TUGAS AKHIR – EC184801

SISTEM INFORMASI GEOGRAFIS BERBASIS GOOGLE EARTH API MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS

Nicky Permana Putra
NRP 07211540000014

Dosen Pembimbing
Dr.Supeno Mardi Susiki Nugroho,ST.,MT.
Susi Juniastuti, ST., M.Eng.

DEPARTEMEN TEKNIK KOMPUTER
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2020

TUGAS AKHIR – EC184801

SISTEM INFORMASI GEOGRAFIS BERBASIS GOOGLE EARTH API MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS

Nicky Permana Putra
NRP 07211540000014

Dosen Pembimbing
Dr.Supeno Mardi Susiki Nugroho,ST.,MT.
Susi Juniastuti, ST., M.Eng.

DEPARTEMEN TEKNIK KOMPUTER
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya

LEMBAR PENGESAHAN

Sistem Informasi Geografis Berbasis Google Earth API Menggunakan Algoritma *K-Nearest Neighbors*

Tugas Akhir ini disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana Teknik di Institut Teknologi Sepuluh Nopember Surabaya

Oleh: Nicky Permana Putra (NRP: 07211540000014)

Tanggal Ujian : 9 Juli 2020 Periode Wisuda : September 2020

Disetujui oleh:

Dr. Supeno Mardi Susiki Nugroho, ST., MT. (Pembimbing I)
NIP: 19700313 199512 1 001

Susi Juniastuti, ST., M.Eng. (Pembimbing II)
NIP. 19650618 199903 2 001

Prof. Dr. Ir. Yoyon Kusnendar Suprapto, M.Sc. (Penguji I)
NIP: 19540925 197803 1 001

Mochamad Hariadi, ST., M.sc., Ph.D. (Penguji II)
NIP. 19691209 199703 1 002

Reza Fuad Rachmadi, ST., M.T. (Penguji III)
NIP. 19850403 201212 1 001

Mengetahui
Kepala Departemen Teknik Komputer

Dr. Supeno Mardi Susiki Nugroho, ST., MT.
NIP: 19700313199512 1001

ABSTRAK

Nama Mahasiswa	:	Nicky Permana Putra
Judul Tugas Akhir	:	Sistem Informasi Geografis Berbasis Google Earth API menggunakan Algoritma *K-Nearest Neighbors*
Pembimbing	:	1.Dr. Supeno Mardi Susiki Nugroho, ST., MT 2.Susi Juniastuti, ST., M.Eng.

Sistem Informasi Geografis merupakan sistem komputer yang memiliki kemampuan untuk membangun, menyimpan, mengelola dan menampilkan informasi berupa peta geografis suatu daerah. Sistem Informasi Geogafis ini akan digunakan untuk melakukan pengelompokan statistik – statistik daerah dan dikelompokkan pada kategori yang ada untuk kemudian ditampilkan dalam bentuk peta tematik atau peta digital menggunakan algoritma K-Nearest Neighbor. Data – data statistik yang nantinya akan dimasukkan dapat berupa data tingkat kemiskinan, data tingkat pengangguran dari suatu daerah, dll. Algoritma K-NN akan mengelompokkan data sesuai kategori dan saat data-data tersebut sudah terkelompokkan maka peta akan muncul dengan daerah yang sudah diberi warna dimana daerah-daerah tersebut sudah tergolongkan sesuai dengan kategorinya. Hal ini dapat diterapkan di pemerintahan guna untuk mempermudah dalam pengambilan kebijakan dalam suatu daerah.

Kata Kunci: Sistem Informasi Geografis, GIS, *Google Earth* , API, Klasifikasi, *K-Nearest Neighbors.*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT atas segala limpahan berkah dan hidayah-Nya, penulis dapat menyelesaikan penelitian ini dengan judul **"Sistem Informasi Geografis Berbasis Google Earth API menggunakan Algoritma *K-Nearest Neighbors*"**

Penelitian ini disusun dalam rangka pemenuhan bidang riset di Departemen Teknik Komputer, FTEIC-ITS serta digunakan sebagai persyaratan menyelesaikan pendidikan S1. Penelitian ini dapat terselesaikan tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Keluarga, Ibu, Bapak dan Saudara tercinta yang telah memberikan dorongan spiritual dan material dalam penyelesaian buku penelitian ini.
2. Bapak Dr. Supeno Mardi Susiki Nugroho, S.T., M.T. dan Ibu Susi Juniastuti, ST., M.Eng. selaku dosen pembimbing yang telah meluangkan waktu untuk berdiskusi, memberi semangat, arahan dan nasehat selama proses pengerjaan tugas akhir ini.
3. Bapak dan ibu dosen Departemen Teknik Komputer atas pengajaran, bimbingan, serta perhatian yang diberikan kepada penulis selama ini.
4. Dan teman-teman Teknik Komputer 2015, teman-teman E55 serta teman-teman Laboratorium Komputasi Multimedia B401 yang tidak dapat saya sebutkan satu-persatu.

Kesempurnaan hanya milik Allah SWT, untuk itu penulis memohon segenap kritik dan saran yang membangun. Semoga penelitian ini dapat memberikan manfaat bagi kita semua. Aamiin.

Surabaya, 29 Oktober 2019

Penulis

Halaman ini sengaja dikosongkan

DAFTAR ISI

Abstrak **i**

KATA PENGANTAR **iii**

DAFTAR ISI **v**

DAFTAR GAMBAR **vii**

DAFTAR TABEL **ix**

NUMENKLATUR **xi**

1 PENDAHULUAN **1**
1.1 Latar belakang 1
1.2 Permasalahan 1
1.3 Tujuan 2
1.4 Batasan masalah 2
1.5 Sistematika Penulisan 2

2 TINJAUAN PUSTAKA **5**
2.1 *Related Works* 5
2.1.1 *Soil Data Clustering by Using K-means and Fuzzy K-means Algorithm* 5
2.1.2 *Clustering with GIS : An Attempt to Classify Turkish District Data* 7
2.2 Teori Penunjang 9
2.2.1 Sistem Informasi Geografis 9
2.2.2 *Application Programming Interface* (API) . . 10
2.2.3 Data Sistem Informasi Geografis 11
2.2.4 *K-Nearest Neighbors* (KNN) 13

3 DESAIN DAN IMPLEMENTASI SISTEM **15**
3.1 Desain Sistem 15
3.2 Alur Implementasi Sistem 18
3.3 Request *Google Earth Engine* API dengan QGIS . . 19
3.4 Pembagian Kecamatan di Kota Surabaya 20

3.5 Pembentukan Data Vektor dan Data Atribut Kecamatan 21
3.6 Penggabungan Data Vektor dengan Data Atribut menjadi Data Tabular 22
3.7 Konversi Data Tabular menjadi Data CSV pada QGIS 26
3.8 Training Data untuk Klasifikasi dengan Algoritma KNN 27
3.8.1 Import *library* dan Komponen-Komponen Python 27
3.8.2 Import Dataset CSV 28
3.8.3 *Split dataset* menjadi *Test* dan *Train* 28
3.8.4 Melakukan Proses *Scaling* Dataset 29
3.8.5 Melakukan *Training* Data 29
3.9 Visualisasi Peta Hasil Klasifikasi 30
3.10 Proses Deployment Aplikasi 30

4 PENGUJIAN DAN ANALISA 31
4.1 Pengujian Sistem Klasifikasi 31
4.1.1 Pengujian Akurasi Sistem Klasifikasi 31
4.1.2 Pengujian Klasifikasi Pada Sistem Algoritma K-NN 33
4.1.3 Pengujian Hasil klasifikasi Pada Peta Digital ditingkat Kecamatan dan Kelurahan 39

5 PENUTUP 45
5.1 Kesimpulan 45
5.2 Saran 45

DAFTAR PUSTAKA 47

DAFTAR GAMBAR

2.1 Metodologi *clustering* menggunakan *fuzzy k-means algorithm* [1] 6
2.2 Peta Digital beserta *marker* [1] 6
2.3 Hasil Klasifikasi algoritma *K-Means* [2] 8
2.4 Hasil Klasifikasi algoritma SOM [2] 8
2.5 Sistem Informasi Geografis QGIS 10
2.6 Data Atribut berupa Tabel 12
2.7 Data Vektor berupa *polygon* 13
2.8 Data Raster berupa *Image* 13

3.1 Blok Diagram Sistem 16
3.2 QGIS dengan *integrated* GEE pyhton API 19
3.3 Kota Surabaya 20
3.4 Shapefile dari Kecamatan Tegalsari 21
3.5 Titik dari Kecamatan Tegalsari yang Berisi Data Atribut 22
3.6 Proses Penggabungan Data Atribut dengan Data vektor 23
3.7 Shapefile Gabungan dari Kecamatan Tegalsari . . . 24
3.8 Shapefile Gabungan dari beberapa Kecamatan . . . 25
3.9 Proses Konversi Data 26
3.10 Proses Import Library 27
3.11 Proses Import Data CSV 28
3.12 Proses *Split* Data 28
3.13 Proses *Scaling* Data 29
3.14 Proses *Training* Data 29
3.15 Proses visualisasi Peta Digitial 30

4.1 Proses *Test* Akurasi Nilai K 32
4.2 Plot Hasil Akurasi Nilai K terhadap Nilai K 33
4.3 Proses Klasifikasi dengan K=3 dan metode Minkowski [3] 33
4.4 Proses Klasifikasi dengan K=7 dan metode Minkowski [3] 34

4.5 Proses Klasifikasi dengan K=15 dan metode Minkowski [3] 34
4.6 Hasil Klasifkasi Skema 1 35
4.7 Hasil klasifikasi Skema 2 36
4.8 Hasil klasifikasi Skema 3 37
4.9 Hasil klasifikasi Skema 4 38
4.10 Zona-zona yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Maret 2020 . . 39
4.11 Zona-zona yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 April 2020 . . . 40
4.12 Zona-zona yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Mei 2020 41
4.13 Zona-zona pada kelurahan yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Maret 2020 pada tingkat Kelurahan 42
4.14 Zona-zona pada kelurahan yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Mei 2020 43

DAFTAR TABEL

3.1 Struktur Tabel CSV 26

Halaman ini sengaja dikosongkan

NUMENKLATUR

Xi	: Koordinat X titik i (Tidak ada satuan)
Yi	: Koordinat Y titik i (Tidak ada satuan)
i	: Titik yang merepresentasikan data tertentu (Tidak ada satuan)
n	: Banyak data (Tidak ada satuan)
q	: Jarak Chebysev (Tidak ada satuan)
K	: Jumlah Klaster (Tidak ada satuan)

Halaman ini sengaja dikosongkan

BAB 1
PENDAHULUAN

Penelitian ini di latar belakangi oleh berbagai kondisi yang menjadi acuan. Selain itu juga terdapat beberapa permasalahan yang akan dijawab sebagai luaran dari penelitian.

1.1 Latar belakang

Perkembangan teknologi informasi hari ini berkembang dengan sangat pesat. Internet adalah salah satu produk teknologi informasi yang selalu di konsumsi oleh mayoritas manusia di muka bumi, namun masih ada kurangnya informasi pada suatu daerah tertentu baik itu terkait dengan permasalahan sosial, ekonomi atau geografis. Hal ini menyebabkan beberapa daerah mengalami perkembangan yang lambat dibandingkan dengan daerah lain. Perkembangan internet juga banyak membawa perkembangan di segala aspek kehidupan manusia saat ini, salah satunya adalah Geographic Information System atau bisa disebut SIG, namun SIG masih jarang digunakan untuk menganalisa data – data statistik yang ada pada suatu daerah. Sebelum adanya Sistem Informasi Geografis (SIG) ini, sejumlah informasi tentang geografis suatu daerah disajikan secara manual.

Untuk menghasilkan peta dengan hasil kategori yang komprehensif dapat digunakan metode – metode pengelompokkan data. Metode untuk mengelompokkan data-data tertentu tersebut dapat menggunakan salah satu algoritma pengelompokan yaitu K-Nearest Neighbors dengan hasil berupa kategori tertentu yang sudah ditentukan lalu data yang sudah tergolongkan akan direpresentasikan dengan warna tertentu. Data tersebut nantinya akan ditampilkan dalam bentuk peta digital dimana setiap daerah akan diberi warna sesuai data yang sudah tergolongkan. Hal ini dapat memudahkan pemerintah daerah untuk mengetahui informasi pada daerah tersebut.

