisis dengan metode RQDA dan diperoleh hasil 5,37% berbicara tentang bentang alam. Sedangkan titik lokasi media sosial dibandingkan dengan data DEMNAS memiliki skor akurasi 437,8 yang divalidasi dengan metode RMSE dan tidak direpresentasikan mendekati 1,0. Disarankan bahwa data media sosial masih jauh untuk dapat menjadi alternatif sumber data lanskap.

Kata Kunci – Crowdsourcing, Bentang Lahan, Media Sosial, RMSE, DEM, Geoinformatika

*) **penulis korespondensi:** Budi Fajar Supriyanto

Email: fajar.deroom@gmail.com

I. PENDAHULUAN

Media sosial adalah aplikasi yang memungkinkan penggunaannya untuk membuat dan berbagi konten atau terlibat

dalam jejaring sosial[1]. Salah satu fitur di media sosial adalah pengguna dapat berbagi lokasi atau geo-located melalui foto yang diunggah. Fitur geo-located di media sosial terintegrasi dengan peta digital, memungkinkan pengguna lain untuk mengikuti atau mengunjungi lokasi yang sama dengan petunjuk yang dapat diperoleh dari peta digital.

Fitur geo-located merupakan bentuk pengembangan dari proses pencarian manual yang ada dan interpretasi pesan-pesan penting[2]. Meskipun pesan yang disampaikan secara geo-located bersifat sementara, namun hal ini dapat ditelaah lebih dalam. Geo-located memiliki kaitan erat dengan masukan dari pengguna media sosial dan sangat bervariasi. Variasi data disebabkan oleh banyaknya lokasi di media sosial, bahkan pengguna memiliki akses untuk menambahkan lokasi. Akses yang tersedia dapat membentuk nama yang berbeda pada satu titik di lokasi yang sama.

Satu titik lokasi dengan berbagai nama merupakan salah satu bentuk variasi data. Data dari media sosial bisa lebih bervariasi. Properti variabel ini dikenal sebagai crowdsourcing[3].

Variasi data crowdsourcing dari media sosial berasal dari unggahan pengguna. Postingan di unggahan media sosial bisa menjadi representasi dunia nyata[4]. Atau apapun yang mereka inginkan, termasuk data, sentimen, fakta, atau bahkan opini dan apapun yang ingin dituliskan pada caption oleh pengguna[5]. Penelitian sebelumnya tentang crowdsourcing menemukan jawaban atas pertanyaan ketika crowdsourcing menjadi mekanisme yang lebih baik untuk suatu masalah. Pembahasan difokuskan pada kemampuan crowdsourcing untuk menemukan jarak terdekat ke suatu lokasi. Penelitian ini merupakan studi literatur sehingga data mining merupakan kecenderungan dari setiap makalah yang menjadi acuan. Kesimpulannya adalah bahwa crowdsourcing dapat digunakan sebagai mekanisme penyelesaian masalah untuk batasan-batasan tertentu dan dengan batasan-batasan yang ketat, tetapi batasan-batasan apa yang tidak diketahui dan untuk topik apa[6].

Peran crowdsourcing dalam penelitian telah digunakan. Seperti Rogstadius & Kostakos [7] meneliti crowdsourcing untuk menjadi sumber data lokasi keberadaan peristiwa berbahaya dan informasi tentang keberadaan korban. Data crowdsourcing diambil dari media sosial Twitter dengan

memanfaatkan API, dan upload dipilih yang hanya menyertakan koordinat lokasi.

Crowdsourcing juga digunakan untuk pengembangan game. Crowdsourcing berupa data lokasi dari Foursquare dan Facebook. Lokasi media sosial diselidiki perannya sebagai kontribusi pada proses komputasi di lingkungan game yang terkait dengan informasi geospasial. Desain game disebut Urbanopoly, namun belum menyajikan sejauh mana pengaruh perilaku manusia di media sosial dengan efeknya pada game [8].

Data tentang bentang lahan dapat diperoleh dari berbagai sumber. Misalnya dari Data Elevation Model (DEM), yang biasa dirilis oleh perusahaan tertentu atau pemerintah. Pemerintah Indonesia telah merilis data DEM yang disebut DEMNAS.

Akuisisi data lainnya adalah penggunaan penginderaan jauh dengan bantuan drone, serta dengan teknologi LIDAR. Beberapa metode pengambilan data tersebut selama ini telah banyak digunakan untuk kajian bentang lahan. Masalah yang sering dihadapi dalam akuisisi data tersebut, adalah waktu yang dibutuhkan cenderung lama serta biaya yang sangat mahal.

Media sosial memiliki potensi yang besar apabila digunakan sebagai data alternatif. Salah satu contoh media sosial yang berpotensi adalah Instagram, dimana ada fitur geo-located yang memberikan informasi lokasi pengambilan gambar. Fitur yang mirip juga ditemukan pada Twitter dan Facebook. Namun belum banyak kajian yang dilakukan dengan sumber data yang berasal dari media sosial.

Penelitian ini bertujuan untuk mengolah data crowdsourcing dari media sosial Facebook, Instagram, dan Twitter terkait fitur geo-located. Begitu juga dengan tulisan pada caption yang diunggah. Kami bertujuan untuk mengungkap sejauh mana data crowdsourcing dari media sosial dapat menyediakan data lanskap, dan apakah data lanskap yang diperoleh dapat digunakan untuk sumber data alternatif untuk studi geoinformatika.

II. PENELITIAN YANG TERKAIT

Peran media sosial dalam dunia pemasaran adalah menjadi wahana kegiatan promosi agar lebih efisien[9]. Penelitian lain dengan sumber data dari media sosial Twitter, dapat memberikan informasi posisi bencana yang terjadi lebih cepat. Dalam upaya kolaborasi data dari Twitter sebagai informasi teks dan data dari Flickr sebagai informasi posisi. Jika nilai bukti adalah 99,9%, maka penelitian ini memberikan akurasi 49% [10]. Media sosial Twitter memiliki sifat keterbatasan dalam menulis caption. Jadi pengguna Twitter cenderung menulis teks pendek, di bawah 160 karakter. Penelitian tentang crowdsourcing Twitter adalah tentang parafrase teks yang diunggah pengguna pada topik yang sama tetapi menggunakan kata-kata yang berbeda. Data yang berasal dari Twitter dilakukan dengan mengunduh API publik. Hasilnya adalah korelasi antara satu caption dengan caption lainnya yang memiliki topik yang sama namun menggunakan kata yang berbeda dengan koefisien korelasi Pearson mencapai 0,735. Meskipun tidak disebutkan prosedur penyortiran yang dilakukan[12].

Beberapa penelitian yang telah dilakukan menghadapi berbagai tantangan. Kendala yang paling dominan adalah dari faktor akurasi titik lokasi. Perbedaan letak titik ini memunculkan informasi baru dan dikaitkan dengan referensi baru[10]. Diperlukan metode komparatif agar akurasi data dapat lebih dipertanggungjawabkan[13]. Crowdsourcing gratis dan siapa pun dapat memberikan data. Variasi dan banyaknya data membuat sulit untuk memilih calon data mana yang terbaik sesuai dengan kebutuhan yang harus dipenuhi[6].

