



**PENGEMBANGAN MODEL SISTEM
INFORMASI GEOGRAFIS UNTUK DATA
SPASIAL KESEHATAN MASYARAKAT
MENGUNAKAN SEMANTIK AI**

**DIAH TRIBUDI LESTARI
NPM: 99218019**

**PROGRAM DOKTOR TEKNOLOGI INFORMASI
UNIVERSITAS GUNADARMA**

2021

Daftar Isi

Abstrak	iii
Abstract	iv
Kata Pengantar	v
1 Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	6
1.5 Kontribusi Penelitian	7
2 Telaah Pustaka	8
2.1 Informasi Geospasial	8
2.1.1 Data Spasial	9
2.2 Sistem Informasi Geografis (SIG)	10
2.2.1 Geografi di Bidang Kesehatan	14
2.2.2 Kesalahan yang mudah diketahui pada GIS	15
2.3 Semantic Geospasial	18
2.4 <i>Artificial Intelligence</i> (AI)	21
2.5 Program Indonesia Sehat dengan Pendekatan Keluarga (PIS-PK)	22
2.5.1 Indeks Keluarga Sehat (IKS)	24
2.6 <i>Machine Learning</i>	25
2.7 Klasifikasi Teks	27
2.8 Metode Klasifikasi	28
2.8.1 K-nearest neighbors (KNN)	28
2.8.2 Naïve bayes (NB)	28

2.9 Rangkuman Hasil Penelitian Terkait	29
3 Metode Penelitian	33
3.1 Tahapan Penelitian	33
3.2 Memahami Indikator IKS	34
3.3 Memahami Indikator IKS yang Berhubungan dengan Teknik Sipil	35
3.4 Pengembangan Model	43
3.4.1 Akuisisi Data	44
3.4.2 Preproccesing Data	44
3.4.3 Klasifikasi Machine Learning	45
3.4.4 Sistem Informasi Geografis	45
3.4.5 Sistem Informasi Geospasial	45
3.4.6 Semantik Geospasial	46
3.5 Rencana Kerja	46
4 Hasil dan Pembahasan	47
4.1 Akuisisi Data	47
4.2 Preprocessing	51
4.2.1 Tahap pemeriksaan data	52
4.2.2 Penghapusan data	52
4.2.3 Pelabelan data	52
5 Kesimpulan dan Saran	53
5.1 Kesimpulan	53
5.2 Saran	53
Daftar Pustaka	53
LAMPIRAN	55
RIWAYAT HIDUP	56

Daftar Tabel

2.1	Rangkuman Hasil Penelitian Terkait	29
3.1	Indikator Keluarga Sehat	35
3.2	Indikator IKS Teknik Sipil	36
3.3	Rencana Kerja	46
4.1	Data Laporan IKS Januari 2020	47
4.2	Label Data	52

Daftar Gambar

2.1	Model Manajemen Data Spasial	10
2.2	Komponen SIG	11
2.3	Organisasi SIG	12
2.4	SIG Description (De Mers, 1997)	14
2.5	Contoh Analisis Spasial Kasus Malaria Subang (SIMKES,2009)	15
2.6	(a) Fragmen Ontology Proyek NSF GeoLink (Adila Krisna- dhi,et al., 2015); (b) Bagian dari Gazetteer	20
2.7	Indikator Keluarga Sehat (IKS)	23
2.8	Penjabaran Visi & Misi Presiden Menjadi Program Indonesia Sehat	24
2.9	Hasil Indeks Keluarga Sehat Kab. Ngawi	25
2.10	Skema Artifical Intelligence dan Machine Learning	27
3.1	Tahapan Penelitian	34
3.2	Tahap Pengembangan Model	44
4.1	Data IKS Setiap Wilayah (a)	49
4.2	Data IKS Setiap Wilayah (b)	49
4.3	Data IKS Setiap Wilayah (c)	50
4.4	Data IKS Setiap Wilayah (d)	51