1.2 Permasalahan

Berdasarkan latar belakang tersebut adanya permasalahan berupa kurangnya informasi terkait suatu daerah tertentu, kemudian

juga kurangnya *tools-tools* dalam menganalisa data-data statisik dan data spasial suatu daerah tersebut, serta kurangnya penggunaan sistem informasi geografis di pemerintahan.

1.3 Tujuan

Tujuan dari Tugas Akhir ini yaitu untuk menampilkan informasi atau visualisasi data covid-19 pada tingkat kecamatan di Surabaya menggunakan *Google Earth API* dan algoritma *K-Nearest Neighbors*.

1.4 Batasan masalah

Untuk memfokuskan masalah yang akan diangkat maka perlu dilakukan pembatasan masalah. Batasan-batasan tersebut diantara lain:

1. Dasar aplikasi yang digunakan adalah Google Earth API.
2. sistem hanya mengelompokkan data yang sudah terinputkan.
3. Daerah yang ditentukan yaitu Kota Surabaya.
4. Hanya beberapa kecamatan yang akan digunakan

1.5 Sistematika Penulisan

Laporan penelitian tugas akhir ini disusun secara sistematis dan memiliki struktur yang mudah dipahami oleh pembaca maupun seseorang yang hendak meneruskan penelitian ini. Alur sistematika penulisan laporan penelitian ini yaitu :

1. BAB I Pendahuluan Bab ini berisi uraian tentang latar belakang permasalahan, permasalahan yang terjadi, tujuan penelitian, batasan masalah dan sistematika laporan.
2. BAB II Tinjauan Pustaka Bab ini berisi tentang uraian teori-teori yang berkaitan dengan permasalahan yang dibahas dalam penelitian ini. Teori-teori ini digunakan sebagai dasar dalam penilitian, yaitu Sistem Informasi Geografis, penggunaan *Application Programming Interface* (API), Data Spasial, dan *K-Nearest Neighbors*(K-NN).
3. BAB III Desain dan Implementasi Sistem Bab ini berisi penjelasan-penjelasan terkait eksperimen yang dilakukan dan langkah-langkah pengambilan data speta digital dan pengelompokan data hingga ditampilkan dalam suatu peta digital.

4. BAB IV Pengujian dan Analisa Bab ini menjelaskan tentang pengujian eksperimen yang dilakukan terhadap proses pengambilan data peta digital, pengklasifikasian data spasial dan analisanya.
5. BAB V Penutup Bab ini berisi kesimpulan yang diambil dari penelitian dan pengujian yang dilakukan. Saran dan kritik yang membangun untuk mengembangkan penelitian lebih lanjut dituliskan pada bab ini.

Halaman ini sengaja dikosongkan

BAB 2
TINJAUAN PUSTAKA

Demi mendukung penelitian ini, dibutuhkan beberapa teori penunjang dan *related works* sebagai bahan acuan dan refrensi. Dengan demikian penelitian ini menjadi lebih terarah.

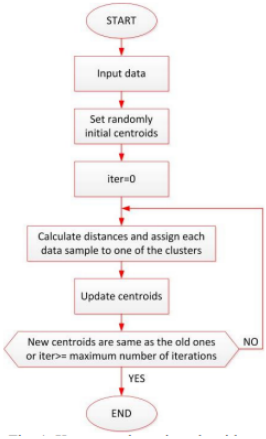
2.1 *Related Works*

Penelitian mengenai pengelompokan data spasial sudah pernah dilakukan sebelumnya. Beberapa referensi yang bisa dijadikan acuan dalam peneilitian ini akan dijelaskan pada beberapa subbab berikut.

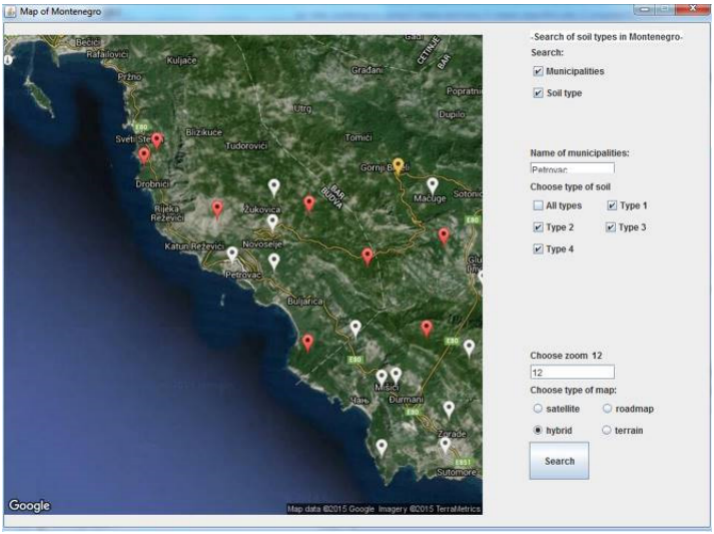
2.1.1 *Soil Data Clustering by Using K-means and Fuzzy K-means Algorithm*

Riset ini [1] mengembangkan suatu peta digital untuk menentukan *clustering* terhadap jenis tanah dengan menggunakan algoritma *fuzzy k-means* ini ditujukan untuk mempermudah dalam melakukan pengelompokan data tanah berdasarkan karakteristik unsur kimia didalam tanah dengan data representasi yang didapatkan dari lapangan. Data yang diinputkan berupa *database* yang berisi dengan karakteristik tanah selanjutnya ditentukannya titik pusat secara acak untuk dilakukan penghitungan *k-means*. Setelah titik pusat ditentukan akan dilakukan pengitungan dan menetapkan hasil penghitungan tersebut sebagai titik pusat baru. Proses ini akan diulang terus menerus sampai terbentuk titik pusat yang sama dengan data yang tersedia seperti yang digambarkan dalam gambar 2.1.

Data sampel yang ada di database kemudian diproyeksikan di *Google Maps* dimana data-data tersebut memiliki koordinat dan koordinat tersebut dikonversikan menjadi garis bujur dan garis lintang. Data sampel diberi label menggunakan *markers* dan warna dari *marker* berdasarkan pada jenis tanah yang ada dalam *database* [1]. Ini memungkinkan mencari peta dengan jenis tanah, serta menyesuaikan *zoom* pada peta tersebut. Berbagai jenis peta juga tersedia untuk proses visualisasi yaitu : peta jalan, satelit, *hybrid* dan *terrain* seperti gambar 2.2.



Gambar 2.1: Metodologi *clustering* menggunakan *fuzzy k-means algorithm* [1]



Gambar 2.2: Peta Digital beserta *marker* [1]

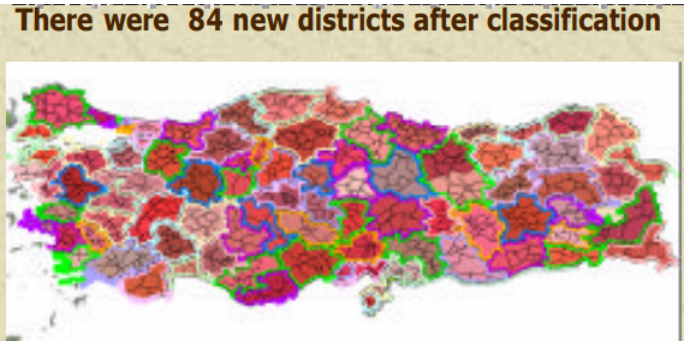
2.1.2 *Clustering with GIS : An Attempt to Classify Turkish District Data*

Riset ini [2] mengembangkan peta dengan klasifikasi dari data-data distrik dari kota turki menggunakan 2 metode yang berbeda yaitu teknik pengelompokan spasial dan non-spasial.Riset ini menggunakan beberapa software yaitu ArcGIS, CrimeStat, dan Matlab. Data-data geografis yang digunakan untuk membentuk peta digital untuk hasil akhir nantinya yaitu :

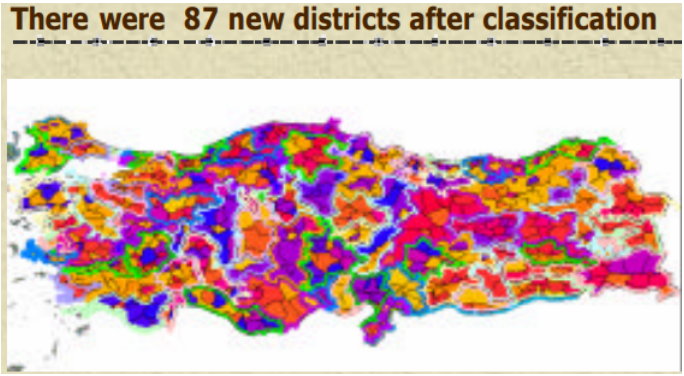
1. Peta Distrik berupa *Polygon Data*
2. Peta Titik Tengah dari distrik berupa *Point Data*
3. Peta Citra Satelit (*MRE Image*)
4. Peta Tektonik dari Kota Turki
5. Peta Perbatasan Kota Turki
6. Peta Perkembangan Cekungan Kota Turki
7. Peta Kontur Kota Turki berupa *Polygon Data*

Pengelompokan non-spasial menggunakan algoritma *K-Means* dan Pengelompokan spasial menggunakan algoritma SOM (*Self-Organizing Map.* Kelebihan daripada algoritma SOM daripada K-means yaitu algoritma SOM memiliki fitur "*Topology-Preserving*", memiliki tingkat efektifitas yang tinggi terhadap *High Dimensional Space*, serta dapat mengolah data baru lebih cepat, namun algoritma ini memiliki kekurangan pada bagian penggabungan data hasil pengelompokan menggunakan SOM dengan data geografis yang tersedia dan perlu adanya suatu visualisasi khusus terhadap data hasil pengelompokan algoritma SOM. Sedangkan algoritma *K-Means* lebih cepat dalam pengelompokan dan relatif mudah dimana pada riset ini dimana algoritma ini menentukan jarak-jarak terdekat antara titik tengah dari peta distrik-distrik Turki.

Data-data yang ada pada riset ini ada beberapa macam tergantung pada jenis algoritma yang digunakan. Algoritma *K-Means* menggunakan komponen PCA dari data tabular, sedangkan Algoritma SOM menggunakan 36 variabel yang telah dipiilih untuk dilakukan klasifikasi. Hasil dari kelasifikasi dari 2 metode ini memiliki 2 hasil yang berbeda dimana algortima *K-Means* menghasilkan 84 Distrik baru Seperti gambar 2.3 dan Algoritma SOM menghasilkan 87 distrik baru seperti gambar 2.4.



Gambar 2.3: Hasil Klasifikasi algoritma *K-Means* [2]



Gambar 2.4: Hasil Klasifikasi algoritma SOM [2]

2.2 Teori Penunjang

Dalam penelitian tugas akhir ini juga dibutuhkan beberapa teori penunjang sebagai bahan acuan dan referensi. Dengan adanya teori pendukung ini diharapkan penelitian tugas akhir akan menjadi lebih terarah.

2.2.1 Sistem Informasi Geografis

Sistem Informasi Geografis (SIG) atau *Geographic Information Sistem* (GIS) sendiri merupakan sistem berbasis komputer yang biasanya digunakan untuk menyimpan, memanipulasi, dan menganalisa informasi geografis. SIG dapat menggabungkan berbagai jenis data pada satu titik tertentu yang ada di bumi, menghubungkannya, menganalisanya, hingga memetakan hasilnya.Teknologi ini juga banyak digunakan untuk tujuan investigasi ilmiah,pengelolaan sumber daya, perencanaan rute, kartografi, dan perencanaan pembangunan. Ruang lingkup sistem informasi geografis terdiri atas :

1. Input Data
   Proses ini digunakan untuk menginputkan data spasial ataupun data non-spasial. Data spasial dapat berupa data analog ataupun digital, jika data yang ada berupa peta analog maka peta tersebut harus dirubah menjadi peta digital menggunakan tools *digitizer*. selain proses digitasi dapat juga digunakan proses overlay dengan proses scanning pada peta analog

2. Manipulasi Data
   Tipe data yang diperlukan SIG mungkin perlu dimanipulasi agas sesuai dengan sistem yang dipergunakan, oleh karena itu Sistem informasi geografis dapat mengedit baik data spasial ataupun non-spasial.