Penelitian sebelumnya terkait media sosial telah membahas crowdsourcing yang digunakan sebagai sumber data untuk menentukan basis lokasi suatu topik tertentu. Data diambil dari lokasi yang dibagikan oleh pengguna smartphone dan diunggah di Twitter & Foursquare. Data yang kemudian digunakan berupa siapa pengguna, lokasi pengguna, waktu upload dilakukan dan teks pada caption. Penelitian ini menghasilkan pendekatan yang lebih murah dari pendekatan yang sudah ada[14].

Hubungan yang diunggah oleh pengguna, baik teks maupun data foto, yang berkaitan dengan kondisi pengguna dan cara berkomunikasi[15]. Alasan seseorang menggunakan Facebook[16], terhadap hubungan antara frekuensi penggunaan Facebook dengan kebiasaan siswa dalam belajar[17].

Crowdsourcing dari Instagram digunakan untuk penelitian dalam perencanaan kota dan deteksi urbanisasi[18].

Penelitian dilakukan dengan data yang diambil dari Instagram geo-located. Data dikumpulkan dalam rentang hari yang berbeda dan analisis menggunakan metode rata-rata DBSCAN dan DBSCAN 1 hari. Kekuatan penelitian ini adalah membandingkan penelitian sebelumnya dengan topik yang sama yaitu media sosial. Meski ditemukan kelemahan, proposal yang diambil memiliki sumber data yang berbeda dari media sosial. Perencanaan pengembangan energi berkelanjutan terkait nilai sosial dan budaya memanfaatkan data dari Instagram[19]. Pada penelitian ini diusulkan suatu metode untuk mengetahui ketahanan alam di sekitar bendungan dari informasi di media sosial penduduk sekitar. Data diambil dari unggahan di Instagram tentang Bendungan SiteC, British Columbia, dan Bendungan Mactaquac, New Brunswick di Kanada sebagai wilayah studi. Hasil pembahasan lanskap terbanyak adalah topik keindahan sebanyak 22%. Karena tidak adanya data pembandingan, akurasi hasil ini tidak diketahui.

Perpaduan analisis geotagging dan unggahan pada caption untuk mengetahui kecenderungan perbincangan publik di USA. Data diambil dari Twitter dengan menggunakan API. Ditemukan bahwa kata kunci penyamanan dan penyamanan adalah yang paling banyak digunakan. Penelitian ini mempresentasikan hasil dengan baik melalui berbagai visualisasi termasuk hubungan cluster antar hashtag. Namun tidak disebutkan berapa atau berapa persen kata kunci yang dominan[20].

Fitur geo-located di media sosial, misalnya Twitter, dapat digunakan untuk mengidentifikasi perilaku pengguna atau karakteristik pengguna. Misalnya, fitur geo-located dapat memberikan informasi waktu favorit mengunjungi tempat hiburan, hingga tempat wisata yang paling banyak dikunjungi di akhir pekan. Data crowdsourcing digunakan dari Twitter dan menghasilkan peta persebaran lokasi wisata di London

serta kecenderungan bahasa yang digunakan. Hasilnya adalah bahasa yang digunakan oleh pengguna 8,9% dalam bahasa Inggris dan bahasa lainnya adalah bahasa asing. Sedangkan unggahan yang menyertakan informasi geo-located sebesar 7,5%. Keunggulan dari penelitian ini adalah memiliki skema pemilahan yang terstruktur sehingga diperoleh sampel terbaik sesuai dengan skenario penelitian. Meskipun tulisan ini menyebutkan geotagging dan geo-located berulang kali, namun hal ini berpotensi menjadi interpretasi yang berbeda [21].

Bersumber dari berbagai penelitian tentang crowdsourcing yang telah dilakukan, kami menemukan adanya kesenjangan yaitu hubungan dengan lanskap. Topik lanskap berbeda dengan penelitian sebelumnya. Sehingga penelitian ini dilakukan untuk mengkaji sejauh mana crowdsourcing dapat menjadi data alternatif untuk lanskap. Kebaruan yang akan diangkat dalam penelitian ini adalah perolehan data dari media sosial sebagai alternatif sumber data. Data alternatif dengan jumlah data yang sangat besar dan sebagai bentuk akuisisi data selain yang sudah ada saat ini.

III. METODE PENELITIAN

A. Data & Proses

Data yang digunakan dalam penelitian ini bersumber dari media sosial Facebook, Instagram, dan Twitter. Ketiga media sosial tersebut memiliki fitur geo-located sehingga dapat diketahui posisi koordinat lokasi pada saat pengguna mengambil gambar. Metadata lokasi tersebut kemudian diterjemahkan menjadi *located* dan terpublikasi pada unggahan pengguna, serta dapat diakses oleh publik.

Dalam penelitian ini, dibutuhkan data berupa gambar/foto. Titik lokasi atau koordinat, serta teks yang merupakan caption dari sebuah unggahan. Sehingga untuk memenuhi ketiga unsur tersebut, sosial media Facebook, Instagram dan Twitter adalah yang paling tepat karena memiliki fitur yang mendukung dan dapat menjawab ketiga unsur yang dibutuhkan dalam penelitian ini.

Pada penelitian ini, kami mengekstrak data secara manual dari media sosial dengan mengunjungi situs resmi Facebook, Instagram, dan Twitter menggunakan browser Google Chrome dan Mozilla Firefox. Dan lokasi titik/koordinat x dan y dalam format desimal diperoleh dari aplikasi google map.

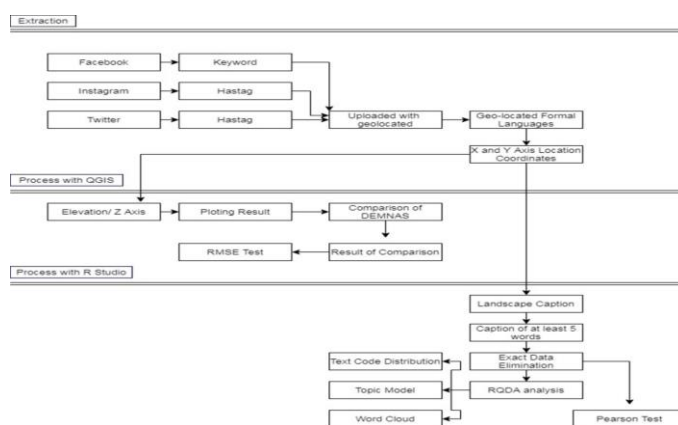
Data yang akan diproses adalah geo-located yang diberikan oleh pengguna, serta keterangan yang menyertainya. Apakah lokasi geo-located dan keterangan yang disediakan mencerminkan lanskap yang sesuai. Data geo-located berupa koordinat dibandingkan dengan Digital Elevation Model dan Batimetri Nasional (DEMNAS). DEMNAS adalah Digital Elevation Model yang resmi dirilis oleh pemerintah Indonesia.