Bab 1

Pendahuluan

1.1 Latar Belakang

Indonesia merupakan Negara yang terletak di daerah tropis dengan jumlah penduduk hampir mencapai 220 juta jiwa. Kondisi lingkungan yang kurang baik, faktor ekonomi dan perilaku kesehatan masyarakat yang masih buruk, sehingga angka kasus penyakit tropis masih tinggi [Putu Kurniawan,2014]. Kementerian Kesehatan berusaha membuat program yang dinamakan Indeks Keluarga Sehat (IKS). IKS dilakukan dengan strategi pendekatan keluarga, keluarga yang sehat akan diukur dengan 12 indikator keluarga sehat.

Perhitungan kedua belas indikator keluarga sehat dari setiap keluarga yang besarnya berkisar antara 0 sampai dengan 1. Keluarga yang tergolong dalam keluarga sehat adalah dengan $IKS > 0,8$ [Kementerian Kesehatan RI,2016]. Capaian IKS masih tergolong rendah untuk beberapa daerah di Indonesia, hasil perhitungan IKS dari 9 provinsi sasaran awal yaitu Sumatera Utara, Sumatera Selatan, Lampung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten dan Sulawesi Selatan per 8 Juni 2017 didapatkan keluarga yang memiliki IKS di atas 0,8 sebesar 0,163 dari 570.326 keluarga [Pusdatin,2016].

Program Indonesia Sehat dengan pendekatan keluarga merupakan salah satu dari Agenda ke-5 Nawa Cita, yaitu Meningkatkan Kualitas Hidup Manusia Indonesia. Program Indonesia Sehat dengan pendekatan keluarga ini selanjutnya sudah dituangkan ke dalam bentuk rencana jangka menengah yang merupakan penjabaran dari Rencana Pembangunan Jangka Menengah Nasional 2015-2019, melalui Keputusan Menteri Kesehatan R.I. No-

mor HK.02.02/Menkes/52/2015 tentang Rencana Strategis Kementerian Kesehatan Tahun 2015-2019.

Puskesmas sebagai fasilitas pelayanan kesehatan pertama merupakan kunci dalam pelaksanaan pembangunan kesehatan menuju Indonesia Sehat melalui pendekatan keluarga. Pendekatan keluarga adalah pendekatan pelayanan kesehatan yang dilakukan oleh Puskesmas yang mengintegrasikan Upaya Kesehatan Perseorangan (UKP) dan Usaha Kesehatan Masyarakat (UKM) secara berkesinambungan kepada keluarga berdasarkan siklus hidup dari pelayanan kesehatan pada ibu hamil sampai lansia. Pelaksanaan pembangunan kesehatan menuju Indonesia Sehat melalui pendekatan keluarga melibatkan peran serta jaringan, jejaring Puskesmas dan masyarakat. Upaya yang dilakukan dengan memanfaatkan sumber daya yang ada di masyarakat salah satunya ialah Upaya Kesehatan Bersumberdaya Masyarakat (UKBM). Pengembangan dan pembinaan UKBM yang ada di desa dilakukan dalam rangka pemberdayaan kemandirian masyarakat. Puskesmas memiliki peranan dalam membina UKBM untuk menyelaraskan pelaksanaan pembangunan.