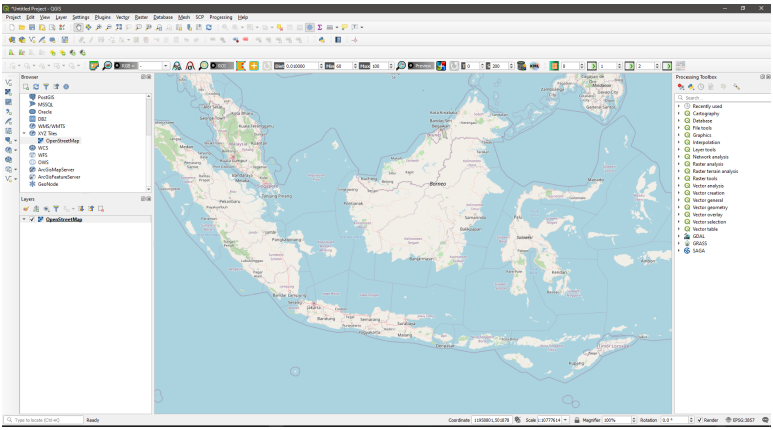
3. Manajemen Data
   Setelah data spasial atau non-spasial diiputkan maka data-data tersebut dapat diolah oleh user untuk dikembangkan lebih lanjut. Penggunaan data non-spasial meliputi penggunaan DBMS untuk menyimpan data skala besar.

4. *Query* dan Analisis
   Query dalam sistem informasi geografis adalah proses analisis yang dilakukan secara tabular. Terdapat dua jenis analisis dalam sistem inforasi geografis yaitu Analisis *proximity* dan Ana-

lisis *Overlay*. Analisis *proximity* merupakan analisis geografi yang berdasar pada jarak antara layer dimana digunakan proses *buffering* dalam membangun lapisan pendukung disekitar layer dalam jarak tertentu untuk menentukan dekatnya hubungan antar layer. Analisis *Overlay* merupaka proses penyatuan data dari lapisan layer yang berbeda, secara sederhana proses ini disebut sebagai operasi visual yang membutuhkan lebih dari satu layer untuk digabungkan.

5. Visualisasi
   Hasil akhir dari proses-proses yang telah dilakukan dapat terwujudkan dengan adanya peta atau grafik karena peta efektif untuk memberikan informasi geografis

Data khusus yang diolah oleh sistem ini adalah data spasial yakni data yang berorientasi pada geografis. Gambar 2.5 adalah contoh Sistem Informasi Geografis yaitu aplikasi QGIS



Gambar 2.5: Sistem Informasi Geografis QGIS

2.2.2 *Application Programming Interface* (API)

Application Programming Interface (API) adalah sekumpulan definisi subrutin, protokol komunikasi, dan alat untuk membangun perangkat lunak. API memungkinan *developer* untuk mengintegrasikan dua bagian atau lebih dari suatu aplikasi yang berbeda secara

bersamaan. Tujuan penggunaan API adalah untuk mempercepat proses pengembangan dengan menyediakan *function* secara terpisah sehingga *developer* tidak perlu membuat fitur yang sama. API yang bekerja pada tingkat sistem operasi membantu aplikasi berkomunikasi dengan *layer* dasar dan satu sama lain mengikuti serangkaian protokol atau spesifikasi. API terdiri atas beberapa macam yaitu:

1. *Libraries* dan *Frameworks*
2. Sistem Operasi
3. *Remote* APIs
4. Web APIs

   Contoh dari API yaitu Rest API dan *Google Cloud* API.

2.2.3 Data Sistem Informasi Geografis

Terdapat 2 proses dalam sistem informasi geografis dalam lingkup data [4] yaitu input dan manajemen data. Seperti yang disebutkan diatas dalam sistem informasi geografis data terdapat data spasial dan data non-spasial. Data-data tersebut secara umum terbagi menjadi 4 jenis data dalam sistem informasi geografis, yaitu :

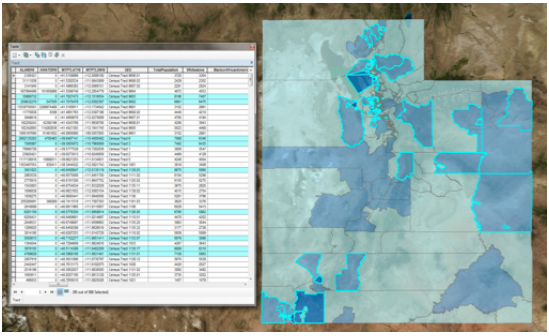
1. Data Spasial

   Data spasial adalah data yang memiliki referensi ruang kebumian (georeference) di mana berbagai data atribut terletak dalam berbagai unit spasial seperti garis bujur *(longitude)* dan garis lintang *latitude*. Sekarang ini data spasial menjadi media penting untuk perencanaan pembangunan dan pengelolaan sumber daya alam yang berkelanjutan pada cakupan wilayah continental, nasional, regional maupun lokal. Pemanfaatan data spasial semakin meningkat setelah adanya teknologi pemetaan digital dan pemanfaatannya pada Sistem Informasi Geografis (SIG)

2. Data Atribut

   Data Atribut merupakan data yang menjabarkan aspek dari suatu geografis dalam bentuk deskripsi atau penjelasan yang terperinci seperti Gambar 2.6. Data ini terangkai dalam bentuk kata-kata, angka, serta tabel, contohnya adalah data luas wilayah, data kepadatan penduduk dan sebagainya. Data

Atribut dapat digabungkan dalam satu tabel sehingga dapat menghasilkan suatu analisis tertentu terhadap data tersebut. Data atribut juga bisa disebut sebagai tabular data dimana data ini nantinya dapat dibentuk menjadi sebuah *geodatabase* .



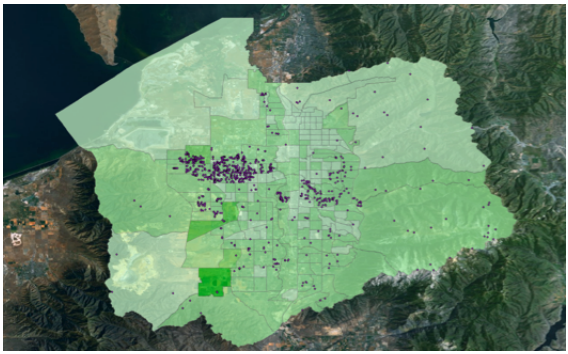
Gambar 2.6: Data Atribut berupa Tabel

3. Data Vektor

   Data Vektor berupa data yang merepresentasikan fitur lokasi dan bentuk geografis melalui titik, garis, dan poligon didalam sistem informasi geografis. Fitur vektor bentuknya direpresentasikan menggunakan sistem geometri dimana geometrinya terbentuk dari titik yang saling terhubung satu sama lain, fitur ini disebut sebgai *vertex*. *Vertex* ini mendekripsikan posisinya menggunakan sumbu x, y, dan z. Kegunaan data vektor ini untuk menganalisa ketepatan posisi suatu wilayah ataupun mendefinisikan hubungan spasian dari beberapa fitur. Gambar2.7 merupakan contoh dari data vektor.

4. Data Raster

   Data Raster atau sering juga disebut dengan *Grid Cell* merupakan data yang dihasilkan dari sistem penginderaan jauh atau citra digital seperti gambar 2.8. Pada data raster objek geografis direpresentasikan sebagai struktur *Grid Cell* yang disebut dengan pixel (*picture element*)



Gambar 2.7: Data Vektor berupa *polygon*

Gambar 2.8: Data Raster berupa *Image*

2.2.4 *K-Nearest Neighbors* (KNN)

Algoritma *K-Nearest Neighbors* (KNN) adalah sebuah algoritma sederhana untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data serta menyimpa semua kasus yang tersedia dan mengklasifikasikan kasus baru berdasarkan ukuran kesamaan jarak. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah kasus diklasifikasikan berdasarkan mayoritas dari

tetangganya dimana kasus tersebut akan ditetapkan *K-Nearest Neighbors*. Semakin besar nilai K maka semakin akurat hasil klasifikasi terhadap suatu kasus tersebut. Terdapat 3 fungsi persamaan jarak pada KNN yaitu persamaan Euclidean, persamaan Manhattan, dan persamaan Minkowski [5]. Persamaan Euclidean ditunjukkan pada Persamaan 2.1, Persamaan Manhattan ditunjukkan pada Persamaan 2.2 dan Persamaan Minkowski ditunjukkan pada Persamaan 2.3.

$$Euclidean = \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2} \tag{2.1}$$

Keterangan:
1. Xi = Koordinat X titik i
2. Yi = Koordinat Y titik i
3. n = Banyak data

$$Manhattan = \sum_{i=1}^{n}|Xi - Yi| \tag{2.2}$$

Keterangan:
1. Xi = Koordinat X titik i
2. Yi = Koordinat Y titik i
3. n = Banyak data

$$Minkowski = (\sum_{i=1}^{n}(|Xi - Yi|)^q)^{1/q}, q \geq 1 \tag{2.3}$$

Keterangan :
1. Xi = Koordinat X titik i
2. Yi = Koordinat Y titik i
3. n = Banyak data
4. q = Jarak Chebysev

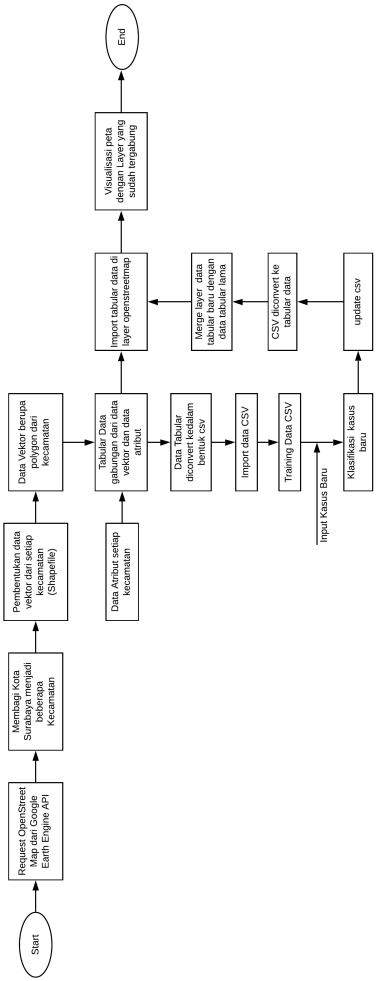
BAB 3
DESAIN DAN IMPLEMENTASI SISTEM

Penelitian ini dilaksanakan sesuai dengan desain sistem berikut dengan implementasinya. Desain sistem merupakan konsep dari pembuatan dan perancangan perangkat lunak dan kemudian diwujudkan dalam bentuk blok-blok alur yang harus dikerjakan.

3.1 Desain Sistem

Tugas akhir ini merupakan salah satu bentuk dari pemanfaatan beberapa Sistem Informasi Geografis (SIG) seperti ArcGIS dan QGIS, *Application Programming Interface* (API) yakni *Google Earth Engine* API serta algoritma *K-Nearest Neighbors* (KNN) untuk menganalisa dan mengklasifikasikan data-data SIG baik itu berupa data vektor, data raster, data spasial, atau data atribut. Sistem ini menghasilkan visualisasi dan ilustrasi berupa peta digital Kota Surabaya yang terbagi menjadi banyak kecamatan-kecamatan.

Berdasarkan diagram Gambar 3.1, dalam sistem QGIS akan merequest layer basemap berupa openstreetmap kepada *google earth engine* API. Untuk mengakses request tersebut perlu dilakukan instalasi *plugin Google Earth Engine Python* didalam *environtment* QGIS. Basemap ini yang akan mejadi layer dasar daripada peta digital nantinya. Peta akan difokuskan pada Kota Surabaya dimana Kota Surabaya memiliki rincian data spasial seperti garis bujur 112.768845 dan garis lintang -7.250445 serta tingkat skala yang digunakan pada peta QGIS adalah 1:106440. Setelah tersedia peta Kota Surabaya maka akan dibagi menjadi kecamatan. Kota Surabaya sendiri memiliki 31 kecamatan.