Sehingga dapat dilihat apakah geo-located dari social media crowdsourcing dapat merepresentasikan lanskap tersebut. Penelitian ini memilih media sosial Facebook, Instagram, dan Twitter. Media sosial banyak digunakan oleh masyarakat Indonesia dengan studi kasus Malang Raya. Malang merupakan salah satu daerah di provinsi Jawa Timur, Indonesia yang memiliki bentang alam yang beragam. Ada pemandangan dataran rendah, pantai, dan pegunungan. Serta penggunaan lahan yang beragam, baik untuk kawasan perkotaan, kawasan ekonomi, pendidikan, dan berbagai lokasi wisata.

Metode yang digunakan tampak pada Gambar 1. Dimulai dengan akuisisi data menggunakan kata kunci di Facebook, tagar di Instagram, dan Twitter. Langkah ini dimaksudkan untuk mendapatkan data postingan pengguna yang pernah atau pernah berada di wilayah Malang Raya dari unggahan di akun Instagram mereka. Data yang diperoleh di proses untuk mendapatkan titik-titik geo-located dari masing-masing lokasi yang kemudian diolah dengan QGIS untuk dibandingkan dengan DEMNAS.

Data DEMNAS memungkinkan terbentuknya data elevasi. Data DEMNAS digunakan sebagai pembanding karena data ini dikeluarkan oleh pemerintah Republik Indonesia, dan secara legal standing tingkat validitas tertinggi sesuai dengan amanat UU No.4 Tahun 2011 tentang BIG (Badan Informasi Geospasial) Badan Geospasial Indonesia[22].

DEMNAS dapat diunduh secara gratis dengan mengunjungi website <http://tides.big.go.id/DEMNAS/>, dan mendaftar sebagai anggota terlebih dahulu. Data DEMNAS berupa gambar kemudian diolah menggunakan software QGIS untuk menentukan ketinggian suatu tempat dengan hasil dalam satuan m (meter). Dalam pembahasan ini, kami menyajikan pengenalan beberapa jenis data yang digunakan dalam penelitian ini. Semua data digunakan baik dalam tahap akuisisi data, pengolahan data hingga output yang dihasilkan.



Gbr. 1 Bagan Alir Metodologi Yang Diusulkan

Data Lokasi dari Media Sosial

Ada tiga media sosial untuk ekstraksi data crowdsourcing dalam penelitian ini, yaitu Facebook, Instagram, dan Twitter. Dari masing-masing media sosial tersebut, terdapat perbedaan pada masing-masing karakter. Lokasi di Facebook, disematkan di sisi kanan nama pengguna. Lokasi Instagram ada di menu pilihan saat proses upload foto dan pengguna bisa memilih atau menambahkan lokasi sesuai keinginan. Sedangkan untuk Twitter lokasinya dipersempit karena tidak semua lokasi bisa dipilih seiring dengan regulasi Twitter untuk merahasiakan lokasi dan membaca metadata dari foto yang diunggah pengguna.

Titik Koordinat Lokasi/Geografis

Geo-located yang dimaksud adalah sebuah fitur di media sosial yang di dalamnya terdapat informasi geografis dengan koordinat yang disematkan di setiap pesan[23]. Fitur ini memungkinkan pengguna media sosial untuk berbagi posisi atau lokasi di mana mereka berada berdasarkan data yang diberikan oleh pihak ketiga. Pada penelitian ini, data

koordinat dari media sosial berupa titik koordinat x dan y dengan format desimal. Data titik dari setiap lokasi kemudian diplot pada peta dengan aplikasi QGIS 2.18 sehingga dapat ditampilkan lokasi setiap titik dari data koordinat tersebut. X, Y mengoordinasikan data dari setiap unggahan di Facebook, Instagram, dan Twitter. Dikelompokkan dalam file berformat CSV dengan bantuan Microsoft Excel. File CSV dari Microsoft Excel dipanggil ke QGIS dengan fitur layer - Comma Delimited, sehingga akan muncul visualisasi koordinat lokasi unggahan di media sosial.

Geo-located merupakan instrumen penting yang menyediakan data berupa koordinat suatu objek. Koordinat objek dapat dibaca di peta yang kemudian dibandingkan dengan DEMNAS, terutama pada elevasi atau letak ketinggian objek. Fitur geolocating pada media sosial Facebook, Instagram, dan Twitter, memberikan informasi lokasi berdasarkan nama. Nama lokasi yang dimaksud dapat diolah menggunakan Google Maps sebagai aplikasi open source untuk mendapatkan koordinat X dan Y. Sedangkan untuk data elevasi dari suatu lokasi digunakan data elevasi dari Google Earth dalam satuan meter. Google Earth sebagai media open-source merupakan pengembangan dari Google map yang dapat memberikan informasi lokasi beserta ketinggian. Aplikasi ini dapat diakses secara gratis dengan browser Google Chrome. Ketinggian di google earth dinyatakan dalam meter. Data ketinggian pada google earth berfungsi sebagai nilai prediksi ketinggian suatu tempat.

koordinat dari media sosial berupa titik koordinat x dan y dengan format desimal. Data titik dari setiap lokasi kemudian diplot pada peta dengan aplikasi QGIS 2.18 sehingga dapat ditampilkan lokasi setiap titik dari data koordinat tersebut. X, Y mengoordinasikan data dari setiap unggahan di Facebook, Instagram, dan Twitter. Dikelompokkan dalam file berformat CSV dengan bantuan Microsoft Excel. File CSV dari Microsoft Excel dipanggil ke QGIS dengan fitur layer - Comma Delimited, sehingga akan muncul visualisasi koordinat lokasi unggahan di media sosial.

Geo-located merupakan instrumen penting yang menyediakan data berupa koordinat suatu objek. Koordinat objek dapat dibaca di peta yang kemudian dibandingkan dengan DEMNAS, terutama pada elevasi atau letak ketinggian objek. Fitur geolocating pada media sosial Facebook, Instagram, dan Twitter, memberikan informasi lokasi berdasarkan nama. Nama lokasi yang dimaksud dapat diolah menggunakan Google Maps sebagai aplikasi open source untuk mendapatkan koordinat X dan Y. Sedangkan untuk data elevasi dari suatu lokasi digunakan data elevasi dari Google Earth dalam satuan meter. Google Earth sebagai media open-source merupakan pengembangan dari Google map yang dapat memberikan informasi lokasi beserta ketinggian. Aplikasi ini dapat diakses secara gratis dengan browser Google Chrome. Ketinggian di google earth dinyatakan dalam meter. Data ketinggian pada google earth berfungsi sebagai nilai prediksi ketinggian suatu tempat.

Data Tekstual/ Semantik

Semantik adalah struktur gramatikal. Itu adalah struktur yang berkorelasi dengan cara di mana setiap kata individu muncul dalam dokumen. Semantik hanya menyiratkan fakta dalam dokumen yang dapat berupa istilah dan dapat diambil

sebagai referensi untuk dokumen itu sendiri atau untuk topik[25].

