

Pemanfaatan Ilmu Statistik dalam Teknologi Sistem Informasi Geografis

Latifa Dwiyanti (18209010)

Program Studi Sistem dan Teknologi Informasi

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

18209010@std.stei.itb.ac.id

Abstract—Statistika sebagai salah satu ilmu yang dengan kemampuan untuk menghitung ketidakpastian dengan tepat, adalah salah satu ilmu yang sulit dipisahkan dalam perkembangan teknologi secara umum. Pengembangan dalam bidang teknologi informasi pun tidak terlepas dari peran besar ilmu statistika. Pada makalah ini akan dijabarkan mengenai salah satu pemanfaatan ilmu statistika dalam pengembangan teknologi sistem informasi geografis. Dijelaskan penggunaan ilmu statistika dalam pengolahan data dan analisis citra. Sistem informasi geografis adalah suatu sistem informasi yang digunakan untuk memasukkan data-data, yang berhubungan dengan posisi-posisi permukaan bumi, yang kemudian akan diintegrasikan, diperiksa, dimanipulasi, dianalisis sehingga dapat ditampilkan dan dipergunakan sebagai suatu acuan.

Kata kunci : statistika, sistem informasi, sistem informasi geografis, pengolahan data, analisis citra.

1. PENDAHULUAN

Sistem informasi adalah suatu komponen yang memungkinkan kita untuk membentuk suatu sistem dengan cara mengolah data, mengumpulkannya, menyimpan, menganalisis, sehingga mampu menghasilkan suatu informasi dalam bidang tertentu. Sistem informasi sangat dibutuhkan sebagai suatu sumber data untuk kebutuhan informasi.

Sistem informasi juga dapat diimplementasikan dalam berbagai bidang keilmuan, salah satunya sistem informasi. Sistem informasi dapat berkoordinasi dengan ilmu pemetaan bumi sehingga dapat menghadirkan sistem informasi geografis yang berfokus pada pemetaan posisi-posisi permukaan bumi.

Dalam proses penyediaan sistem informasi geografis dibutuhkan suatu alur prosedur yang memungkinkan semua komponen pembentuk data dapat terintegrasi dengan sempurna untuk mencapai tujuan yang spesifik. Komponen-komponen pembentuk data adalah input, model, output, basis data, dan komponen pengendali. Kesemua komponen ini perlu diproses dengan baik untuk dapat dijadikan suatu data yang sah, disinilah ilmu statistika turut berperan penting.

Statistika adalah suatu ilmu yang memungkinkan kita untuk menerka-nerka jawaban berdasarkan informasi yang tak lengkap. Hal ini sangat sesuai dengan kebutuhan sistem informasi geografis yang menuntut kita untuk

mengolah banyak data menjadi suatu informasi yang bersifat fleksibel, efektif, dan efisien.

2. PEMBAHASAN

2.1. Sistem Informasi Geografis

Salah satu contoh penerapan sistem informasi terdapat pada penyediaan sistem informasi geografis. Suatu sistem yang mengolah segala informasi mengenai keadaan permukaan bumi. Dewasa ini sudah terdapat bermacam-macam teknologi yang memungkinkan kita untuk memproses dan membuat pemetaan bumi. Teknologi itu kita kenal dengan nama *Global Positioning System (GPS)*, *Remote Sensing (RS)*, dan *Geographic Information System (GIS)*.

Global Positioning System (GPS) adalah suatu sistem yang memungkinkan satelit untuk berotasi pada bumi dan memetakan lokasi pada permukaan bumi dengan memanfaatkan letaknya berdasarkan koordinat.

Remote Sensing (RS) adalah teknologi yang memanfaatkan pesawat udara untuk mengambil gambar permukaan bumi.