Kementrian Kesehatan di dalam melakukan pencegahan dan pemberantasan penyakit akan sangat efektif bilamana mendapat dukungan dari sebuah sistem informasi untuk meninjau penyebaran penyakit, karena sistem informasi dapat menyediakan informasi epidemiologi yang peka terhadap perubahan yang terjadi dalam kasus penyebaran penyakit. *Surveilans epidemiologi* adalah suatu proses pengamatan yang dilakukan secara berkesinambungan dan sistematis terhadap penyebaran penyakit yang akan kemudian disebarluaskan kepada pihak-pihak yang bertanggung jawab dalam pencegahan penyakit dan masalah kesehatan lainnya. Sistem informasi geografis merupakan rekaman fenomena atau obyek-obyek keruangan yang selanjutnya diolah menjadi Informasi Geospasial (IG). IG merupakan suatu sistem yang dapat digunakan untuk pemasukan, penyimpanan, menampilkan, dan keluaran informasi geografis berikut atribut-atributnya IG memiliki kemampuan yang sangat baik dalam memvisualisasikan data spasial berikut, memodifikasi bentuk, warna, ukuran, symbol. Sistem informasi geografis dapat digunakan oleh berbagai bidang ilmu, pekerjaan dan peristiwa. Banyak sekali masalah yang dapat ditangani oleh sistem informasi geografis [Guruh Sapto N,2013].

Sistem Informasi Geografis berperan penting dalam sektor kesehatan masyarakat, IG menggabungkan algoritma, analisis spasial, geo-statistik

dan pemodelan, menjadikan teknologi sistem informasi geografis alat untuk memprediksi pola penyakit dan parasite asosiasi ekologi [higgs,2004; Guo dkk 2005; Garcia- Rangel dan Pettorelli,2013]. Munculnya dengan metode Sistem informasi geospasial (SIG) baru-baru ini telah meningkatkan pemahaman tentang hubungan spasial antara kesehatan dan tempat, sehingga SIG ini dapat dijadikan alat yang efektif untuk menangani masalah perencanaan kesehatan masyarakat.

Perkembangan teknologi Informasi Geospasial (IG) ini juga telah memicu tumbuhnya industri IG dunia beberapa dekade terakhir ini di berbagai bidang layanan, namun terkait dengan Industri IG di Indonesia masih terdapat kendala untuk dapat tumbuh sebagai industri mandiri, di antaranya disebabkan lemahnya daya saing industri IG. UU No.4 tahun 2011 tentang informasi geospasial telah disahkan pada tanggal 21 april 2011. Lahirnya undang-undang ini menjamin ketersediaan dan akses terhadap informasi geospasial yang dapat dipertanggung jawabkan. IG merupakan bagian penting dalam mewujudkan sistem informasi yang dapat dimanfaatkan untuk mendukung proses perencanaan, pelaksanaan, dan evaluasi [Nurhayati Imas,2017]

Fenomena yang terjadi saat ini dengan kondisi masyarakat yang kurang memahami peran penting dari IKS seperti di daerah Desa Payaman, 2016 memiliki 196 KK atau 79% yang tidak memiliki sarana pembuangan air limbah yang sesuai dengan persyaratan kesehatan, dengan menyalurkan limbah cair pada kawasan rumah tangga diperlukan saluran berupa saluran pembuangan air limbah maupun resapan, sehingga tidak mencemari air tanah, tidak menimbulkan sarang nyamuk, dan sakit diare.

Analisis spasial menggunakan Sistem Informasi Geografis dimaksudkan untuk mendukung pengambilan keputusan surveilans dan penanggulangan penyakit. Proses pengambilan keputusan harus benar-benar mengetahui informasi terakhir mengenai situasi penyakit, populasi berisiko dan trend terjadinya kasus di masa datang di wilayahnya. Kebutuhan khusus bagi layanan penanggulangan penyakit berbasis lingkungan bervariasi sesuai dengan perbedaan situasi lingkungan dan epidemiologi. Dibutuhkan pemahaman secara cepat berdasarkan informasi yang ada dan penyediaan informasi terbaru yang berguna untuk memandu kegiatan di lapangan. Misalnya kapan dan di mana harus mengintervensi dan intervensi apa yang paling efektif, bagaimana suatu intervensi menjadi layak walau dengan sumber daya yang

terbatas. Guna pengambilan keputusan penting, maka sudah selayaknya para pengambil keputusan memperoleh informasi yang mudah dipahami dan dapat dipercaya [sunaryo,2015]. Sesuai dengan upaya Kementerian Kesehatan menyelenggarakan program IKS sebagai upaya mewujudkan masyarakat Indonesia yang berperilaku sehat, hidup dalam lingkungan sehat, serta mampu menjangkau pelayanan kesehatan yang bermutu untuk mencapai derajat kesehatan yang setinggi-tingginya, sehingga dalam rangka penyelenggara program IKS dengan pendekatan keluarga, ditetapkan 12 indikator utama sebagai penanda status kesehatan sebuah keluarga [Pusdatin,2016]