Gambar 3.1: Blok Diagram Sistem

Berdasarkan diagram Gambar 3.1, dalam sistem QGIS akan merequest layer basemap berupa openstreetmap kepada *google earth engine* API. Untuk mengakses request tersebut perlu dilakukan instalasi *plugin Google Earth Engine Python* didalam *environtment* QGIS. Basemap ini yang akan mejadi layer dasar daripada peta digital nantinya. Peta akan difokuskan pada Kota Surabaya dimana Kota Surabaya memiliki rincian data spasial seperti garis bujur 112.768845 dan garis lintang -7.250445 serta tingkat skala yang digunakan pada peta QGIS adalah 1:106440. Setelah tersedia peta Kota Surabaya maka akan dibagi menjadi kecamatan. Kota Surabaya sendiri memiliki 31 kecamatan.

Setiap Kecamatan dibuat dari kumpulan titik-titik koordinat yang kemudian digabungkan menjadi satu fitur *polygon*. Fitur ini merupakan data vektor yang merepresentasikan bentuk serta perbatasan dari setiap kecamatan yang ada di Kota Surabaya. Garis perbatasan akan terbentuk jika adanya titik atau*vertices* yang identikal atau sama saling bertumpukan dalam satu layer. Layer ini memiliki format file ESRI shapefile. Format Shapefile adalah format umum dari sebuah layer dalam sebuah sistem informasi geografis.Format ini membantu dalah hal penyimpanan data-data titik vektor yang membentuk kecamatan

Setiap Kecamatan juga memiliki data atribut yang berhubungan dengan kecamatan tersebut. Data atribut tersebut berupa *string* dan angka. untuk proses penggabungan data tersebut maka perlu dilakukan sebuah *merger* antara data vektor dengan data atribut. Proses ini menghasilkan data tabular. Data tabular berisi titik-titik *polygon* yang membentuk wilayah kecamatan tersebut dan data atribut yang ditambahkan kedalamnya. Data-data tersebut disimpan dalam bentuk tabel tapi masih mempertahankan bentuk *polygon* yang terdiri dari titik-titik yang saling terhubung. Data Atribut juga memiliki format ESRI Shapefile sehingga dapat digabungkan menjadi satu layer dengan proses *merge*.

Setelah Data tabular antar kecamatan terbentuk dalam format shapefile maka akan dilakukan konversi dari data tabular menjadi data *Comma Separated Value* (CSV). Proses konversi data dilakukan di ArcGIS dengan menggunakan *tools* manajemen data. Data CSV tersebut akan menjadi data input untuk melakukan proses *tra-*

ining data pada algoritma *K-Nearest Neighbors* (KNN). Data tabular menjadi dataset awal untuk menentukan prediksi dan klasifikasi pada kasus baru nantinya. Setelah dilakukan training maka data hasil prediksi akan diklasifikasikan dan dirubah menjadi CSV.

Setelah hasil prediksi dikonversi ke CSV data tersebut kemudian dirubah menjadi data tabular kembali. Setelah data terubah data tersebut kemudian akan di*merge* dengan data tabular yang sebelumnya untuk menghasilkan data tabular yang baru dimana data yang baru berisi kasus terbaru yang sudah terklasifikasikan. data yang tergabung akan divisualisasikan dalam bentuk peta digital.

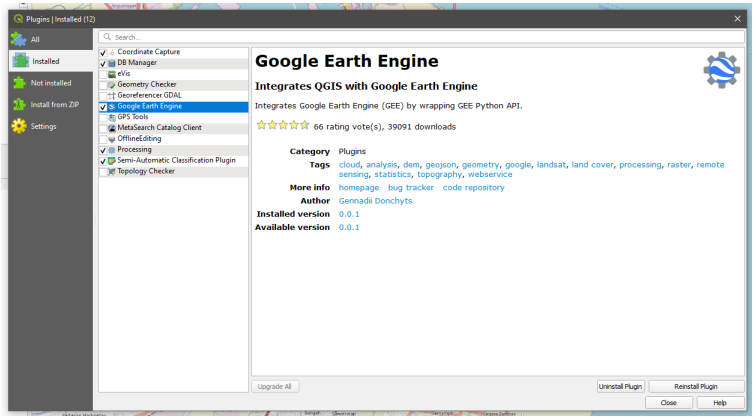
3.2 Alur Implementasi Sistem

Pengerjaan tugas akhir ini dibagi menjadi beberapa tahapan berdasarkan metodologi penelitian, yaitu :

1. Request *Google Earth Engine* API dengan QGIS
2. Pembagian Kecamatan di Kota Surabaya
3. Pembentukan Data Vektor dan Data Atribut Kecamatan
4. Penggabungan Data Vektor dengan Data Atribut Menjadi Data Tabular
5. Konversi Data Tabular menjadi Data CSV pada QGIS
6. Training Data untuk Klasifikasi dengan Algoritma KNN
7. Visualisasi Peta Hasil Klasifikasi

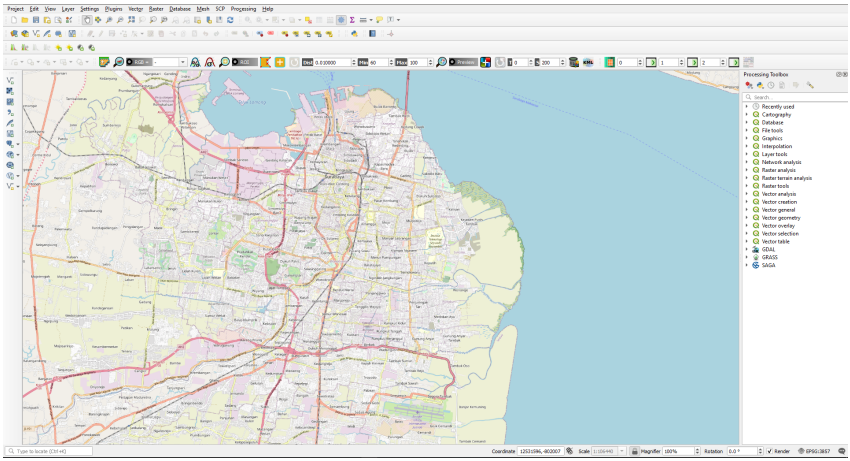
3.3 Request *Google Earth Engine* API dengan QGIS

Pada Penelitian tugas akhir ini perangkat lunak sistem informasi geografis yang digunakan adalah QGIS. QGIS merupakan sistem informasi geografis yang memiliki *environtment* python untuk pengembangannya. QGIS juga memiliki *plugin internal* dan dapat melakukan instalasi *plugin* yang berhubungan dengan *environment-nya* yaitu *Google Earth Engine* python API seperti Gambar 3.2.



Gambar 3.2: QGIS dengan *integrated* GEE pyhton API

Integrated Google Earth Engine API mempermudah melakukan request peta baik itu berupa peta berjenis *open street map* ataupun peta digital berupa citra satelit dari google. Setelah request akan muncul peta *openstreetmap* dalam layer. Peta akan difokuskan pada *longitude* dan *latitude* dari Kota Surabaya dan skala peta akan dikunci pada ukuran 1:106440 seperti Gambar 3.3



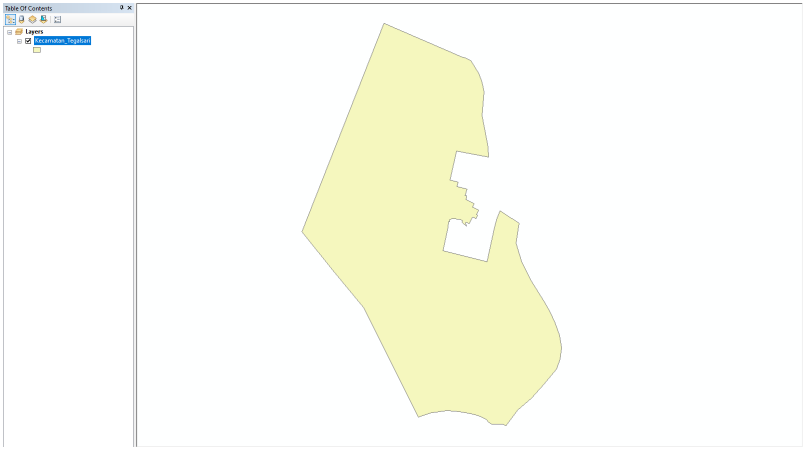
Gambar 3.3: Kota Surabaya

3.4 Pembagian Kecamatan di Kota Surabaya

Pada penelitian tugas akhir ini akan dilakukan proses pembagian wilayah Kota Surabaya menjadi beberapa kecamatan. Surabaya memiliki 31 kecamatan dalam 5 wilayah Surabaya.5 wilayah Kota Surabaya terdiri atas Surabaya Pusat, Surabaya Utara, Surabaya Timur, Surabaya selatan, dan Surabaya Barat. Surabaya pusat terdiri atas kecamatan Tegalsari, kecamatan Simokerto, Kecamatan Genteng, dan kecamatan Bubutan. Surabaya Utara terdiri atas kecamatan Bulak, kecamatan Kenjeran, kecamatan Semampir, kecamatan Pabean cantian, dan kecamatan Krembangan. Surabaya Timur terdiri atas kecamatan Gubeng, kecamatan Gunung Anyar, kecamatan Sukolilo, kecamatan Tambaksari, kecamatan Mulyorejo, kecamatan Rungkut, dan kecamatan Tenggilis Mejoyo. Surabaya Barat terdiri atas kecamatan Benowo, kecamatan Pakal, kecamatan Asemrowo, kecamatan Sukomanunggal, kecamatan Tandes, kecamatan Lakarsantri, dan kecamatan Sambikerep. Surabaya Selatan terdiri atas kecamatan Dukuh Pakis, kecamatan Gayungan, kecamatan Jambangan, kecamatan Karang Pilang, kecamatan Sawahan, kecamatan Wiyung, kecamatan Wonocolo, dan kecamatan Wonokromo. Untuk pembentukan data secara representatif akan diambil beberapa kecamatan dari setiap wilayah di Kota Surabaya.

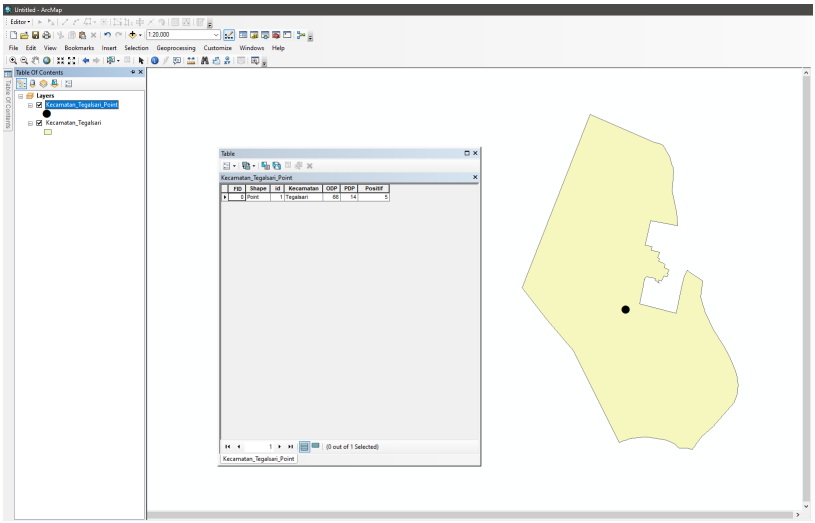
3.5 Pembentukan Data Vektor dan Data Atribut Kecamatan

Data vektor dapat terdiri atas titik atau *vertices*, garis, *polygon*, dan beberapa koleksi fitur. Pembentukan data vektor ini akan dilakukan menggunakan sistem informasi geografis ArcGIS. Dalam pembentukan kecamatan akan dimulai dari penempatan titik-titik tertentu pada koordinat tertentu lali titik-titik tersebut dihubungkan sehingga membentuk suatu garis. Garis tersebut dihubungkan dari titik awal sampai kembali lagi ke menutup ke titik awal sehingga terbentuklah *polygon* yang tertutup merepresentasikan bentuk wilayah yang identikal dengan kecamatan tersebut seperti gambar 3.4. Setelah *polygon* terbentuk pada layer kemudian layer disimpan dalam format ESRI Shapefile sesuai Gambar 3.5



Gambar 3.4: Shapefile dari Kecamatan Tegalsari

Untuk pembentukan data atribut pada kecamatan dibentuk melalui sebuah titik didalam daerah data *polygon* yang terbentuk diproses pembentukan kecamatan. Untuk pembentukan data atribut menggunakan *tools Attribute Table Panel*. Atribut yang akan dimasukkan kedalam titik tersebut yaitu nama kecamatan, Orang Dalam Pengawasan (ODP), Pasien Dalam Pengawasan (PDP), dan Pasien Positif dari Covid-19 sesuai data yang tersedia.