Semantik dalam penelitian ini diambil dari caption. Caption adalah tulisan yang ditulis oleh pengguna saat mengunggah foto yang ditampilkan pada Tabel 1. Caption dapat berupa penjelasan atas foto yang diunggah, atau sebagai informasi pelengkap. Namun di media sosial, caption bisa berupa hal-hal lain yang tidak berkaitan dengan informasi yang ada di dalam foto.

DEM

Digital Elevation Model (DEM) adalah data digital yang merepresentasikan bentuk permukaan bumi. Salah satu data DEM yang dapat diunduh secara gratis adalah JAXA'S ALOS 3D World yang memiliki akurasi vertikal 5 meter[26]. Pada penelitian ini, data DEM dirilis secara resmi oleh pemerintah Indonesia melalui Badan Informasi Geospasial yang disebut dengan DEMNAS.

DEMNAS adalah Model Elevasi Digital yang diakui secara resmi oleh pemerintah Indonesia dan berlaku secara nasional. DEM nasional dibangun dari beberapa sumber data termasuk data IFSAR (resolusi 5m), TERRASAR-X (resolusi 5m) dan ALOS PALSAR (resolusi 11,25m), dengan menambahkan data Masspoint ke hasil stereo-plotting. Resolusi spasial DEMNAS adalah 0,27 detik busur, menggunakan datum vertikal EGM2008.

Proses Penyaringan Data

Alat yang digunakan dalam proses ini adalah Microsoft Excel, Google Chrome, atau Mozilla Firefox tray yang diakses melalui PC atau laptop. Proses pemilihan data dari Facebook, Instagram, dan Twitter adalah:

1. Mencari data dengan kemungkinan kata kunci pada caption, pada akuisisi data Instagram dan Twitter menggunakan hashtag,

2. Hanya data yang diunggah dengan fitur geo-located yang akan digunakan. Lokasi geo-lokasi yang menggunakan bahasa formal dan koordinat dapat ditemukan,

3. Hanya data yang diunggah dengan minimal lima kata yang akan digunakan,

4. Penghapusan data yang memiliki kemiripan/identik. Pendekatan URL digunakan untuk menyortir data yang identik. Tahapan untuk proses sortasi adalah:

- a. sebuah. Menggunakan Microsoft Excel. Semua data dikumpulkan dalam 1 lembar kerja, dengan penyesuaian kolom yang sesuai dengan isi,

- b. Pada kolom URL, penyortiran dilakukan dengan fitur Conditional Formatting - Highlight Cell Rule - Duplicate Value, sehingga URL yang dobel/identik akan ditandai dengan warna tertentu. Dipilih satu jika ada URL ganda.

Di media sosial, Facebook memilih kata kunci sebagai media pencarian. Pasalnya, peran hashtag di Facebook tidak sedominan di Instagram dan Twitter. Sangat jarang pengguna Facebook yang menggunakan hashtag. Pada fitur pencarian di Facebook, penggunaan lebih banyak kata kunci dapat diupayakan lebih lanjut di lokasi Malang, dan Batu sesuai dengan wilayah studi kasus. Namun, fitur berbentuk kerucut ini tidak ditemukan saat menggunakan hashtag.

Proses penyaringan di Twitter dan Instagram menggunakan hashtag. Keduanya memiliki kesamaan fitur

hashtag yang dominan digunakan oleh pengguna. Trending topik di Twitter bahkan berdasarkan seberapa banyak penggunaan hashtag. Di Instagram dan Twitter, ada banyak dan beragam tagar. Menyertakan hastag berisi kata Malang dan Batu menurut daerah untuk studi kasus.

Visualisasi Quantitatif

Analisis berbasis LDA (laten Dirichlet alokasi) dapat memberikan hasil yang berkonsentrasi pada parameter yang telah ditentukan [28], dan dapat memberikan data spesifik pada dokumen yang berisi banyak tag [29]. Identifikasi kata-kata yang digunakan pengguna media sosial pada caption dianalisis dengan pendekatan LDA. Sehingga dapat divisualisasikan kata-kata yang paling populer atau paling banyak digunakan dari semua data. Kata-kata yang muncul dengan frekuensi tinggi akan muncul di WordCloud.

Wordcloud adalah metode visualisasi teks yang dapat langsung muncul sesuai dengan frekuensi teks yang tinggi. Metode tipikal ini bekerja secara statis dan dapat menampilkan ringkasan teks secara murni[30]. Frekuensi kata yang dimunculkan dalam visualisasi word cloud tergantung pada skala penelitian. Pada penelitian ini, kata yang muncul pada kata cloud visualization adalah kata pada caption dengan frekuensi minimal 10 kali. Pengaturan frekuensi ditulis dalam pengkodean di R Studio.

Coding R Studio juga mampu menjalankan analisis kuantitatif terhadap teks pada caption yaitu pemodelan topik. Topic Modelling dihasilkan dari analisis kategori teks yang paling sering muncul. Posisi teratas adalah teks dengan frekuensi kemunculan tertinggi. Metode ini dijalankan dengan RQDA (paket R untuk Analisis Data Kualitatif). Visualisasi pemodelan topik didasarkan pada kategori-kategori yang dibentuk (berupa kode). Studi ini terbentuk dalam 18 kategori. Proses pembentukan kategori tersebut adalah:

1. Pelajari teks dalam keterangan secara manual. Disimpulkan bahwa caption berbicara tentang apa. Setiap caption yang memiliki kecenderungan yang sama atau mendekati dikelompokkan dalam kategori yang sama. Memberi nama kategori tersebut sekaligus mempresentasikan apa yang dibicarakan, misalnya penampilan, fotografi, lanskap,
2. Jika kategori yang ada masih belum mengakomodir ide dari sebuah caption, maka dibentuklah kategori baru untuk mengakomodir apa tema dari caption tersebut,
3. Proses pada poin 2 berulang hingga semua caption dapat masuk dalam kategori. Dan dibentuk menjadi 18 kategori.

IV. HASIL DAN PEMBAHASAN

Data Sorting

Hampir setiap tahapan terjadi penyusutan sesuai ketentuan yang diinginkan. Terdapat perbedaan jumlah data, sesuai dengan karakteristik masing-masing media sosial (Tabel I). Instagram memiliki jumlah sampel terbanyak karena di media sosial sangat mudah mencari data sesuai ketentuan.

TABEL I
JUMLAH DATA DAN DEPRESIASI PADA SETIAP TAHAP SORTASI

No	Stage	Facebook	Instagram	Twitter
1	Data dengan kemungkinan kata kunci pada caption, pada akuisisi data Instagram dan Twitter menggunakan hashtag,	223	13,542	861
2	Hanya data yang diunggah dengan fitur geo-located yang akan digunakan,	216	3,586	264
3	Lokasi geo-lokasi yang menggunakan bahasa formal dan koordinat dapat ditemukan,	204	3,542	254
4	Hanya data yang diunggah dengan setidaknya lima kata,	185	3,175	230
5	Penghapusan data dengan kesamaan identik. Pendekatan URL digunakan untuk menyortir data yang identik	153	3,153	230

Dalam proses pemilihan data ini, terdapat beberapa permasalahan, terutama dalam hal ketidakpastian. Ketidakpastian dapat didefinisikan sebagai ukuran pemahaman pengguna tentang perbedaan antara isi dataset dan fenomena yang dapat diamati yang diyakini mewakili data [31]. Ketidakpastian yang dapat muncul dalam penelitian ini adalah faktor ambigu, di mana pengguna mengunggah foto lokasi A tetapi berikan lokasi geografis di lokasi B.