Semantic Network pertama kali dikembangkan untuk *Artificial Intelligence* (AI) sebagai cara untuk merepresentasikan memori dan pemahaman bahasa manusia. Struktur *semantic net* berupa grafik dengan *node* (simpul) dan *arc* (ruas) yang menghubungkannya. Semantik network sering disebut dengan semantik AI. [Nava'atul Fadillah, Novrido Charibaldi, Herlina Jayadianti, 2010]. *Artificial Intelligence* (AI) merupakan representasi dari pengetahuan yang berhubungan erat dengan teknologi berbasis komputer dan menekankan pada kemampuan komputer untuk meniru manusia dalam melakukan pembelajaran dan mengatasi masalah [Heryadi and Irwansyah, 2020].

Semantik geospasial adalah bidang yang mengadopsi perspektif penelitian yang unik terhadap masalah geospasial. Sampai batas tertentu, semantik geospasial dapat dibandingkan dengan statistik geospasial: keduanya bisa diterapkan ke berbagai masalah di seluruh domain dan keduanya memiliki kumpulan metode uniknya sendiri (mis., pemodelan ontologis dan *natural language processing* (NLP) untuk semantik geospasial). Dalam beberapa tahun terakhir, banyak sekali penelitian tentang semantik geospasial yang telah dilakukan, tujuannya adalah untuk menggambarkan peta yang memberikan gambaran umum tentang enam utama area penelitian pada semantik geospasial [Yingjie Hu,2017].

Semantik geospasial menambahkan kata sifat geospasial dan penambahan ini membatasi dan memperluas area awal semantik yang berlaku. Di satu sisi, semantik geospasial berfokus pada ekspresi yang lebih terkait dengan geografi daripada ekspresi umum apa pun; di sisi lain, semantik geospasial memungkinkan studi tidak hanya ekspresi linguistik tetapi juga arti dari tempat-tempat geografis, data geospasial, dan GeoWeb [Yingjie Hu,2017]

Penelitian [Guruh Sabdo N, 2013] tentang SIG kesehatan bahwa sistem

penyampaian informasi yang ada saat ini kurang efektif dan belum menampilkan kenampakan sebaran DBD secara geografis, sehingga tidak diketahui daerah mana saja yang terkena dampak DBD terendah hingga tertinggi. Oleh karena itu diperlukan metode dalam penyelesaiannya dimana data yang ditampilkan juga diikuti data geografis wilayah terdampak. SIG salah satunya dimana sebaran DBD di Wilayah Kota Solo diolah dengan menggabungkan data spasial dengan data non spasial yang berupa *attribute* data lalu diimplementasikan dengan mengkonvert kedalam database menggunakan *open-svgsmapserver*. Data spasial pada penyakit DBD dikumpulkan dan dimasukkan di tingkat desa dimana surveilans epidemiologi dikumpulkan secara rutin, data lingkungan dan fasilitas surveilans penyakit akan diintegrasikan ke dalam SIG untuk memudahkan pengguna dalam mempelajari penyebaran spasial penyakit DBD.

Selain di bidang kesehatan, SIG pun dapat membantu permasalahan penanganan banjir dengan cara memberi informasi mengenai kondisi fisik suatu daerah meliputi kemiringan lereng, jenis tanah, penggunaan lahan, tingkat kerentanan banjir dan jumlah rumah yang harus dievakuasi apabila wilayah tersebut terjadi banjir. Selain itu dapat membantu pemerintah maupun donator dalam penyaluran bantuan agar lebih efektif [Agus Anggoro, 2011]. Dibangun SIG yang dapat meningkatkan kreativitas dan inovasi juga mengacu pengembangan perangkat lunak nasional melalui pengembangan SIG berbasis web, sehingga dapat membantu program pemerintah terutama untuk penanganan bencana.