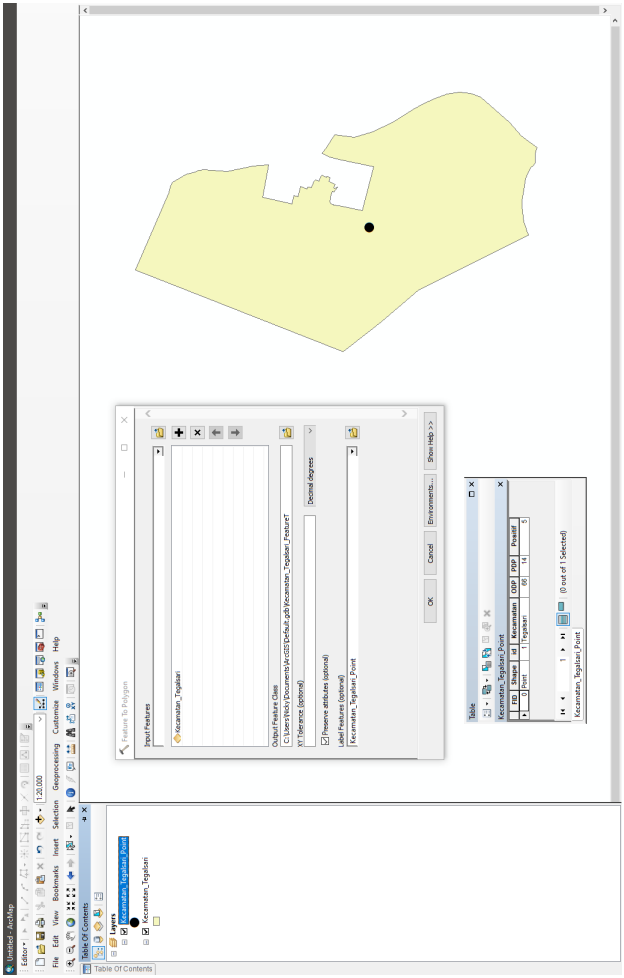


Gambar 3.5: Titik dari Kecamatan Tegalsari yang Berisi Data Atribut

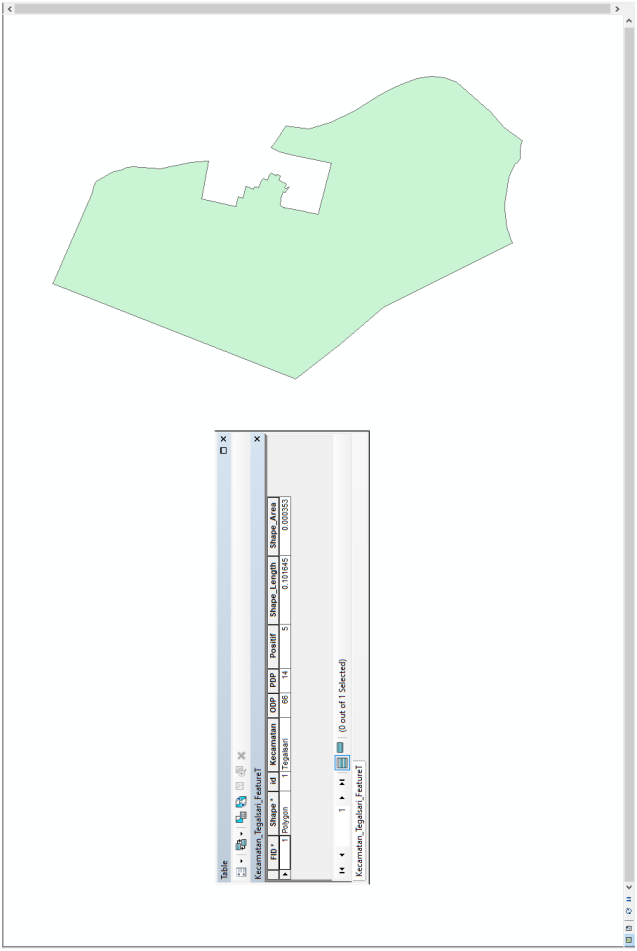
3.6 Penggabungan Data Vektor dengan Data Atribut menjadi Data Tabular

Untuk menghasilkan data tabular maka diperlukan proses untuk merubah dan menggabungkan semua fitur dari data vektor yang berbentuk poligon dan data atribut yang berbentuk titik menjadi satu poligon dengan mempertahankan atribut dari kedua data seperti Gambar 3.6. Proses ini menghasilkan fitur poligon yang berbentuk kecamatan tersebut namun menghilangkan *visual* titik dari data atribut, sedangkan pada sisi atribut kedua atribut yang dihasilkan oleh data vektor dan data atribut tergabung menjadi satu

seperti gambar 3.7

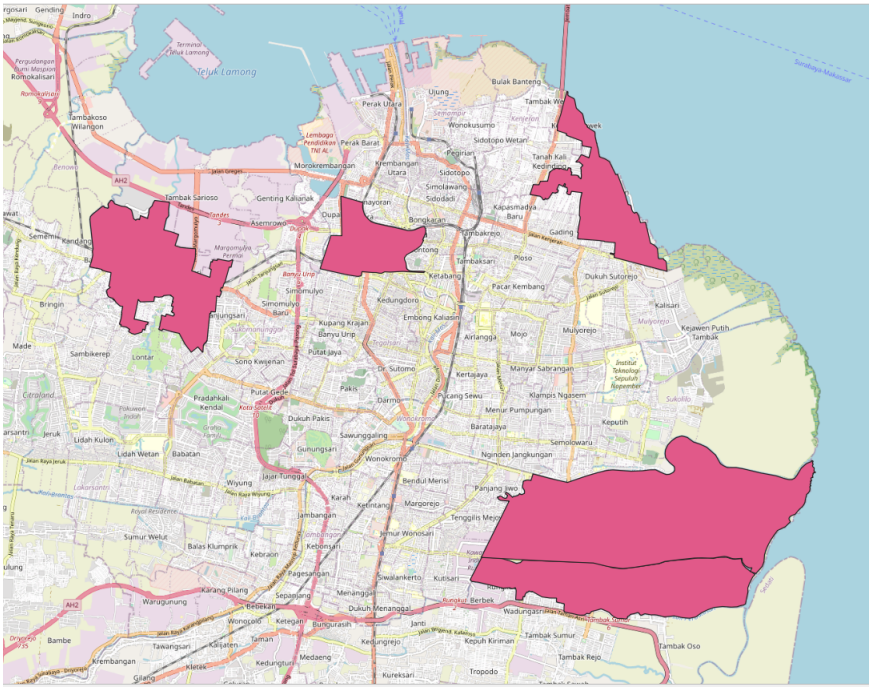


Gambar 3.6: Proses Penggabungan Data Atribut dengan Data vektor



Gambar 3.7: Shapefile Gabungan dari Kecamatan Tegalsari

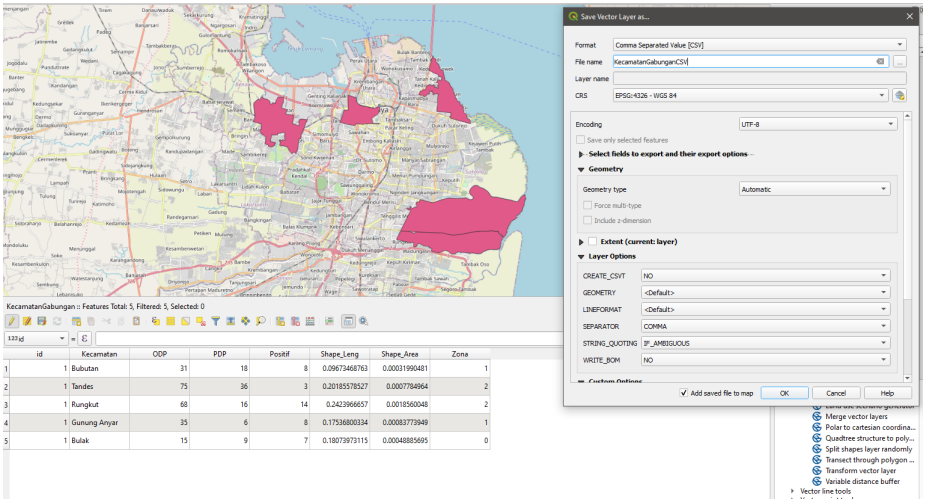
Dua proses diatas pembentukan dan penggabungan dilakukan berulang kali untuk menghasilkan beberapa kecamatan. Kecamatan yang berhasil tergabungkan seperti gambar 3.8 adalah kecamatan Bubutan, kecamatan Tandes, kecamatan Bulak, Kecamatan Rungkut, dan Kecamatan Gununganyar. Data gabungan berupa shapefile ini kemudian divisualisasikan didalam QGIS serta dilakukan penambahan atribut zona pada setiap kecamatan setelah data-data atribut kecamatan tergabungkan.



Gambar 3.8: Shapefile Gabungan dari beberapa Kecamatan

3.7 Konversi Data Tabular menjadi Data CSV pada QGIS

Setelah data gabungan berhasil terbentuk dan tervisualisasi maka akan dilakukan proses konversi dari bentuk data ESRI Shapefile menjadi data *Comma Separated value* (CSV) pada Gambar 3.9. Data csv (Tabel 3.1) ini nantinya akan menjadi dataset awal atau data inputan untuk melakukan *training data* dengan algoritma *K-Nearest Neighbors*.



Gambar 3.9: Proses Konversi Data

Tabel 3.1: Struktur Tabel CSV

1	id	Kecamata	ODP	PDP	Positif	Shape_Le	Shape_Ar	Zona
2	1	Bubutan	31	18	8	0.096735	0.00032	1
3	1	Bulak	15	9	7	0.18074	0.000489	0
4	1	Gunung A	35	6	8	0.175368	0.000838	1
5	1	Rungkut	68	16	14	0.242397	0.001856	2
6	1	Tandes	75	36	3	0.201856	0.000778	2

3.8 Training Data untuk Klasifikasi dengan Algoritma KNN

K-Nearest Neighbors adalah bentuk paling sederhana dari sebuah proses *learning* dan klasifikasi. Dataset yang akan di*training* data *Comma Separated Value* yang telah dikonversikan dari data Tabular. Untuk algoritma *K-Nearest Neighbors* menggunakan library sklearn. Proses-proses *learning* terdiri atas :

1. Import *Library* dan Komponen-Komponen Python
2. Import Dataset CSV
3. *split dataset* menjadi *Test* dan *Train*
4. Melakukan Proses *Scaling* Dataset
5. Melakukan Training Data dan Klasifikasi

3.8.1 Import *library* dan Komponen-Komponen Python

Komponen-komponen utama *environtment* python pada gambar 3.10 yang di *import* adalah *pandas*, *numpy*, dan *matplotlib*. *Library* utama untuk algoritma *K-Nearest Neighbors* adalah sklearn. Sklearn berisi perintah-perintah untuk melakukan *training*, memprediksi, dan pengklasifikasian.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
from sklearn.metrics import confusion_matrix
from sklearn.metrics import f1_score
from sklearn.metrics import accuracy_score
```

Gambar 3.10: Proses Import Library

3.8.2 Import Dataset CSV

Dataset yang di*import* adalah data kecamatan gabungan pada Gambar 3.11 yang telah dikonversikan kedalam bentuk format *Comma Separated Value* (CSV) menggunakan perintah dari library pandas.

```
: dataset = pd.read_csv('D:\KULIAH\TA BOS\Sistem Informasi Geografis\
  print(dataset.head())

     id     Kecamatan  ODP  PDP  Positif  Zona
0   1       Bubutan   31   18        8     1
1   1         Bulak   15    9        7     0
2   1  Gunung Anyar   35    6        8     1
3   1       Rungkut   68   16       14     2
4   1        Tandes   75   36        3     2
```

Gambar 3.11: Proses Import Data CSV

3.8.3 *Split dataset* menjadi *Test* dan *Train*

Dataset akan dibagi menjadi dua bagian Test dan Train sesuai Gambar 3.12. Data dibagi menjadi 2 variabel dimana masing-masing variabel akan memiliki Test dan Train. Data variabel X berisikan kolom ODP, PDP, Positif, dan zona, sedangkan Variabel Y berisi kolom zona.Untuk variabel X perintah pada library panda akan mengeksekusi dan mengecek semua baris yang ada pada dataset tetapi dibatasi hanya dari kolom dua sampai kolom lima dimana berisi ODP, PDP, Positif, dan zona. Untuk variabel Y akan dicek semua baris dataset namun hanya dibatasi pada kolom lima saja.Ukuran data test ditentukan sebesar duapuluh persen terhadap dataset.