Postingan yang mengikuti kriteria penyortiran, caption yang menyertainya bisa diproses. Hanya teks terpilih yang berada di bawah skenario penyortiran. Semua keterangan yang sesuai diformat dalam file teks dengan bantuan notepad. File teks yang telah disiapkan kemudian diproses dengan metode analisis RQDA yang dijalankan dengan R Studio. Sehingga dapat diketahui topik atau kata-kata apa yang paling banyak diunggah oleh pengguna media sosial.

Topic Modelling

Tantangan yang didapat selama penelitian ini adalah ragam informasi yang dituliskan dalam caption oleh pengguna media sosial. Pada tahap penyortiran, banyak karakter yang harus dihilangkan, seperti karakter spasi berlebihan, tanda tangan, dan enter. Pengguna media sosial sering menggunakan spasi dan enter untuk menambah area kosong pada caption mereka.

Penulisan lokasi yang tidak baku juga menjadi tantangan dalam penelitian ini. Bahkan di satu titik lokasi, mereka memiliki lebih dari dua nama di media sosial yang sama. Berbagai lokasi penamaan dapat dipilih secara bebas oleh pengguna. Padahal penulisan lokasi harus unik agar setiap lokasi memiliki informasinya[31]. Kepadatan poin pengguna media sosial baik Facebook, Instagram, dan Twitter berbeda-beda pada setiap kategori (Tabel II).

Caption yang digunakan oleh pengguna menjadi salah satu agenda dalam penelitian ini. Maka untuk menganalisis berbagai kemungkinan topik yang dibahas, dari pendekatan kata demi kata digunakan pendekatan WordCloud. Proses WordCloud dijalankan dengan alat R Studio dengan paket RQDA. Alat-alat ini biasanya digunakan dalam analisis data kuantitatif[32]. Sehingga didapatkan visualisasi WordCloud seperti pada Tabel II.

TABEL II
KATA TERPOPULER DARI MASING-MASING MEDIA SOSIAL

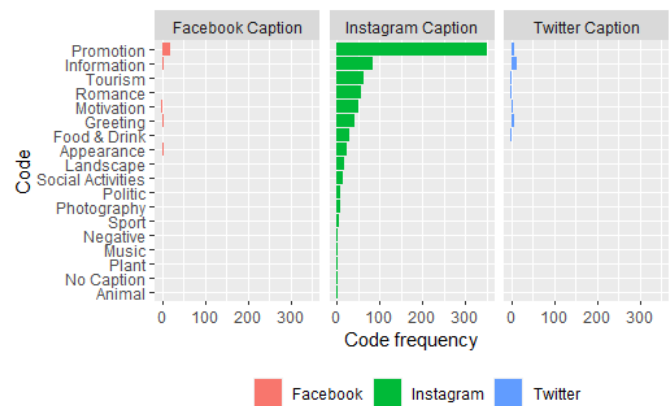
(a) Kata populer di Facebook	(b) Kata populer di Instagram	(c) Kata populer di Twitter

Dari aktivitas pengguna di media sosial Facebook, kata "batu" paling banyak muncul kemudian "Kota". Sedangkan di Instagram, trend yang muncul adalah kata "Malang" dan diikuti dengan "promosi". Adapun hasil di Twitter yang menjadi populer adalah "informasi" dan "malang". Konten promosi ditemukan di Facebook, Instagram, dan Twitter. Jumlah konten menjadi dominan dengan persentase 24,41% dari seluruh data yang masuk sebanyak 3.536 sampel. Sedangkan kategori lanskap sebesar 5,37%. Proporsi masing-masing kategori dalam persentase ditunjukkan pada Tabel III.

Bentang alam yang dimaksud dalam penelitian ini adalah caption yang berisi kata-kata tentang alam. Misalnya suhu berbicara tentang panas, dingin, tentang bentang alam tinggi, rendah, lembah, bukit, puncak gunung, tentang letak gunung, hutan, sumber air panas, pantai, savana. Hasil frekuensi bentang alam di facebook, instagram dan Twitter menempati posisi kesembilan dalam 18 kategori. Seperti yang ditunjukkan pada Gambar 2.

Posisi pengguna media sosial akan lebih mudah dipahami dengan adanya data informasi geo-located. Data geo-located yang kami peroleh diilustrasikan dalam peta berikut dengan kategori yang telah ditentukan. Penentuan titik masuk ke dalam kategori mana sesuai dengan tulisan pada caption. Kategori dibuat untuk membentuk kelompok atas

kecenderungan pengguna memberikan informasi di media sosial. Kategori tersebut ditunjukkan pada Tabel IV.



Gbr. 2 Code Distribution Dari Setiap Caption Sosial Media

TABEL III
JUMLAH UNGGAHAN UNTUK SETIAP KATEGORI

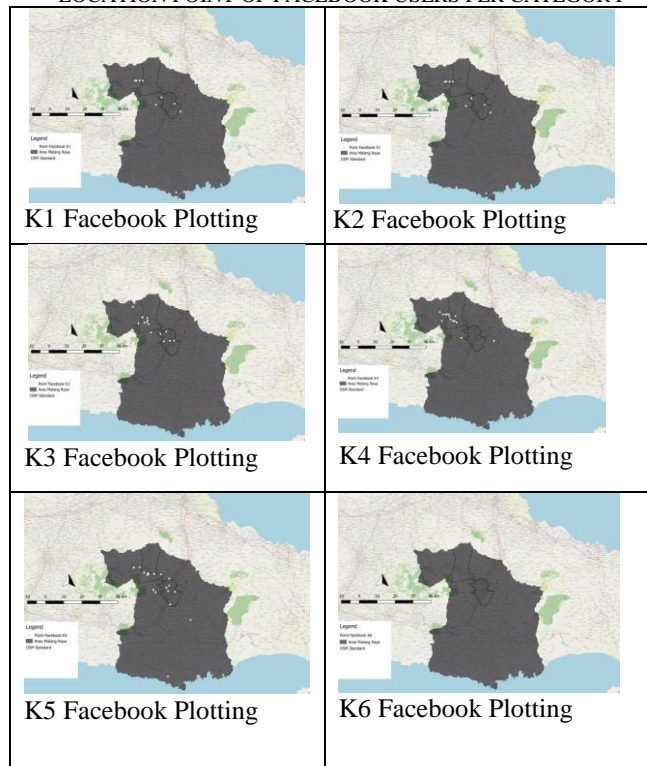
No	Code	Facebook	Instagram	Twitter	% upload
1	Social Activities	5	67	9	2.29
2	Landscape	13	173	4	5.37
3	Food & Drink	0	128	7	3.82
4	Photography	3	40	2	1.27
5	Animal	0	10	0	0.28
6	Information	21	253	87	10.21
7	Motivation	9	459	25	13.94
8	Music	0	24	0	0.68
9	Negative	0	18	1	0.54
10	No Caption	0	3	3	0.17
11	Sport	1	32	1	0.96
12	Appearance	16	160	4	5.09
13	Romance	4	404	10	11.82
14	Politic	0	23	2	0.71
15	Promotion	41	790	32	24.41
16	Tourism	9	241	8	7.30
17	Plant	0	16	1	0.48
18	Greeting	31	312	34	10.66
Total Unggahan		3536			100

Dari 18 kategori yang di unggah pada sebuah caption. Ada kesamaan dalam kelompok tertentu. Sehingga 18 kategori dipersempit menjadi 6 kelompok dengan kemiripan yang khas. Banyak kesamaan yang ditemui adalah penerimaan makna ganda dalam caption. Misalnya informasi yang disampaikan pada caption mengandung unsur promosi dengan menggunakan metode soft selling suatu produk atau jasa.