Baru-baru ini, beberapa desain Building Information Model (BIM) untuk menyimpan informasi semantic tentang gedung dan 3D model informasi geospasial untuk merepresentasikan objek dunia nyata yang telah dilaporkan. Tujuan dari model ini adalah untuk menawarkan sarana mendefinisikan objek spasial dengan geometri dan semantik. *Industry Foundation Classes* (IFC) dan *City Geography Markup Language* (CityGML) saat ini dianggap sebagai dua model semantic yang paling menonjol dalam desain dan objek dunia nyata [ilmu geo-informasi 3D, LNG & C, springer Verlag, hlm, 2009]. Berdasarkan latar belakang diatas maka dalam penelitian ini akan dibangun model sistem informasi geospasial dari data IKS dengan menggunakan pendekatan semantik geospasial. Data SIG ini dapat dimanfaatkan untuk membantu pemerintah dari segi kesehatan maupun infrastruktur bangunan.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka perlu dirumuskan masalah-masalah penelitian yang harus dicari solusinya, yaitu:

1. Bagaimana mengembangkan model semantik sistem informasi geospasial berdasarkan indikator nilai IKS pada wilayah Indonesia.
2. Bagaimana mengekstraksi data IKS untuk mengembangkan model sistem informasi geospasial tentang kesehatan masyarakat.
3. Bagaimana mengidentifikasi resiko kesehatan masyarakat terhadap infrastruktur pada wilayah Indonesia.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan dalam sistem informasi geospasial berdasarkan indikator nilai IKS
2. Pendekatan metode semantik geospasial yang digunakan adalah *semantic network*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah mengembangkan model semantik informasi geospasial kesehatan masyarakat, bertujuan khusus:

1. Mengembangkan model semantic sistem informasi geospasial berdasarkan indikator nilai IKS pada wilayah Indonesia.
2. Melakukan ekstraksi data IKS untuk mengembangkan model sistem informasi geospasial tentang kesehatan masyarakat.
3. Mengidentifikasi resiko kesehatan masyarakat terhadap infrastruktur pada wilayah Indonesia.

1.5 Kontribusi Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi, sebagai berikut:

1. Menghasilkan sebuah model sistem informasi geografis tentang kesehatan masyarakat pada suatu wilayah
2. Menghasilkan semantik geospasial dari sistem informasi geografis kesehatan masyarakat sebagai penunjang keputusan
3. Hasil dari analisis semantik geospasial dapat membantu pemerintah dalam hal ini Kementrian Kesehatan atau Kementrian Pekerjaan Umum dalam penanganan masalah Kesehatan dan infrastruktur pada suatu wilayah

Bab 2

Telaah Pustaka

2.1 Informasi Geospasial

Informasi Geospasial atau ruang kebumian adalah aspek keruangan yang menunjukkan lokasi, letak, dan posisi suatu objek atau kejadian yang berada di bawah, pada, atau di atas permukaan bumi yang dinyatakan dalam sistem koordinat tertentu. Data Geospasial yang selanjutnya disingkat DG adalah data tentang lokasi geografis, dimensi atau ukuran, dan/atau karakteristik objek alam dan/atau buatan manusia yang berada di bawah, pada, atau di atas permukaan bumi. Informasi Geospasial yang selanjutnya disingkat IG adalah DG yang sudah diolah sehingga dapat digunakan sebagai alat bantu dalam perumusan kebijakan, pengambilan keputusan, dan/atau pelaksanaan kegiatan yang berhubungan dengan ruang kebumian.