```
X = dataset.iloc[:, 2:5]
y = dataset.iloc[:, 5]
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, test_size=0.2)
print(X)
print(X_train)
print(X_test)
print(len(y_train))
print(y_train)
print(y_test)
```

Gambar 3.12: Proses *Split* Data

3.8.4 Melakukan Proses *Scaling* Dataset

Gambar 3.13 merupakan proses*scaling* atau yang biasa disebut *Feature Scaling*. Proses merupakan proses perubahan data integer untuk membuat sebuah standar terhadap dataset tersebut, angka-angka didalam dataset akan menjadi angka diantara -1 sampai 1. Proses ini merupakan tahap *pre-processing* dalam sebuah *learning*.

```
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.transform(X_test)
print (X_train)
print (X_test)
```

```
[[-1.28811922  0.06045142 -1.41699309]
 [ 1.828691   -0.99470071  0.22193868]
 [-0.66475718  0.76388618  0.49509397]
 [-1.2491591   1.02767421 -1.1438378 ]
 [ 1.28324921  0.85181552 -0.8706825 ]
```

Gambar 3.13: Proses *Scaling* Data

3.8.5 Melakukan *Training* Data

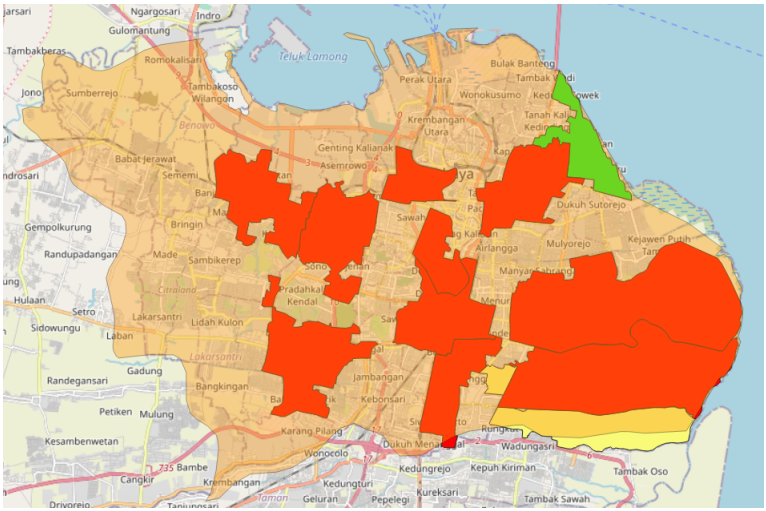
Training data dilakukan berulang kali dengan nilai K=1 sampai dengan K=20 untuk nantinya dilihat akurasi dari setiap nilai K sesuai Gambar 3.14.

```
k_range = range(1,21)
scores = {}
scores_list = []
for k in k_range :
    knn = KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train, y_train)
    y_pred = knn.predict(X_test)
    scores[k]=metrics.accuracy_score(y_test,y_pred)
    scores_list.append(metrics.accuracy_score(y_test, y_pred))
```

Gambar 3.14: Proses *Training* Data

3.9 Visualisasi Peta Hasil Klasifikasi

Data Hasil *learning* dan klasifikasi nantinya akan disimpan dalam bentuk CSV baru lalu dikonversi kembali menjadi tabular data dengan mempertahankan atribut yang tidak masuk dalam proses *learning* dan klasifikasi. Data yang telah diprediksi akan dimasukkan kedalam peta digital pada QGIS seperti Gambar 3.15.



Gambar 3.15: Proses visualisasi Peta Digitial

3.10 Proses Deployment Aplikasi

Untuk proses deployment perlu adanya penyamaan *environment* antara algoritma *K-Nearest Neighbors* dengan *environment* dari Google Earth API dalam bahasa python sehingga dapat berfungsi dengan baik. Untuk proses deployment diperlukan juga dimasukkannya sistem autentikasi dari google agar aplikasi dapat diakses melalui web.Untuk percobaan daripada sistem autentikasi kita perlu melakukan *subscription* pada API Google dan akan didapatkan token berupa kode autentikasi unik untuk *user*.

BAB 4
PENGUJIAN DAN ANALISA

Pada bab ini dilakukan pengujian serta analisa terhadap desain sistem yang telah dirancang. Pengujian dilakukan dengan maksud dan tujuan untuk mengetahui tingkat efektivitas serta efisiensi desain sistem sehingga dapat menarik kesimpulan terhadap penelitian tugas akhir ini. Secara garis besar, pengujian yang akan dilakukan pada bagian ini adalah sebagai berikut :

1. Pengujian Akurasi Sistem Klasifikasi - meliputi tingkat akurasi klasifikasi berdasarkan nilai K.
2. Pengujian Klasifikasi Sistem Algoritma K-NN - meliputi klasifikasi klasifikasi terhadap kasus baru yang belum terklasifikasikan.
3. Pengujian Hasil Klasifkasi - meliputi perubahan zona pada peta digital setelah proses Klasifkasi

4.1 Pengujian Sistem Klasifikasi

Sistem klasifikasi dala tugas akhir ini akan diuji dalam hal akurasi dan tingkat efektifitas klasifikasi dan klasifikasi terhadap suatu skema pada data yang ada.

4.1.1 Pengujian Akurasi Sistem Klasifikasi

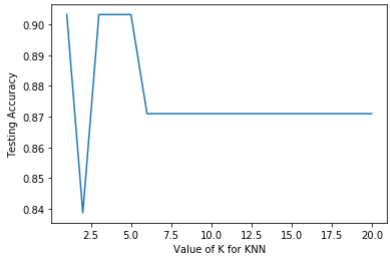
Akurasi dari sistem algoritma *K-Nearest Neighbors* dapat berubah sesuai dengan nilai K yang ditentukan. Perlunya dilakukan iterasi dalam proses *training* data dapat menemukan nilai K yang paling optimum. Proses *split* data juga membantu meningkatkan kualitas akurasi dari dataset dikarenakan tanpa proses ini maka akan terjadi fenomena *overfit* dimana ada beberapa data yang pengklasifikasiannya tidak sesuai. Secara umum semakin besar nilai K tanpa dilakukan proses pemisahan data semakin buruk, dan nilai k yg kecil smakin baik hasil akurasinya dikarenakan algoritma melakukan klasifikasi pada dataset yang sama tanpa ada data lain diluar dataset yang ditraining. Dengan pemisahan dataset menjadi test dan train pada proses 3.8.3 maka nilai K memiliki nilai yang bervariasi dan setiap nilai K memiliki nilai akurasi tersendiri dan bisa dikatakan stabil jika nilai K sudah memiliki nilai akurasi yang sama.

Pengujian ini ditujukan untuk menentukan mana nilai K yang memiliki akurasi paling tinggi dalam klasifikasi setelah melakukan proses *training* data. Sesuai Gambar 4.1 dan plot pada Gambar 4.2 dapat diketahui bahwa ketika K bernilai 2 memiliki akurasi klasifikasi sebesar 83 persen, K bernilai 1, 3, 4, 5 memiliki akurasi klasifikasi sebesar 90 persen dan K bernilai 6 sampai 20 memiliki akurasi sebesar 87 persen. K dengan nilai lebih dari 6 memiliki nilai akurasi yang lebih stabil daripada nilai K yang lainnya.

```
In [6]: k_range = range(1,21)
        scores = {}
        scores_list = []
        for k in k_range :
            knn = KNeighborsClassifier(n_neighbors=k)
            knn.fit(X_train, y_train)
            y_pred = knn.predict(X_test)
            scores[k]=metrics.accuracy_score(y_test,y_pred)
            scores_list.append(metrics.accuracy_score(y_test, y_pred))

In [7]: scores

Out[7]: {1: 0.9032258064516129,
         2: 0.8387096774193549,
         3: 0.9032258064516129,
         4: 0.9032258064516129,
         5: 0.9032258064516129,
         6: 0.8709677419354839,
         7: 0.8709677419354839,
         8: 0.8709677419354839,
         9: 0.8709677419354839,
         10: 0.8709677419354839,
         11: 0.8709677419354839,
         12: 0.8709677419354839,
         13: 0.8709677419354839,
         14: 0.8709677419354839,
         15: 0.8709677419354839,
         16: 0.8709677419354839,
         17: 0.8709677419354839,
         18: 0.8709677419354839,
         19: 0.8709677419354839,
         20: 0.8709677419354839}
```

Gambar 4.1: Proses *Test* Akurasi Nilai K

```
In [8]: plt.plot(k_range, scores_list)
        plt.xlabel('Value of K for KNN')
        plt.ylabel('Testing Accuracy')

Out[8]: Text(0, 0.5, 'Testing Accuracy')
```



Gambar 4.2: Plot Hasil Akurasi Nilai K terhadap Nilai K

4.1.2 Pengujian Klasifikasi Pada Sistem Algoritma K-NN

Klasifkasi pada suatu data baru dapat dilakukan jika sudah sudah melalui proses *training* dataset diawal dan sudah didapatkan beberapa nilai K dengan akurasi predikisi yang bervariasi. Sebelum dilakukan Klasifkasi Jumlah K harus ditentukan terlebih dahulu. Proses pengujian ini akan menggunakan 3 nilai K yang berbeda untuk menguji Klasifikasi yang paling akurat sesuai dengan skema didunia nyata. Terdapat 3 hasil pada proses pengujian akurasi yaitu 83 persen dengan K bernilai 2, 87 persen ketika K bernilai 6 sampai 20, dan 90 persen dengan K bernilai 1, 3, 4, 5. Metode untuk penghitungan jarak antara titik menggunakan metode Minkowski.

```
knn = KNeighborsClassifier(n_neighbors=3, metric='minkowski')
knn.fit(X,y)

KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=3, p=2,
                     weights='uniform')
```

Gambar 4.3: Proses Klasifikasi dengan K=3 dan metode Minkowski [3]

```
knn = KNeighborsClassifier(n_neighbors=7, metric='minkowski')
knn.fit(X,y)
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=7, p=2,
                     weights='uniform')
```

Gambar 4.4: Proses Klasifikasi dengan K=7 dan metode Minkowski [3]