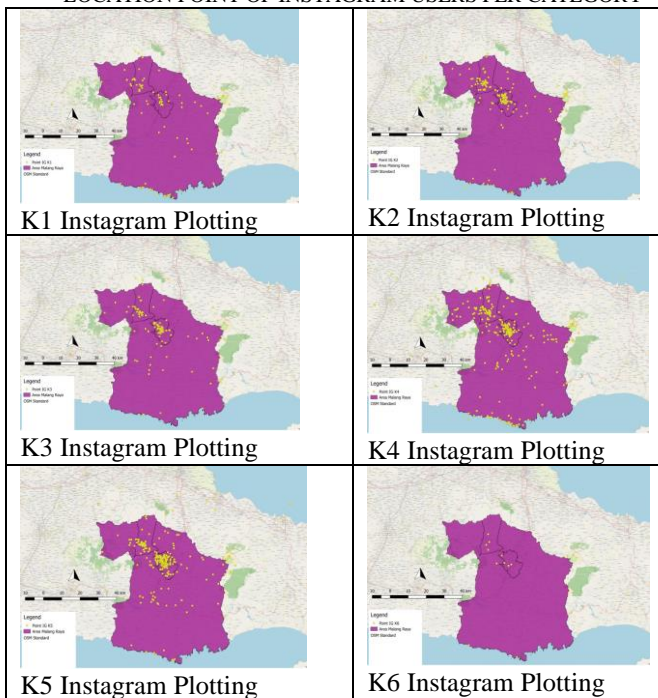
Pengelompokan kategori didasarkan pada kesamaannya. Kategori 1 (K1) Hewan, Tumbuhan, Lanskap. Kategori 2 (K2) Pariwisata, Fotografi, Makanan & Minuman. Kategori 3 (K3) Kegiatan Sosial, Penampilan, Musik, Olahraga. Kategori 4 (K4) Romansa, Motivasi, Salam. Kategori 5 (K5) Promosi, Informasi, Politik. Kategori 6 (K6) Tanpa Keterangan/Slank, Negatif.

Penyajian distribusi koordinat dari 6 kategori tersebut, akan memberikan informasi yang lebih mudah dipahami dimana suatu trend terjadi. Misalnya di mana sebagian besar kunjungan turis terjadi atau di mana sebagian besar pengguna media sosial berbicara tentang politik.

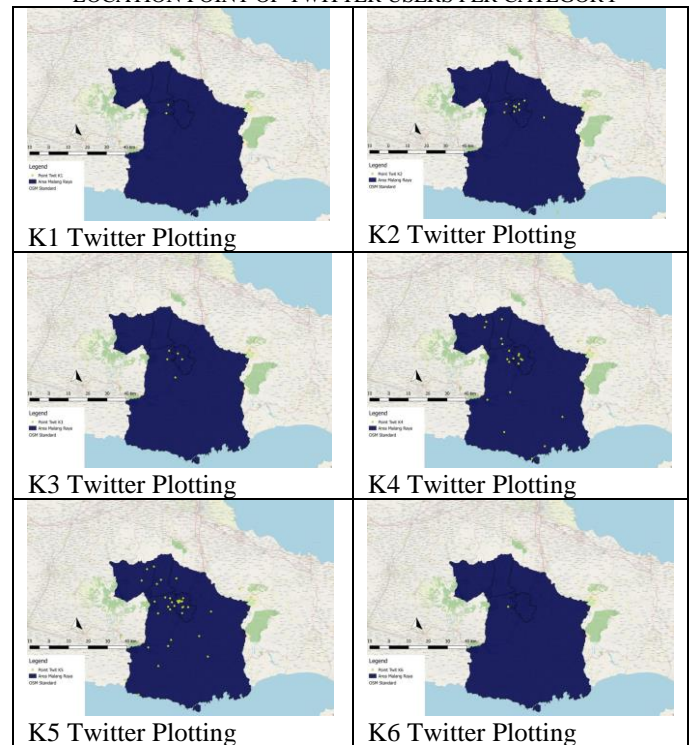
TABEL IV
LOCATION POINT OF FACEBOOK USERS PER CATEGORY



TABEL V
LOCATION POINT OF INSTAGRAM USERS PER CATEGORY



TABEL VI
LOCATION POINT OF TWITTER USERS PER CATEGORY



Pengguna media sosial Facebook membicarakan kategori 1 (Tabel IX) yang tersebar dari Malang Selatan ke utara hingga ke barat. Di Malang selatan, banyak lokasi wisata yang berhubungan erat dengan alam, seperti pantai, perkebunan kopi, dan perkebunan cengkeh. Begitu juga sebaran di Malang bagian barat, merupakan tempat dengan kontur banyak tempat wisata air terjun hingga air panas. Terdapat kesamaan sebaran Kategori 3,4 dan 5. Berada di wilayah Kota Malang. Kemudian di perbatasan Kota Malang dan Kota Batu mengarah ke utara. Di daerah-daerah tersebut terdapat berbagai lokasi. Mulai lokasi wisata buatan hingga lokasi wisata alam dan hutan serta lokasi pendakian.

Pada sebaran data Instagram (Tabel X), pada Kategori 2,3 dan 5 terdapat perbedaan yaitu banyaknya titik yang ada di luar wilayah Malang Raya (Kota Malang, Kabupaten Malang, Kota Batu). Pengguna media sosial mengunggah caption tentang pariwisata atau informasi lain tentang Malang namun memberikan input lokasi di luar Malang.

Sebaran kategori 1 s/d 6. Konsentrasi pengguna media sosial Instagram di Kota Malang menyebar ke Kota Batu. Adapun bagian selatan Malang berhimpitan dengan pesisir dan pulau-pulau kecil di sekitarnya. Kategori 5 yang meliputi Promosi, Informasi, dan Politik terkonsentrasi di Kota Malang dengan persebaran yang sangat padat. Sedangkan yang merata hampir di seluruh wilayah adalah Kategori 4.