Beberapa perkembangan di bidang geodesi dan geomatika setelah adanya UU no 4 tahun 2011 yaitu: menjamin ketersediaan dan akses terhadap Informasi Geospasial yang dapat dipertanggung jawabkan. Undang-undang tentang Informasi Geospasial ini menjadi aturan yang mengikat bagi seluruh pemangku kepentingan, sehingga dapat dimanfaatkan untuk menjaga keutuhan Negara Kesatuan Republik Indonesia (NKRI) dan meningkatkan kesejahteraan masyarakat. Sebagai pendukung pengelolaan sumber daya alam dan sumber daya lainnya di negeri ini bagi kemakmuran seluruh rakyat Indonesia, di masa kini dan masa yang akan datang. Dengan disebarluaskannya Informasi Geospasial akhirnya pengetahuan kalangan umum akan berbagai sumber daya yang ada di Indonesia meningkat. Berdasarkan rangkuman UU geospasial itu untuk mendukungnya secara menyeluruh, dibutuhkan banyak

tenaga geomatika di seluruh wilayah Indonesia.

Dari pengetahuan yang berkaitan dengan keruangan geospasial tersebut, kemudian untuk membuat peta dan melakukan analisa, dibutuhkan penggunaan software Sistem Informasi Geografi (SIG).

2.1.1 Data Spasial

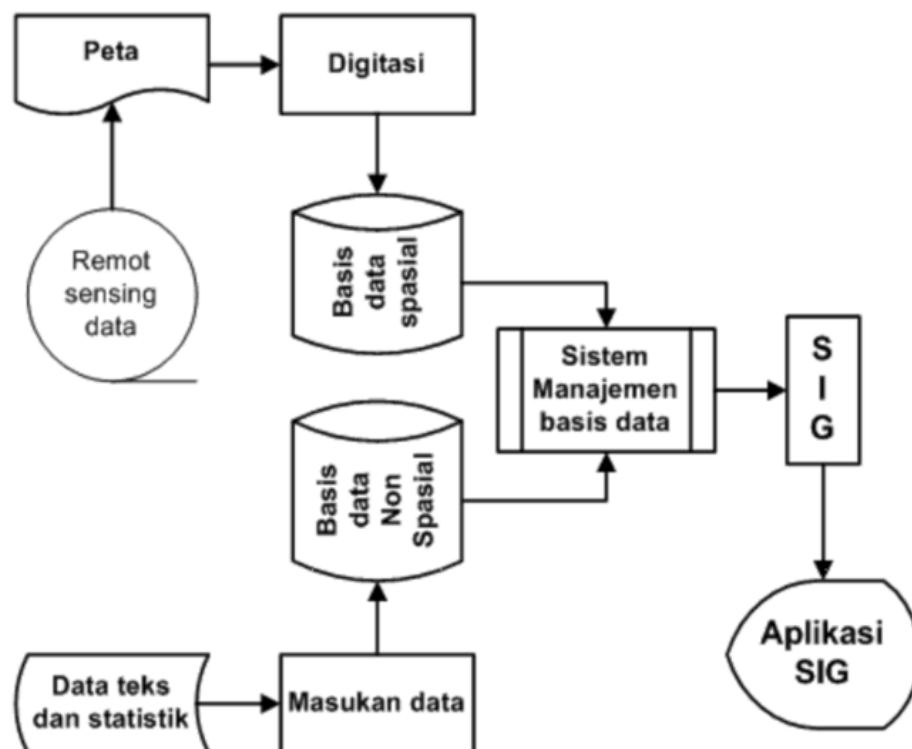
Data spasial mempunyai pengertian sebagai suatu data yang mengacu pada posisi, objek, dan hubungan diantaranya dalam ruang bumi. Data spasial merupakan salah satu item dari informasi, di mana di dalamnya terdapat informasi mengenai bumi, termasuk permukaan bumi, perairan, kelautan, dan bawah atmosfer [Rajabidfard dan Williamson,2000]. Data spasial dan informasi turunannya digunakan untuk menentukan posisi dari identifikasi suatu elemen dipermukaan bumi [Radjabidfard,2001]. Lebih lanjut lagi Mapping Science Committee (1995) dalam [Rajabidfard,2001] menerangkan mengenai pentingnya peranan posisi lokasi, yaitu (1) pengetahuan mengenai lokasi yang berdekatan dan (2) lokasi memungkinkan diperhitungkannya jarak, pembuatan peta dan memberikan arahan dalam membuat keputusan spasial yang bersifat kompleks.