```
knn = KNeighborsClassifier(n_neighbors=15, metric='minkowski')
knn.fit(X,y)
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=15, p=2,
                     weights='uniform')
```

Gambar 4.5: Proses Klasifikasi dengan K=15 dan metode Minkowski [3]

Variabel yang akan diklasifkasi adalah Zona dimana terdapat 3 zona klasifikasi yaitu Zona Hijau, Zona Kuning, dan Zona Merah dengan inputan Orang Dalam Pengawasan (ODP), Pasien Dalam Pengawasan (PDP), dan Pasien Positif COVID-19.

Pengujian ini dilakukan untuk membandingkan hasil klasifikasi zona dengan nilai K=3 dan K=7 tanpa melalui proses normalisasi data (Gambar 4.3 dan Gambar 4.4) dengan K=15 melalui proses normalisasi data (Gambar 4.5) terhadap hasil yang diharapkan dengan beberapa variasi skema pada input data ODP, PDP dan Positif.

Skema n+10 pada setiap input						
Input			Hasil yang Diharapkan	Hasil KLasifikasi(K=3) Tanpa Normalisasi	Hasil KLasifikasi(K=7) Tanpa Normalisasi	Hasil Klasifikasi (K=15) Dengan Normalisasi
ODP	PDP	Positif				
0	0	0	Green Zone	Green Zone	Yellow Zone	Green Zone
10	10	10	Green Zone	Green Zone	Yellow Zone	Green Zone
20	20	20	Green Zone	Green Zone	Yellow Zone	Green Zone
30	30	30	Yellow Zone	Green Zone	Yellow Zone	Yellow Zone
40	40	40	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
50	50	50	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
60	60	60	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
70	70	70	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
80	80	80	Yellow Zone	Yellow Zone	Red Zone	Yellow Zone
90	90	90	Red Zone	Yellow Zone	Red Zone	Yellow Zone
100	100	100	Red Zone	Red Zone	Red Zone	Red Zone
110	110	110	Red Zone	Red Zone	Red Zone	Red Zone
120	120	120	Red Zone	Red Zone	Red Zone	Red Zone

Gambar 4.6: Hasil Klasifkasi Skema 1

Pada skema 1 semua data input ditambah 10 untuk setiap iterasinya (n+10) dan menghasilkan Klasifkasi sesuai Gambar 4.6. Percobaan awal adalah ketika data input ODP, PDP, dan Positif bernilai 0 nilai K=3 dapat mengklasifkasi bahwa zona tersebut zona hijau sedangkan ketika nilai K=7 mengklasifikasi zona tersebut zona kuning.Ketika K=7 hasil klasifikasi banyak berupa zona kuning dan mencapai zona merah ketika nilai input berada pada angka 80 orang, sedangkan K=3 terjadi perubahan klasifikasi dari zona hijau ke zona kuning pada saat nilai input berada pada angka 40 dan berubah dari zona kuning ke zona merah pada saat nilai input berada pada angka 100. Hasil klasifikasi ketika k=7 menunjukkan hasil yang lebih sensitif terhadap zona merah pada saat adanya perubahan data inputan, sedangkan ketika k=3 hasil yang ditunjukkan memiliki tingkat sensitifitas yang sedikit berbeda dengan dapat mengklasifikasi 3 zona. Nilai K=3 mengahasilkan klasifikasi yang hampir sama dengan hasil yang diharapkan. Nilai K=15 memiliki hasil yang sedikit berbeda namun dapat mengklasifikasikan 3 zona dengan baik dan memiliki sensitifitas terhadap perubahan data yang relatif sama dengan K=7.

Skema n+20 pada input ODP						
Input			Hasil yang Diharapkan	Hasil KLasifikasi(K=3) Tanpa Normalisasi	Hasil KLasifikasi(K=7) Tanpa Normalisasi	Hasil Klasifikasi (K=15) Dengan Normalisasi
ODP	PDP	Positif				
0	0	0	Green Zone	Green Zone	Yellow Zone	Green Zone
20	10	10	Green Zone	Green Zone	Yellow Zone	Green Zone
40	20	20	Green Zone	Green Zone	Yellow Zone	Green Zone
60	30	30	Green Zone	Yellow Zone	Yellow Zone	Yellow Zone
80	40	40	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
100	50	50	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
120	60	60	Red Zone	Red Zone	Yellow Zone	Yellow Zone
140	70	70	Red Zone	Red Zone	Red Zone	Yellow Zone
160	80	80	Red Zone	Red Zone	Red Zone	Yellow Zone
180	90	90	Red Zone	Red Zone	Red Zone	Red Zone
200	100	100	Red Zone	Red Zone	Red Zone	Red Zone
220	110	110	Red Zone	Red Zone	Red Zone	Red Zone
240	120	120	Red Zone	Red Zone	Red Zone	Red Zone

Gambar 4.7: Hasil klasifikasi Skema 2

Pada skema 2 semua data input ditambah 10 untuk setiap iterasinya (n+10) namun data input ODP ditambahkan 20 (n+20) untuk untuk setiap iterasinya dan menghasilkan klasifikasi sesuai Gambar 4.7. Percobaan awal adalah ketika data input ODP, PDP, dan Positif bernilai 0 nilai K=3 dapat mengklasifikasi bahwa zona tersebut zona hijau sedangkan ketika nilai K=7 mengklasifikasi zona tersebut zona kuning. Ketika K=7 hasil klasifikasi banyak berupa zona kuning dan mencapai zona merah ketika nilai input ODP berada pada angka 140 orang, sedangkan K=3 terjadi perubahan klasifikasi dari zona hijau ke zona kuning pada saat nilai input ODP berada pada angka 60 dan berubah dari zona kuning ke zona merah pada saat nilai input ODP berada pada angka 120. Hasil klasifikasi saat nilai K=3 lebih sensitif terhadap perubahan data ODP dibandingkan saat nilai K=7. Hasil klasifikasi dengan k=3 memiliki hasil yang mendekati dengan hasil yang diharapkan. Nilai K=15 memiliki hasil yang sedikit berbeda namun dapat mengklasifikasikan 3 zona dengan baik dan memiliki sensitifitas terhadap perubahan data yang relatif sama dengan K=3.

Skema n+20 pada input ODP dan PDP						
Input			Hasil yang Diharapkan	Hasil KLasifikasi(K=3) Tanpa Normalisasi	Hasil KLasifikasi(K=7) Tanpa Normalisasi	Hasil Klasifikasi (K=15) Dengan Normalisasi
ODP	PDP	Positif				
0	0	0	Green Zone	Green Zone	Yellow Zone	Green Zone
20	20	10	Green Zone	Green Zone	Yellow Zone	Green Zone
40	40	20	Yellow Zone	Green Zone	Yellow Zone	Green Zone
60	60	30	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
80	80	40	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
100	100	50	Yellow Zone	Red Zone	Red Zone	Yellow Zone
120	120	60	Red Zone	Red Zone	Red Zone	Yellow Zone
140	140	70	Red Zone	Red Zone	Red Zone	Yellow Zone
160	160	80	Red Zone	Red Zone	Red Zone	Red Zone
180	180	90	Red Zone	Red Zone	Red Zone	Red Zone
200	200	100	Red Zone	Red Zone	Red Zone	Red Zone
220	220	110	Red Zone	Red Zone	Red Zone	Red Zone
240	240	120	Red Zone	Red Zone	Red Zone	Red Zone

Gambar 4.8: Hasil klasifikasi Skema 3

Pada skema 3 data input Positif tetap ditambah 10 untuk setiap iterasinya (n+10) namun data input ODP dan PDP ditambahkan 20 (n+20) untuk untuk setiap iterasinya dan menghasilkan klasifikasi sesuai Gambar 4.8. Percobaan awal adalah ketika data input ODP, PDP, dan Positif bernilai 0 nilai K=3 dapat mengklasifikasi bahwa zona tersebut zona hijau sedangkan ketika nilai K=7 mengklasifikasi zona tersebut zona kuning. Ketika K=7 hasil klasifikasi banyak berupa zona merah dan mencapai zona merah ketika nilai input ODP dan PDP berada pada angka 100 orang, sedangkan K=3 terjadi perubahan klasifikasi dari zona hijau ke zona kuning pada saat nilai input ODP dan PDP berada pada angka 60 dan berubah dari zona kuning ke zona merah pada saat nilai input ODP dan PDP berada pada angka 100. Hasil klasifikasi saat nilai K=3 dan K=7 memiliki tingkat sensitifitas yang sama terhadap perubahan data ODP dan PDP pada saat memasuki zona merah, namun nilai K=7 tidak dapat mengidentifikasi zona hijau. Hasil klasifikasi K=3 dan K=7 lebih cepat dalam mengklasifikasi perubahan zona berdasar inpu ODP dan PDP daripada hasil yang diharapkan. Hasil ketika K=15 memberikan hasil klasifikasi yang berbeda dan memiliki sensitifitas yang berbeda dengan K=3 dan K=7 karena melalui proses normalisasi data.

Skema n+20 pada setiap input						
Input			Hasil yang Diharapkan	Hasil KLasifikasi(K=3) Tanpa Normalisasi	Hasil KLasifikasi(K=7) Tanpa Normalisasi	Hasil Klasifikasi (K=15) Dengan Normalisasi
ODP	PDP	Positif				
0	0	0	Green Zone	Green Zone	Yellow Zone	Green Zone
20	20	20	Green Zone	Green Zone	Yellow Zone	Green Zone
40	40	40	Yellow Zone	Yellow Zone	Yellow Zone	Green Zone
60	60	60	Yellow Zone	Yellow Zone	Yellow Zone	Yellow Zone
80	80	80	Yellow Zone	Yellow Zone	Red Zone	Yellow Zone
100	100	100	Yellow Zone	Red Zone	Red Zone	Red Zone
120	120	120	Red Zone	Red Zone	Red Zone	Red Zone
140	140	140	Red Zone	Red Zone	Red Zone	Red Zone
160	160	160	Red Zone	Red Zone	Red Zone	Red Zone
180	180	180	Red Zone	Red Zone	Red Zone	Red Zone
200	200	200	Red Zone	Red Zone	Red Zone	Red Zone
220	220	220	Red Zone	Red Zone	Red Zone	Red Zone
240	240	240	Red Zone	Red Zone	Red Zone	Red Zone

Gambar 4.9: Hasil klasifikasi Skema 4

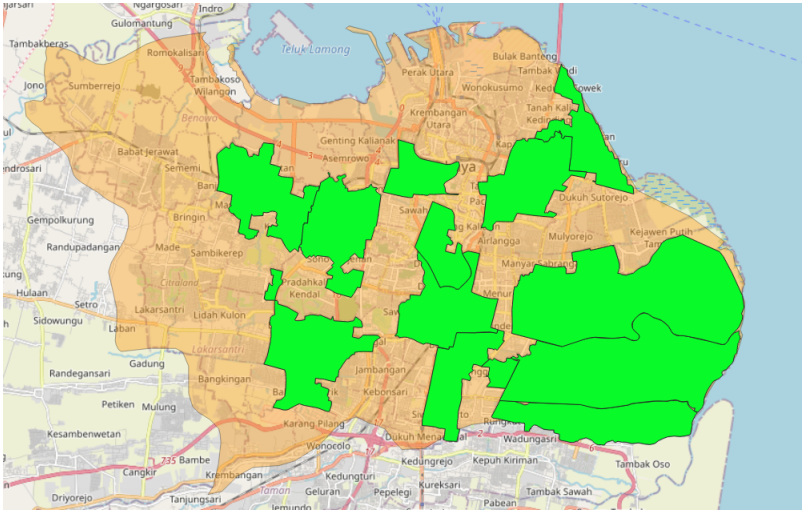