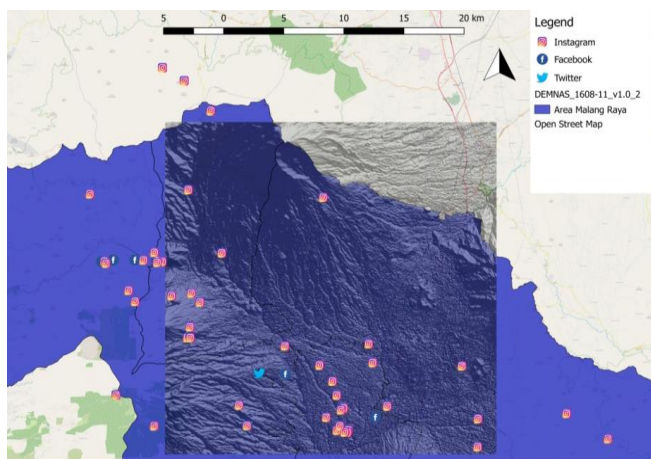
Data koordinat dari media sosial Twitter (Tabel XI), dikelompokkan berdasarkan kategori. Sebaran pengguna Twitter terbanyak terdapat pada kategori 5. Dimana persebarannya hampir merata di seluruh wilayah Malang dan paling padat di Kota Malang. Kepadatan lainnya di wilayah Kota Batu dan sebagian Malang Selatan di sekitar pantai.

Terdapat kesamaan sebaran Kategori 1,2 dan 3 yang semuanya berpusat di Kota Malang dan sedikit menyebar ke arah Kota Batu. Tempat-tempat dengan poin dari data media sosial tidak selalu merupakan bentuk representasi warga di lokasi tersebut. Data geo-located tersebut berasal dari crowdsourcing media sosial, sehingga lokasi titik-titik pada peta merupakan representasi dari pengguna media sosial yang tidak hanya penduduk tetapi juga pendatang atau wisatawan.

Informasi Lanskap/ Bentang Lahan

Data lanskap referensi yang digunakan adalah DEMNAS. Data DEMNAS di-overlay pada QGIS dengan poin-poin dari media sosial dari kategori lanskap. Dengan data elevasi dari DEMNAS, apakah caption atau informasi di media sosial dapat memberikan informasi tentang lanskap tersebut.

Informasi dari media sosial berupa titik-titik yang dikunjungi pengguna. Untuk validasi data tentang lanskap digunakan pendekatan metode RMSE. Perlakuan RMSE berada pada titik-titik yang masuk dalam area data DEMNAS. Ditunjukkan pada Gambar 5, wilayah Malang Raya ditunjukkan dengan warna biru, dan titik kunjungan pengguna media sosial terdapat pada logo media sosialnya. Data DEMNAS yang memiliki data elevasi di-overlay dengan data media sosial.



Gbr. 2 Geo-located Poin dari Data Bentang Lahan

Hasil Validasi

Root Mean Square Error (RMSE) banyak digunakan, termasuk dalam studi iklim dan lingkungan karena kemampuannya untuk mengukur kesalahan rata-rata dari data statistik [33]. Tidak hanya dengan 1 set error tetapi RMSE dapat mendeteksi error pada 3 karakteristik[33]. Tingkat kesalahan dalam RMSE direpresentasikan dalam skala 0,0 - 1,0 dan tingkat kesalahan lebih baik jika mendekati 0,0. Crowdsourcing data dari Facebook, Instagram dan Twitter. Diberi perlakuan RMSE menggunakan rumus (1).

$$RMSE = \left(\frac{\sum(y-y')^2}{n} \right)^{\frac{1}{2}} \tag{1}$$

Di mana :

RMSE = akar rata-rata nilai kesalahan kuadrat

y = nilai hasil penelitian

y ' = nilai prediktif dan n = jumlah data

Nilai hasil penelitian (y) adalah semua data dari media sosial tentang tinggi badan dalam meter. Dengan nilai prediksi (y') adalah data ketinggian tanah dari data DEMNAS dalam meter dan n adalah jumlah data keseluruhan. Untuk mendapatkan hasil RMSE seperti:

TABEL VII
HASIL VALIDASI RMSE

No	Social Media	RMSE Result
1	Facebook	654.4074
2	Instagram	238.1565
3	Twitter	420.9199

Tahap validasi data dari media sosial menggunakan pendekatan human account atau non human account. Setiap akun memiliki karakteristik. Akun manusia atau akun real manusia memposting dengan tagar yang berbeda, sedangkan akun non-manusia akan memposting dengan tagar yang sama dan berulang-ulang [34].

Ciri-ciri akun manusia yaitu: Profil atau nama tidak mengandung karakter khusus, misalnya *%#. Gunakan foto profil yang jelas, pernah menggunakan fitur story, memiliki riwayat postingan lebih dari sepuluh kali, memiliki penekanan atau menambahkan emosi pada postingan dan memposting kurang dari lima kali sehari. Akun yang tidak memenuhi karakter diatas disebut akun non manusia. Termasuk akun yang postinggannya dihapus atau akun yang tidak ditemukan saat dikunjungi untuk kedua kalinya. Pengujian validitas menggunakan pendekatan menarik [35].

Bandingkan jumlah akun manusia nyata dengan jumlah akun non-manusia. Dengan rumus (2) :

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{(n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2)(n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2)}} \tag{2}$$

Dimana :

r_{xy} = Koefisien korelasi antara variabel X and Y

x_i = nilai dari data to = i for variabel grup X yaitu akun manusia

y_i = nilai dari data to = i for variabel grup X yaitu akun bukan manusia

n = jumlah total data

Pengujian validitas membandingkan variabel X yaitu kelompok akun manusia nyata dan variabel Y berisi kelompok akun bukan manusia. Hasil pengujian validitas ditampilkan dalam Tabel VIII.

TABEL VIII
PERBANDINGAN TES VALIDASI

Variabel	Akun Manusia	Akun Bukan Manusia
Facebook	12	1
Instagram	157	16
Twitter	4	0
Number of n	173	17
Coefficient Pearson correlation	0.999956	
Coefficient instrumen validity	0.999956 (r square)	
r table Pearson	0.1191 (valid)	

Instrumen perhitungan valid, dimana r hitung > = r tabel

V. KESIMPULAN

Dari ketiga media sosial yaitu Facebook, Instagram dan Twitter terdapat satu kesamaan. Yaitu promosi kategori upload yang sama tingginya mencapai 24,41% dari total sampel. Sedangkan untuk kepentingan informasi landscape hanya mencapai 5,37% dan porsi terkecil adalah “No Caption”, artinya apa yang tertulis pada kolom caption dapat diinterpretasikan dan tidak dapat dimasukkan ke dalam kategori manapun.

Dari hasil analisis LDA, kata “promosi” muncul di ketiga media sosial tersebut. Dengan kepadatan tertinggi ditemukan di Instagram. Banyak unggahan di Instagram tentang promosi ditemukan sebanyak 790 unggahan. Lanskapnya masih jarang informasi ke pengguna media sosial. Hasil penelitian ini hanya menempatkan lanskap pada posisi yang sangat kecil dengan persentase 5,37. Jika kategori lanskap dipersempit untuk kebutuhan data elevasi atau tinggi tapak, hasil kesalahannya sangat tinggi.