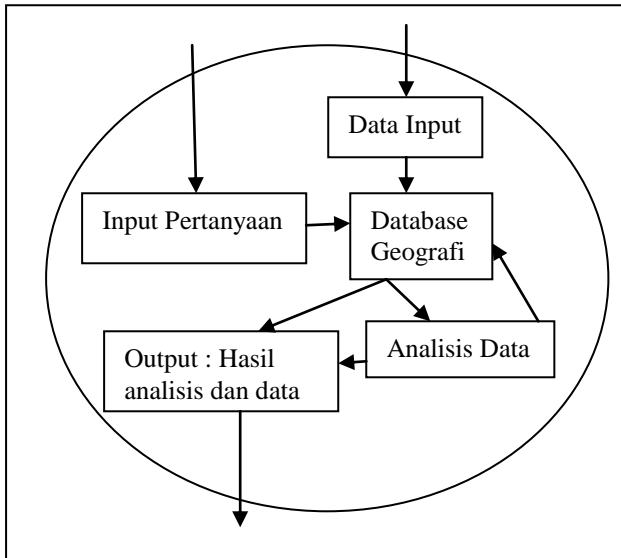
Kedua sistem diatas adalah alat yang dipergunakan *Geographic Information System (GIS)* untuk mendapatkan data yang kemudian akan ia olah dan manipulasi (menggunakan ilmu statistika) untuk kemudian ditampilkan sebagai informasi tentang geografi secara lengkap.

Penyediaan Sistem Informasi Geografi dilatarbelakangi kegunaannya yang variatif. Fungsi utama *Geographic Information System (GIS)* atau Sistem Informasi Geografi (SIG) adalah :

1. Sebagai suatu sistem untuk mengintegrasikan analisis spasial.
2. Sebagai alat untuk mengkomposisikan data geografis dengan cara mengambil, menyimpan, mengamati, menganalisis, memodifikasi dan kemudian menampilkannya.
3. Sebagai media penunjang untuk penyedia data yang dapat digunakan untuk kebutuhan berbagai operasi dan penelitian ilmiah.

Sistem Informasi Geografis merupakan salah satu sistem yang dapat diaplikasikan dalam berbagai sektor kehidupan. Ia menunjang keberlangsungan pembangunan manufaktur suatu wilayah, sebagai penyedia data yang dapat membantu dalam masalah *crime analysis*, *emergency management*, penentuan rute penerbangan, menentukan lokasi penambangan, dan berbagai hal lainnya.

Untuk mengoperasikan Sistem Informasi Geografis dibuat suatu sistem yang mampu mengintegrasikan inputan sekaligus outputan data dengan baik. Di bawah ini akan digambarkan sistem pengoperasian Sistem Informasi Geografis dalam bentuk graf.