Pada skema 4 data input ditambahkan 20 untuk setiap iterasinya (n+20) dan menghasilkan klasifikasi sesuai Gambar 4.9. Percobaan awal adalah ketika data input ODP, PDP, dan Positif bernilai 0 nilai K=3 dapat mengklasifikasi bahwa zona tersebut zona hijau sedangkan ketika nilai K=7 mengklasifikasi zona tersebut zona kuning. Ketika K=7 hasil klasifikasi banyak berupa zona merah dan mencapai zona merah ketika nilai input berada pada angka 80 orang, sedangkan K=3 terjadi perubahan klasifikasi dari zona hijau ke zona kuning pada saat nilai input berada pada angka 40 dan berubah dari zona kuning ke zona merah pada saat nilai input berada pada angka 100. Hasil klasifikasi saat nilai K=7 lebih sensitif terhadap perubahan data dibandingkan saat nilai K=3, namun nilai K=7 tidak dapat mengidentifikasi zona hijau. Hasil klasifikasi dengan k=3 memiliki hasil yang mendekati dengan hasil yang diharapkan. Nilai K=15 memiliki hasil yang sedikit berbeda namun dapat mengklasifikasikan 3 zona dengan baik dan memiliki sensitifitas terhadap perubahan data yang relatif sama dengan K=3 dan K=7.

Pengujian menggunakan beberapa skema ini ditujukan mengetahui *Performance* hasil klasifikasi dari sistem klasifikasi algoritma KNN dengan hasil yang diharapkan. Penggunaan Nilai K sangat mempengaruhi hasil klasifikasi zona karena setiap nilai K memiliki kelemahan dan kelebihan masing-masing, Nilai K=3 memiliki tingkat sensitifitas yang rendah terhadap perubahan data namun dapat menghasilkan klasifikasi 3 zona dengan baik, sedangkan nilai K=7 memiliki tingkat sensitifitas yang tinggi terhadap perubahan data namun tidak dapat mengidentifikasi zona hijau pada hasil klasifikasi walaupun pada proses pengujian akurasi nilai k=7 memiliki nilai akurasi klasifikasi yang paling tinggi.

4.1.3 Pengujian Hasil klasifikasi Pada Peta Digital ditingkat Kecamatan dan Kelurahan

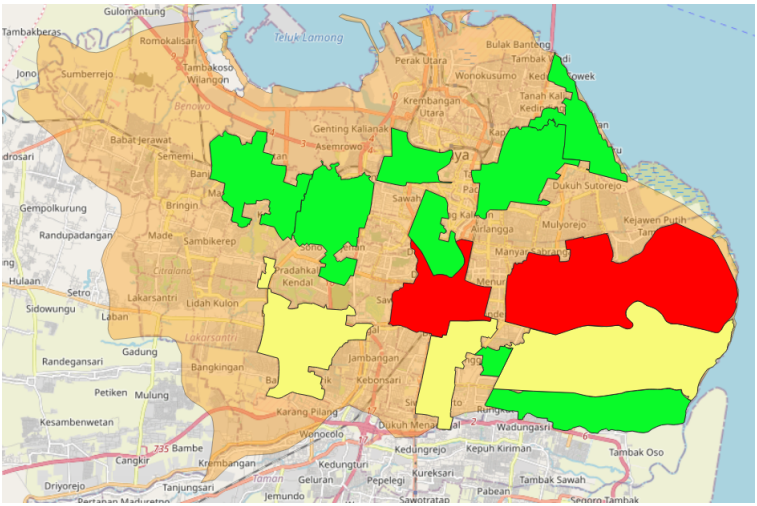
Proses visualisasi dari data-data yang telah dikumpulkan menggunakan peta digital pada sistem informasi geografis. Penggunaan peta digital juga memberikan kemudahan pada analisa terhadap suatu keadaan tertentu.

Pada proses pengujian ini akan dilakukan serangkaian klasifikasi terhadap zona di setiap kecamatan pada waktu yang berbeda. Waktu yang digunakan pada pengujian ini yaitu data persebaran COVID-19 pada tanggal 24 Maret 2020, 24 April 2020, dan 24 Mei 2020. Pengujian ini ditujukan untuk melihat perubahan pada peta digital melalui perbedaan skala waktu.



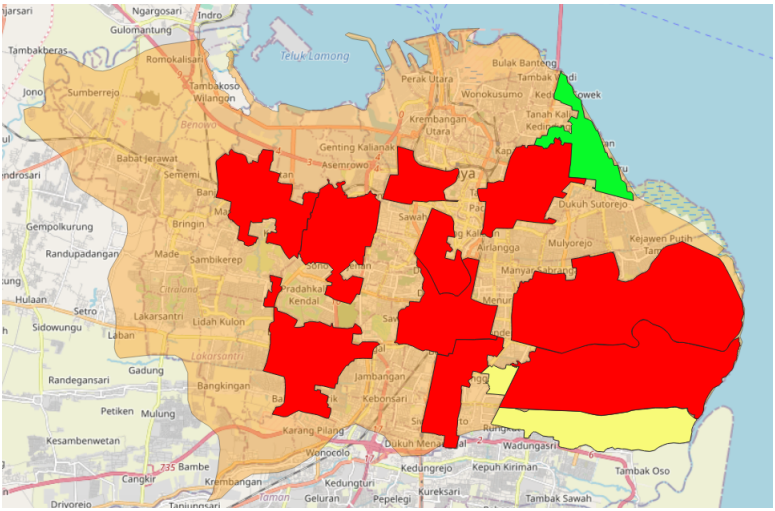
Gambar 4.10: Zona-zona yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Maret 2020

Klasifikasi zona yang dihasilkan oleh algoritma K-NN pada data persebaran COVID-19 pada tanggal 24 Maret 2020 (Gambar 4.10) berupa 13 zona hijau pada 13 kecamatan yang telah ada didalam peta digital. Hasil ini menunjukkan belum adanya perubahan yang signifikan terhadap persebaran COVID-19 pada kota surabaya.



Gambar 4.11: Zona-zona yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 April 2020

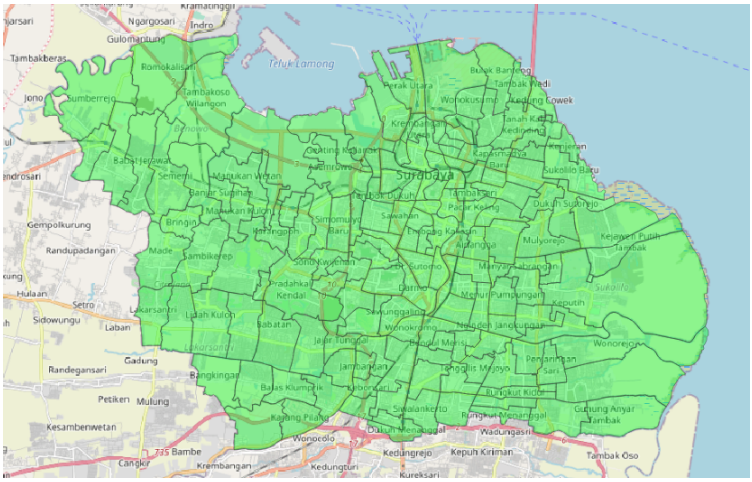
Klasifikasi zona yang dihasilkan oleh algoritma K-NN pada data persebaran COVID-19 pada tanggal 24 April 2020 (Gambar 4.11) berupa 8 zona hijau, 3 zona kuning, dan 2 zona merah pada 13 kecamatan. 2 zona merah yaitu kecamatan Sukolilo dan Kecamatan Wonokromo. 3 zona kuning yaitu Kecamatan Rungkut, kecamatan Wiyung dan kecamatan Wonocolo. 8 zona hijau yaitu kecamatan Tandes, kecamatan Bulak, kecamatan Bubutan, kecamatan Gununganyar, kecamatan Sukomanunggal, kecamatan Tambaksari, kecamatan Tegalsari, dan kecamatan Tenggilis Mejoyo. Hasil klasifikasi menunjukkan bahwa terdapat 2 kecamatan yang sudah memasuki zona beresiko tinggi terhadap penularan dan 3 kecamatan memiliki resiko sedang dalam penularan COVID-19.



Gambar 4.12: Zona-zona yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Mei 2020

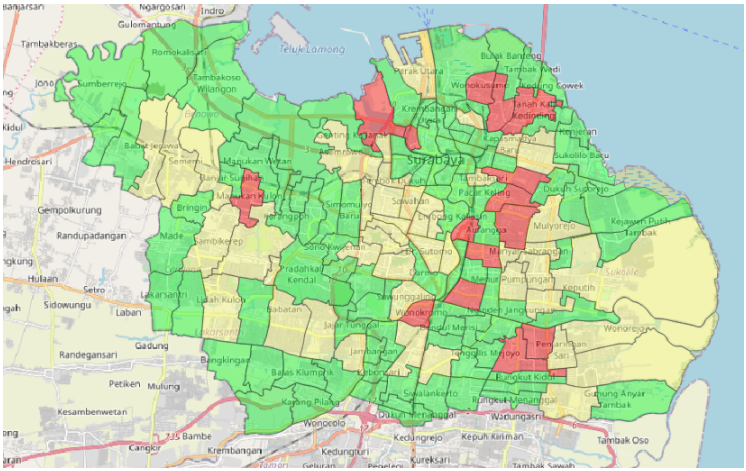
Klasifikasi zona yang dihasilkan oleh algoritma K-NN pada data persebaran COVID-19 pada tanggal 24 Mei 2020 (Gambar 4.12) berupa 1 zona hijau, 2 zona kuning dan 10 zona merah pada 13 kecamatan. 10 zona merah yaitu kecamatan Rungkut, kecamatan Sukolilo, kecamatan wonocolo, kecamatan Wonokromo, kecamatan Tegalsari, kecamatan wiyung, kecamatan Tandes, kecamatan SUkomanunggal, kecamatan Tambaksari, kecamatan Bubutan. 2 zona kuning yaitu kecamatan Gununganyar dan kecamatan Tenggilis Mejoyo. 1 zona hijau yaitu kecamatan Bulak. Terdapat beberapa perubahan hasil klasifikasi pada tanggal 24 Mei 2020 yaitu perubahan beberapa status zona pada kecamatan Rungkut, kecamatan Wonocolo, kecamatan Wiyung berubah dari zona kuning ke zona merah, Kemudian kecamatan Tandes, kecamatan Sukomanungal, kecamatan Bubutan, kecamatan Tegalsari, kecamatan Tambaksari berubah dari zona hijau ke zona merah, lalu kecamatan Gununganyar dan kecamatan Tenggilis mejoyo berubah dari zona hijau ke zona kuning.

Klasifikasi Zona pada tingkat kelurahan memiliki hasil yang berbeda karen pengklasifikasian setiap zona lebih terinci pada setiap input datanya sehingga memberikan hasil yang berbeda dari hasil klasifikasi pada tingkat kecamatan



Gambar 4.13: Zona-zona pada kelurahan yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Maret 2020 pada tingkat Kelurahan

Hasil Klasifikasi pada tingkat kelurahan di tanggal 24 Maret 2020 (Gambar 4.13) menunjukkan bahwa semua zona kelurahan masih berwarna hijau hal ini menunjukkan persebaran COVID-19 masih rendah.



Gambar 4.14: Zona-zona pada kelurahan yang dihasilkan algoritma K-NN pada data persebaran COVID-19 tanggal 24 Mei 2020

Gambar 4.14 merupakan hasil visualisasi setelah perubahan dan penambahan data persebaran COVID-19 terjadi beberapa perubahan zona pada beberapa kelurahan, contohnya seperti Kelurahan Kertajaya dan Kelurahan Mojo pada Kecamatan Gubeng terjadi perubahan dari zona hijau menjadi zona merah lalu Kelurahan Sawahan dan Kelurahan Banyuurip pada Kecamatan Sawahan yang berubah dari zona hijau menjadi zona kuning.

Hasil klasifikasi pada tingkat kelurahan memiliki hasil yang berbeda dikarenakan hasil klasifikasi pada tingkat kecamatan merupakan akumulasi dari beberapa kelurahan sehingga banyak hasil klasifikasi pada tingkat kecamatan berzona merah dikarenakan akumulasi dari beberapa kelurahan dalam kecamatan tersebut yang berzona kuning atau berzona merah.

Halaman ini sengaja dikosongkan

BAB 5
PENUTUP

5.1 Kesimpulan

Dalam penelitian ini, telah diimplementasikan serangkaian prosedur untuk membuat Sistem Informasi Geografis berbasis Google Earth API untuk pengelompokan data spasial menggunakan algoritma *K-Nearest neighbors* dan setelah dilakukan beberapa pengujian data, didapatkan kesimpulan sebagai berikut:

1. Berdasarkan hasil pengujian Akurasi Dari sebuah Klasifikasi ditentukan oleh proses *training* data dan proses pemisahan data menjadi *Test* dan *Train* bukan berdasarkan besar atau kecilnya nilai K.
2. Nilai akurasi prediksi juga bergantung pada variasi data yang telah di*training*kan pada sistem klasifikasi dan proses normalisasi data.
3. Jarak antara input pada data baru mempengaruhi prediksi klasifikasi kedepannya dan beberapa data berpengaruh besar terhadap berubahnya suatu zona pada saat klasifikasi dieksekusi.
4. Perbedaan nilai K pada saat proses prediksi mempengaruhi hasil prediksi berdasarkan variasi skema data input yang diujikan terhadap sistem.

5.2 Saran

Untuk memperbaiki sistem ini dan menyempurnakan hasil pengujian sebaiknya dilakukan beberapa hal berikut :

1. Memperbanyak variasi data pada saat proses *training* data
2. Memperbanyak variasi skema percobaan untuk mengetahui nilai K yang dapat menghasilkan prediksi yang baik sesuai keadaan nyata.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

[1] E. Hot and V. Popović-Bugarin, "Soil data clustering by using k-means and fuzzy k-means algorithm," IEEE, 2016. (Dikutip pada halaman vii, 5, 6).

[2] "Clustering with gis:an attempt to classify turkish district data." https://fig.net/resources/proceedings/fig_proceedings/fig2006/ppt/ts47/ts47_06_aksoy_ppt_0327.pdf. 06 November 2006. (Dikutip pada halaman vii, 7, 8).

[3] "K nearest neighbors - classification." https://www.saedsayad.com/k_nearest_neighbors.html. Terakhir diakses tanggal 26 April 2020. (Dikutip pada halaman vii, viii, 33, 34).

[4] "Geospatial data and resources." https://campusguides.lib.utah.edu/c.php?g=160707&p=1051727. Terakhir diakses tanggal 22 April 2020. (Dikutip pada halaman 11).

[5] "K-nearest neighbors (knn) classification model." https://www.ritchieng.com/machine-learning-k-nearest-neighbors-knn/. Terakhir diakses tanggal 26 April 2020. (Dikutip pada halaman 14).

Halaman ini sengaja dikosongkan