Hasil analisis rata-rata error dengan RMSE untuk media sosial Facebook 654.4074, Instagram 238.1565 dan Twitter 4209199, sangat jauh dari 0.0 - 1.0. Dengan rentang error yang sangat tinggi, disimpulkan bahwa data dari media sosial Facebook, Instagram dan Twitter masih jauh dari yang diharapkan untuk menjadi sumber data alternatif kajian geoinformatika. Peluang penelitian lebih lanjut dapat dilakukan pada pengembangan perilaku pengguna media sosial, tentang data yang diunggah dan memberikan hasil yang lebih baik rata-rata kesalahan.

UCAPAN TERIMA KASIH

Ucapan terimakasih terhatur untuk segala pihak yang membantu & berdoa untuk kesuksesan penelitian ini. Segala dukungan dari keluar, kampus dan berbagai pihak lainnya.

DAFTAR PUSTAKA

- [1] KBBI, “KBBI,” *Kemendikbud*, 2016. [Online]. Available: [https://kbbi.kemdikbud.go.id/entri/media sosial](https://kbbi.kemdikbud.go.id/entri/media%20sosial).
- [2] D. Thom, H. Bosch, S. Koch, M. Worner, and T. Ertl, “Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages,” in *IEEE Pacific Visualization Symposium 2012, PacificVis 2012 - Proceedings*, 2012.
- [3] G. Jossé *et al.*, “Knowledge extraction from crowdsourced data for the enrichment of road networks,” *Geoinformatica*, vol. 21, no. 4, pp. 763–795, 2017.
- [4] S. Asur and B. A. Huberman, “Predicting the future with social media,” in *Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010*, 2010.
- [5] B. Liu, “Sentiment analysis and opinion mining,” *Synth. Lect. Hum. Lang. Technol.*, 2012.
- [6] A. Afuah and C. L. Tucci, “Crowdsourcing as a solution to distant search,” *Academy of Management Review*, 2012.
- [7] J. Rogstadius and V. Kostakos, “Towards Real-time Emergency Response using Crowd Supported Analysis of Social Media,” in *Computational Linguistics*, 2011.
- [8] I. Celino *et al.*, “Urbanopoly - A social and location-based game with a purpose to crowdsource your urban data,” in *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, 2012.
- [9] A. Hassan Zadeh and R. Sharda, “Modeling brand post popularity dynamics in online social networks,” *Decis. Support Syst.*, 2014.
- [10] G. Panteras, S. Wise, X. Lu, A. Croitoru, A. Crooks, and A. Stefanidis, “Triangulating Social Multimedia Content for Event Localization using Flickr and Twitter,” *Trans. GIS*, 2015.
- [11] MASSA P and CAMPAGNA M, “Social Media Geographic Information: Current developments and opportunities in urban and regional planning,” in *Poster-pitch presentation in track 07 Analytical Tools and Support Systems for Planning*, 2014.
- [12] W. Xu, A. Ritter, C. Callison-Burch, W. B. Dolan, and Y. Ji, “Extracting Lexically Divergent Paraphrases from Twitter,” *Trans. Assoc. Comput. Linguist.*, 2014.
- [13] J. Capdevila, M. Arias, and A. Arratia, “GeoSRS: A hybrid social recommender system for geolocated data,” *Inf. Syst.*, 2016.
- [14] P. Shankar, Y. W. Huang, P. Castro, B. Nath, and L. Iftode, “Crowds replace experts: Building better location-based services using mobile social network interactions,” in *2012 IEEE International Conference on Pervasive Computing and Communications, PerCom 2012*, 2012.
- [15] C. Ross, E. S. Orr, M. Sisic, J. M. Arseneault, M. G. Simmering, and R. R. Orr, “Personality and motivations associated with Facebook use,” *Comput. Human Behav.*, 2009.
- [16] R. E. Wilson, S. D. Gosling, and L. T. Graham, “A Review of Facebook Research in the Social Sciences,” *Perspect. Psychol. Sci.*, 2012.
- [17] R. Junco, “The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement,” *Comput. Educ.*, 2012.
- [18] D. R. Domínguez, R. P. Díaz Redondo, A. F. Vilas, and M. Ben Khalifa, “Sensing the city with Instagram: Clustering geolocated data for outlier detection,” *Expert Syst. Appl.*, 2017.
- [19] Y. Chen, J. R. Parkins, and K. Sherren, “Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs,” *Landsc. Urban Plan.*, 2018.
- [20] J. R. Radzikowski, H. Hollen, and S. Fuhrmann, “Using Twitter content to crowdsource opinions on tanning in the United States,” *Kartographische Nachrichten*, 2015.
- [21] G. Lansley and P. A. Longley, “The geography of Twitter topics in London,” *Comput. Environ. Urban Syst.*, vol. 58, pp. 85–96, 2016.
- [22] T. A. D. P. D. A. Sulistiana, “Analisis akurasi vertikal digital elevation model nasional (DEMNAS) studi kasus Kota Medan,” *Indones. Surv. Indones.*, no. January, p. 75, 2019.
- [23] B. Hawelka, I. Sitko, E. Beinart, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, “Geo-located Twitter as proxy for global mobility patterns,” *Cartogr. Geogr. Inf. Sci.*, 2014.
- [24] ESRI, “FAQ: What is the difference between a shapefile and a layer file?,” 2020. [Online]. Available: <https://support.esri.com/en/technical-article/000011516>.
- [25] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *J. Am. Soc. Inf. Sci.*, 1990.
- [26] F. Ramdani, *Pengantar Ilmu Geoinformatika*, 1st ed. Malang: UB Press, 2017.
- [27] BIG, “No Title,” *Tides*, 2018. [Online]. Available: <http://tides.big.go.id/DEMNAS/>. [Accessed: 14-Nov-2019].
- [28] H. M. Wallach, D. Mimno, and A. McCallum, “Rethinking LDA: Why priors matter,” in *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, 2009.

- [29] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora," in *EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009*, 2009.
- [30] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, "Word cloud explorer: Text analytics based on word clouds," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2014.
- [31] P. A. Longley, M. F. G. Good, D. M. J, and D. W. Rhind, *Geographic Information Science and System*, 4th ed. 2015.
- [32] Y. Chandra and L. Shang, "An RQDA-based constructivist methodology for qualitative research," *Qual. Mark. Res.*, 2017.
- [33] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, 2005.
- [34] A. J. Panatra, F. B. Chandra, W. Darmawan, H. L. H. S. Warnars, W. H. Utomo, and T. Matsuo, "Buzzer Detection to Maintain Information Neutrality in 2019 Indonesia Presidential Election," in *Proceedings - 2019 8th International Congress on Advanced Applied Informatics, IIAI-AAI 2019*, 2019.
- [35] B. A. Priatna, "Uji Instrumen Penelitian Dengan MS. Excel dan SPSS," *Penelitian*, no. November, pp. 1–22, 2008.