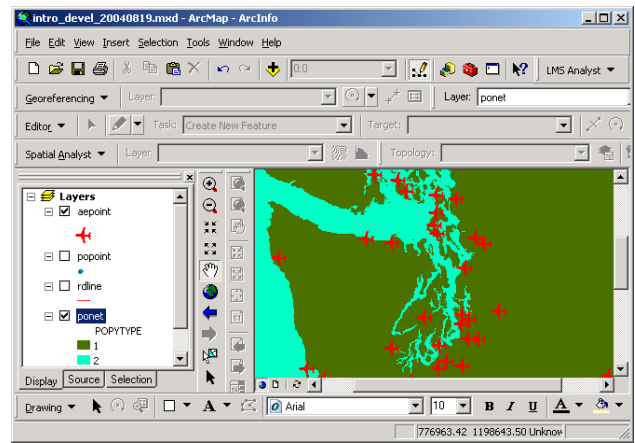


Gambar 2.1.1 Jalur penyediaan data Sistem Informasi Geografis

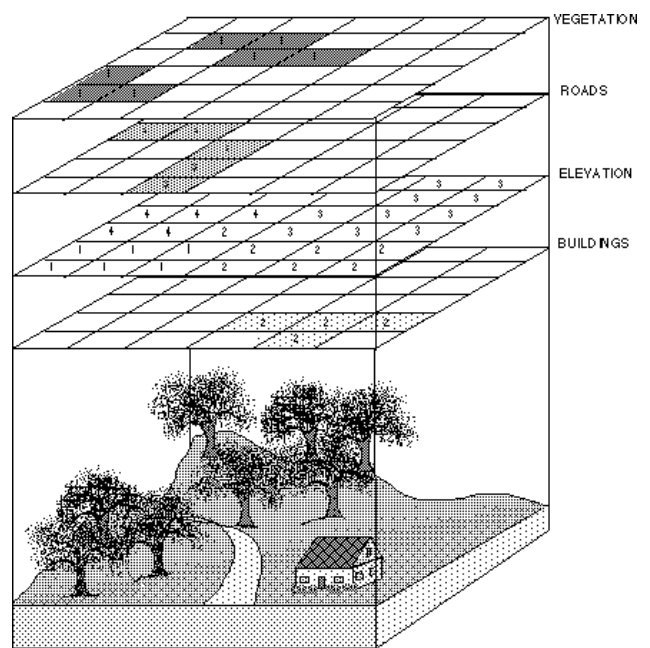
Dalam pembuatan database geografi kita perlu mengerti benar tentang penyediaan data spasial yang melingkupi tentang lapisan-lapisan pada bumi dan posisi-posisi benda pada permukaan bumi. Untuk menggambarkan lapisan dan permukaan bumi, diterapkan dua cara pendekatan yaitu :

1. Vektor : menggambarkan melalui garis dan polygon.
2. Raster : menggambarkan melalui penyediaan ruang gambar berupa cell-cell persegi.

Kedua tipe data ini keduanya dapat merepresentasikan keadaan permukaan bumi dengan sama baiknya. Perbedaan terletak pada hasil keluaran yang dapat diterima oleh pengguna.



Gambar 2.1.2 Penggunaan vector dalam penyediaan sistem informasi geografis.



Gambar 2.1.3 Model diagram dari keadaan nyata bumi menggunakan dataset raster.

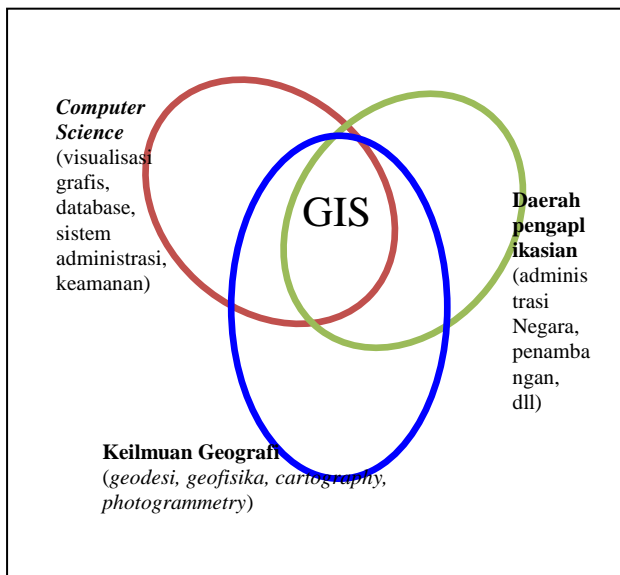
Pada Sistem Informasi Geografi terdapat beberapa faktor yang sangat berpengaruh dalam penilaian kelayakan penyediaan data spasial. Faktor-faktor tersebut adalah :

1. **Proyeksi** : Metode untuk memproyeksikan permukaan 3 dimensi kedalam koordinat X-Y sehingga dapat dipandang sebagai suatu bentuk 2 dimensi.
2. **Skala** : Rasio jarak antara kenyataan dengan penggambaran. Disinilah ilmu statistic berperan, dalam pembuatan skala selalu dicari suatu kisaran implicit yang dapat mempresentasikan kehidupan nyata kedalam gambar.
3. **Akurasi** : Nilai perbandingan antara database dengan kenyataan. Terdapat 3 tolak ukur untuk menentukan keakuratan suatu database yaitu :

- Seberapa tepatnya posisi yang terpetakan dengan keadaan nyata.
- Kekonsistenan antara peta dengan keadaan nyata. Misalnya penggambaran jalan antara Jakarta dan Bandung, harus tepat menghubungkan Jakarta dan Bandung dengan keadaan sama dengan pada kenyataan.
- Seberapa detail data data itu ditampilkan. Seberapa besar ruang gambar data, melingkupi semua daerah.