Karakteristik utama dari data spasial adalah bagaimana mengumpulkannya dan memeliharanya untuk berbagai kepentingan. Selain itu juga ditujukan sebagai salah satu elemen yang kritis dalam melaksanakan pembangunan social ekonomi secara berkelanjutan dan pengelolaan lingkungan. Berdasarkan perkiraan, hamper lebih daari 80% informasi mengenai bumi berhubungan dengan informasi spasial [Wulan,2002].

Rajabidfard dan wiliamson [2000], menernagkan bahwa terdapat dua pendorong utama dalam pembangunan data spasial. Pertama adalah pertumbuhan kebutuhan suatu pemerintah dan dunia bisnis dalam memperbaiki keputusan yang berhubungan dengan keruangan dan meningkatkan efisiensi dengan bantuan data spasial. Faktor pendorong kedua adalah mengoptimalkan anggaran yang ada dengan meningkatkan informasi dan sistem komunikasi secara nyata dengan membangun teknologi informasi soasial. Didorong oleh faktor-faktor tersebut, banyak negara, pemerintahan, dan organisasi memandang pentingnya data spasial, terutama dalam pengembangan informasi spasial, atau yang lebih dikenal dengan Sistem Infromasi Geografis (SIG). tujuannya adalah membantu mengambil keputusan berdasarkan ke-

pentingan dan tujuannya masing-masing, terutama yang berkaitan dengan aspek keruangan.

SIG menyimpan data dalam bentuk ‘peta’ berupa bentuk untuk geometri atau spasial (titik, garis, dan / tau area/ polygon) dan informasi disimpan dalam bentuk atribut. Saat ini SIG dikembangkan dengan menggunakan sistem-sistem manajemen basis data (DBMS) yang telah ada sebelumnya seperti pada Gambar 2.1.



Gambar 2.1: Model Manajemen Data Spasial

2.2 Sistem Informasi Geografis (SIG)

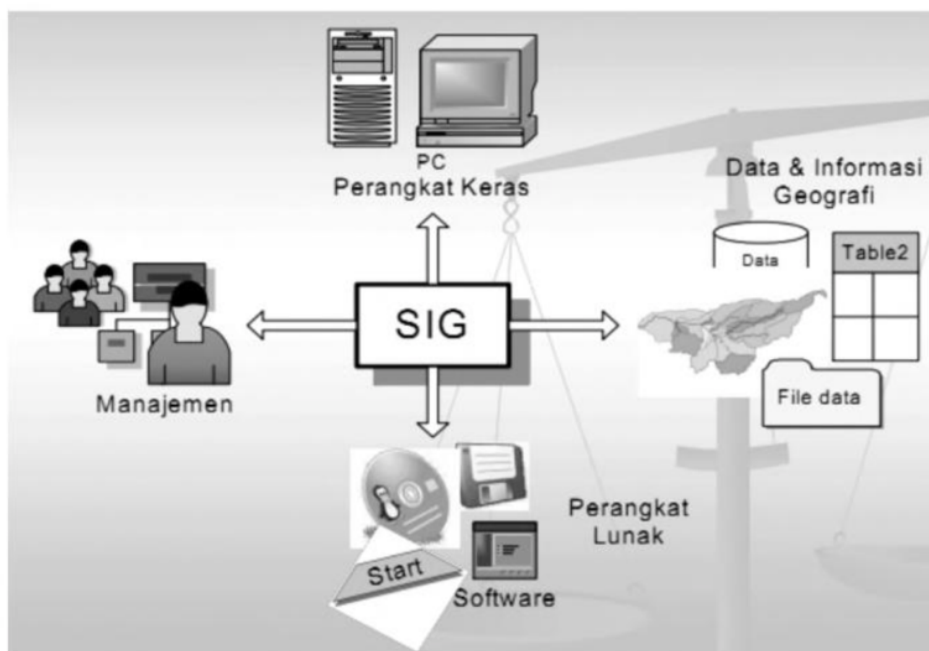
Sistem Informasi Geografi adalah sistem basis data dengan kemampuan analisis untuk data yang tereferensi secara spasial. SIG mempunyai kemampuan untuk mengintegrasikan data spasial dan data atribut sehingga dalam analisisnya mampu menghasilkan informasi yang diinginkan [Murai, 1999 dalam Anggoro, TI. dkk, 2019].