4. **Resolusi** : Kemampuan suatu data untuk menampilkan fitur terkecil dari sistem tersebut. (Berhubungan dengan raster data dan ukuran *pixel*)

Untuk memberikan nilai kelayakan yang tinggi dan memastikan data yang ditampilkan merupakan data yang akurat, maka dalam penyediaan Sistem Informasi Geografi, dibutuhkan keterampilan dari berbagai bidang keilmuan.



Gambar 2.1.4 Ilmu-ilmu yang mendasari Sistem Informasi Geografi

2.2. Penerapan Statistik pada Sistem Informasi Geografis

Secara harfiah statistika adalah suatu ilmu bagaimana kita menkoordinasikan segala data yang telah kita kumpulkan yang kemudian kita integrasi sehingga mampu mempresentasikan data secara keseluruhan. Statistik sendiri adalah suatu ilmu yang memiliki kemampuan untuk menghitung ketidakpastian dengan tepat. Dengan kemampuan itu, ahli statistik mampu membuat pernyataan tegas tentang suatu hal dan lengkap dengan jaminan tingkat kepastian.

Untuk mendapatkan hasil statistic yang baik, seorang ahli statistic perlu berpegang pada 3 hal pokok yang sangat berkaitan dengan dengan statistik, yaitu :

1. Analisis Data
2. Probabilitas
3. Kesimpulan Statistik

2.2.1. Analisis Data Unsur Geografis

Analisis data adalah bagaimana kita mengumpulkan, menyajikan dan mencari data.

Data yang sering kita gunakan untuk menafsirkan realitas adalah bahan mentah. Dalam analisis data sistem informasi geografis dilakukan 4 cara pendekatan yaitu :

1. **Sistematis** : Mengelompokkan pengetahuan geografis menjadi katagori yang kemudian dibahas secara global.
2. **Regional** : Mempelajari hubungan sistematis antara pengetahuan geografis yang telah kita miliki dengan keadaan wilayah tertentu di atas planet.
3. **Deskriptif** : Menjelaskan secara sederhana lokasi suatu masalah beserta populasinya.
4. **Analitis** : Menjawab masalah mengenai populasi pada wilayah geografis tertentu.

Pengetahuan geografis yang kita butuhkan menyangkut unsur fisik geografis berupa atmosfer, litosfer, hidrosfer, pedosfer, laut dan pesisir, biosfer, dan sumber daya alam.

Atmosfir adalah susunan dari beberapa lapisan yang menutupi bumi dari permukaan planet tersebut hingga mencapai jauh dari permukaan. Data data mengenai atmosfer dapat kita dapatkan melalui ilmu atmosfer yang meliputi bidang meteorology, klimatologi dan aeronomi. Sedangkan litosfer adalah lapisan bumi yang paling luar atau kita kenal dengan nama kulit bumi. Dalam kajian tentang litosfer, kita akan mempertimbangkan mengenai struktur lapisan kulit bumi, seisme, proses pengikisan, dan pengendapan. Data data mengenai litosfer dapat kita dapatkan dari keilmuan geologi, geofisika, dan geodesi. Unsur ketiga adalah hidrosfer yaitu lapisan air dipermukaan bumi. Pada bagian hidrosfer kita adapat mempelajari mengenai potensi air permukaan, air tanah, zona pesisir dan laut, morfologi laut, gerakan air laut, dan kualitas fisik air laut. Kesemua data tentang laut dapat kita pelajari berdasarkan bidang keilmuan oseanografi dan hidrologi. Pada bagian tentang pedosfer atau pemahaman tentang tanah kita dapat mengambil banyak datm menggunakan ilmu tanah yang mempelajari lapisan kulit bumi yang terlibat dalam proses pembentukan tanah, disiplin utama dari ilmu tanah adalah edafologi dan pedologi. Hal yang tidak kalah penting adalah data data mengenai laut dan pesisir, hal ini melibatkan keilmuaan mengenai hidrosfer sekaligus pedosfer. Selanjutnya kita juga membutuhkan data-data mengenai biosfer. Biosfer adalah tempat yang memungkinkan kehidupan dan proses biotic berlangsung. Biosfer yang merupakan bagian luar dari planet bumi mencakup hidrosfer, pedosfer, dan atmosfer. Selanjutnya adalah data data yang menyangkut mengenai sumber daya alam.

Kesemua data yang telah kita masukkan dalam berbagai katagori tersebut kemudian akan diproses dengan berbagai prosedur termasuk dengan pengaplikasian ilmu statistik.

Misalnya dalam pencitraan ketinggian suatu daerah dalam sistem operasi geografi. Pertama kita akan mengumpulkan data ketinggian suatu wilayah. Dalam pengaplikasian kedalam sistem digital, kita tidak mungkin menggambarkan ketinggian tempat tersebut dengan sempurna pada setiap centimeter nya. Kita perlu suatu nilai yang dapat merepresentasikan lebih atau kurangnya suatu nilai tersebut. Disini kita menggunakan apa yang ilmu statistik sebut sebagai 'mean'.

Mean adalah nilai rata-rata yang diwakili oleh \bar{x} yang didapat dengan menjumlahkan semua data lalu membagi dengan jumlah pengamatan.

$$\bar{x} = \frac{\text{jumlah data}}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Gambar 2.2.1.1 Rumus Mean

Kita dapat menyingkat cara untuk menyatakan $x_1 + x_2 + \dots + x_n$ dengan huruf kapital Yunani Sigma (\sum).

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

Gambar 2.2.1.2 Rumus Mean dalam Bentuk Sigma

Selain menggunakan rumus statistik berupa pencarian nilai rata rata, kita jugamembutuhkan tools lain berupa penghitungan sebaran. Sebaran data adalah seberapa jauh data cenderung merentang dari pusat. Metode pencarian sebaran yang dapat diterapkan dalam penyimpanan data geografis adalah rentang interkuartil. Inti utama dari pencarian renatang interkuartil adalah dengan membagi data menjadi empat kelompok yang sama besar dan melihat sejauh mana kelompok ekstrim terpisah.

Dalam penggunaan metode interkuartil kita juga memerlukan tools statistic bernama median. Median sendiri adalah "titik tengah" data. Hal ini akan memudahkan pemakian dalam pengelompokan secara kuartil. Apabila semua hal diatas telah sanggup diimplementasikan. Maka kita adapat menghitung median menggunakan rentang interkuartil dengan rumus :

$$IQR = Q_3 - Q_1$$

Gambar 2.2.1.3 Rumus Rentang Interkuartil

Ukuran baku untuk sebaran adalah simpangan baku. Berbeda dengan IQR yang bertumpu pada median, simpangan baku mengukur sebaran mean, sehingga kita dapat menggangapnya sebagai jarak rata-rata dari mean.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Gambar 2.2.1.4 Rumus Varian Sample

$$S = \sqrt{S^2}$$

Gambar 2.2.1.5 Rumus Simpangan Baku

Dibutuhkan juga suatu tools untuk menghitung frekuensi relatif suatu peristiwa dalam jangka panjang. Oleh karena itu dirumuskan :

$$\mu = \sum xp(x)$$

Gambar 2.2.1.6 Rumus untuk Distribusi Probabilitas

Ini juga disebut dengan nilai X yang diharapkan, atau E[x]. Dianggap setiap nilai memiliki bobot probabilitas.

Hal tidak kalah penting dari penerapan rumus-rumus diatas adalah mengenai pemilihan sampel. Karena data yang akan kita proses merupakan data yang sangat banyak, maka untuk mendapatkan data yang dibutuhkan dengan mengalokasikan memori yang sesuai, kita memerlukan apa yang disebut dengan pengambilan sampel. Cara untuk mendapatkan hasil yang diandalkan secara statistic, kita perlu mengambil sampel secara acak.

Sampling sistematis dimulai dari unit yang dipilih secara acak lalu mengambil setiap kelipatan k unit berikutnya. Tanpa adanya pola yang acak pada perhitungan statistika maka jaminan statistic terhadap akurasi penjajakan yang dilakukan tidak akan terlalu baik.

Metode yang umum digunakan dalam pemilihan sampel adalah metode sampel peluang (opportunity sample). Tanpa membuat desain prosedur, penambil sampel peluang hanya perlu mengambil n unit populasi yang pertama kali ditemui.

Perlu diingat untuk pemilihan sample dibutuhkan dua sifat sampel untuk menghasilkan sampel yang baik :

1. Tanpa bias : setiap unit mempunyai peluang yang sama untuk terpilih.
2. Independensi : pemilihan sebuah unit tidak mempengaruhi pemilihan unit lainnya.

Setelah semua sampel dikumpulkan, didata, dan dianalisis, maka diperlukan suatu kesimpulan statistik. Untuk membuat kesimpulan statistik, dibuat sebuah model regresi untuk seluruh populasi yang digambarkan dalam suatu hubungan linier :

$$y = \alpha + \beta x + \epsilon$$

y = variable acak dependen
 α, β = parameter yang tidak diketahui dan akan dicari
 x = variable independen
 ϵ = fluktuasi kesalahan acak

Gambar 2.2.1.7 Model Regresi

Dalam praktek, distribusi ϵ akan berbeda untuk nilai x yang berbeda ; ketinggian tanah 150 cm mungkin beratnya tidak terlalu bervariasi dibanding ketinggian tanah 180 cm. Meskipun demikian kita dapat memakai asumsi untuk menyederhanakannya. Dianggap semua nilai x , besarnya ϵ independen, normal, dan memiliki simpangan baku $\sigma = \sigma(\epsilon)$ dan *mean* yang sama $\mu = 0$.

Kadang-kadang, kesalahan baku dan ambang keyakinan tidak dapat ditemukan. Muncullah *resampling* (sampling-ulang) suatu teknik yang memperlakukan sampel seolah-olah suatu populasi. Teknik seperti ini banyak dikenal sebagai *randomisasi*, pisau lipat (*jackknife*), dan ikatan sepatu boot (*bootstrapping*).

Metode ini juga digunakan dalam penyediaan sistem informasi geografis. Disinilah ilmu computer dan statistic saling berkolaborasi. Statistik sebagai landasan ilmu, dan computer sebagai tools yang dapat membantu memudahkan suatu proses *sampling*.

Untuk melakukan *resampling*, computer melakukan beberapa prosedur. Yaitu :

1. Mengambil sampel dari sampel
2. Membuat perkiraan untuk sampel ulang
3. Mengulang dua langkah di atas berulang kali, mencari sebaran perkiraan sampel ulang.

2.2.2. Analisis Citra

Dalam menampilkan gambar dalam penyediaan database sistem informasi geografi, gambar yang akan ditampilkan sebelumnya diuji dengan menggunakan analisis citra. Analisis citra atau *image analysis* adalah proses mengubah suatu citra (dalam hal ini hasil foto permukaan bumi) ke dalam bentuk digital.

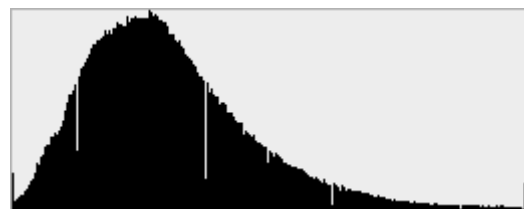
Sebuah citra computer bisa terbentuk dari 1000 kali 1000 piksel, dengan setiap titik data mewakili satu dari 16,7 juta warna di sembarang data. Analisis citra statistik berusaha untuk mencari makna "informasi" seperti ini. Informasi-informasi yang dapat diidentifikasi berupa ukuran file citra, jumlah pixel, dimensi citra, dan resolusi citra. Untuk menampilkan citra yang berkualitas baik, diperlukan beberapa dasar dalam proses *pen-digitalisasi-an*. Hal tersebut berupa histogram, citra brightness, contrast, dan *histogram equalization*.

Histogram adalah fungsi yang menunjukkan jumlah titik pada suatu citra dengan menggunakan pendekatan pada koordinat kartesius. Sumbu X (absis) merepresentasikan

tingkat warna dan sumbu Y (ordinat) merepresentasikan frekuensi kemunculan titik.



Gambar 2.2.2.1 Contoh Penyajian Sistem Informasi Geografis

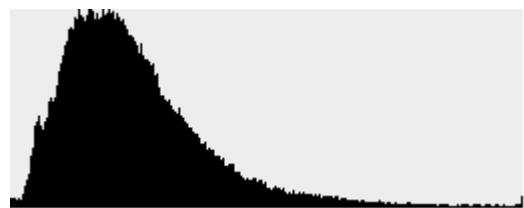


Gambar 2.2.2.2 Histogram Dari Gambar 2.2.2.1

Nilai dari histogram akan berubah seiring kita mengubah tampilan citra.



Gambar 2.2.2.3 Penyajian Gambar 2.2.2.2 dengan Tipe Grayscale



Gambar 2.2.2.4 Histogram Dari Gambar 2.2.2.3

Perbedaan bentuk histogram dapat kita lihat dari gambar 2.2.2.2 dan 2.2.2.4. Perbedaan ini disebabkan oleh adanya perbedaan tampilan warna pada kedua citra.

Untuk melakukan perubahan warna citra, dari citra berwarna menjadi citra hitam-putih, diterapkan ilmu statistic berupa rata-rata atau *mean*.

Pertama dilakukan pencarian nilai rata-rata dari warna primer yaitu biru, hijau, dan merah. Kemudian dilakukan konversi untuk mendapatkan nilai abu.

$$S = \frac{R + G + B}{3}$$

S = nilai rata-rata warna abu
R = nilai rata-rata warna merah
G = nilai rata-rata warna hijau
B = nilai rata-rata warna biru

Gambar 2.2.2.5 Rumus Mengubah Citraan Warna Menjadi *Grayscale*

3. KESIMPULAN

Statistik merupakan suatu alat bantu dalam proses pemecahan suatu masalah. Berbagai perhitungan sederhana berupa pencarian *mean*, simpangan baku, atau sebaran, dapat dikembangkan sehingga dapat digunakan untuk pemecahan masalah yang lebih kompleks. Salah satu penggunaan ilmu statistik dalam penyediaan sistem informasi adalah dalam proses analisis data untuk sistem informasi geografis dan analisis citra dalam visualisasi sistem informasi geografis. Ilmu statistik memungkinkan terciptanya database sistem informasi geografis karena memungkinkan penintegrasian banyak data menjadi suatu nilai yang mempresentasikan keseluruhan data. Hal ini membuat database menjadi lebih sederhana. Penyediaan jumlah sampel beserta nilai sebarannya yang telah dikalkulasi dengan baik dapat memungkinkan terciptanya hasil data sistem informasi geografis yang baik.

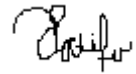
REFERENSI

- [1] Gonick, Larry . dan Smith, Wollcott, *Kartun Statistik*. Jakarta : Kepustakaan Populer Gramedia, 2002.
- [2] Walpole, *Probability & Statistics for Engineers & Scientists*. London : Pearson Education, 2007.
- [3] "Sistem Informasi Geografi (SIG)/ Geographic Information System (GIS)", URL : www.mbojo.wordpress.com/2007/04/08/sitem-informasi-geografi-sig, diakses tanggal 14 Desember 2010 pukul 20:00.
- [4] "Contoh Sederhana Analisis Citra Digital. Oleh : Ismail Adha Kesuma.", URL : www.scribd.com/doc/29586175/Contoh-Sederhana-Analisis-Citra-Digital, diakses tanggal 14 Desember 2010 pukul 20:00.

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 Desember 2010



Latifa Dwiyantri
18209010