SIG mempunyai kemampuan untuk menghubungkan berbagai data pada satu titik tertentu di bumi, menghubungkannya, lalu menganalisa dan

akhirnya memetakan hasilnya. Data yang diolah pada SIG merupakan data spasial yaitu sebuah data yang berorientasi geografis dan merupakan lokasi yang memiliki system koordinat tertentu sebagai referensinya. Sehingga aplikasi SIG dapat menjawab beberapa pertanyaan seperti; kondisi, lokasi, trend, pola dan permodelan [Aronaff, 1989 dalam Anggoro, TI. dkk, 2019]

Istilah “informasi geografis” mengandung pengertian informasi mengenai tempat-tempat yang terletak di permukaan bumi, pengetahuan mengenai posisi dimana suatu objek terletak di permukaan bumi, dan informasi mengenai keterangan-keterangan (atribut) yang terdapat di permukaan bumi yang posisinya diberikan atau diketahui.

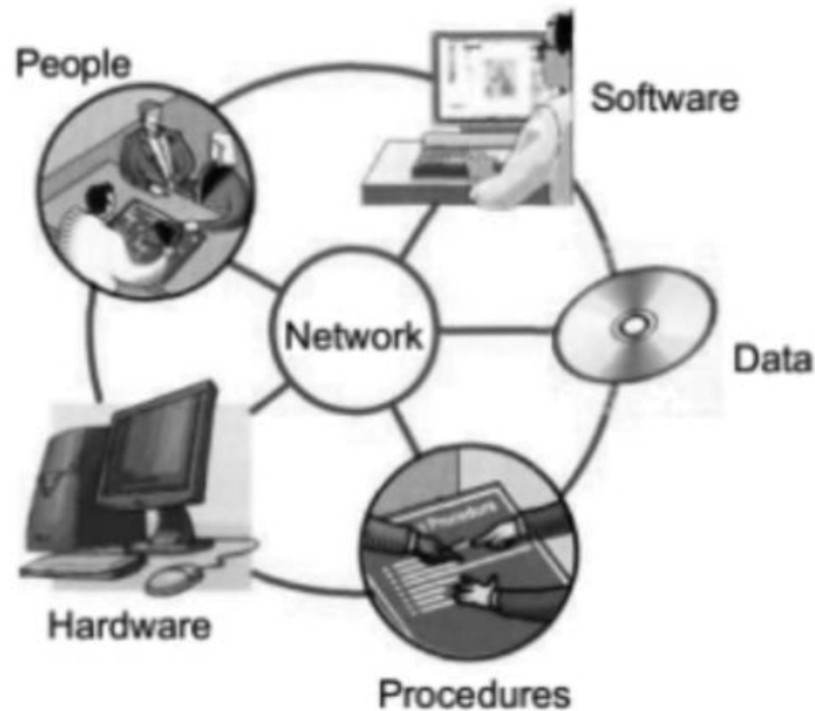
Dengan memperhatikan pengertian sistem informasi, maka SIG adalah suatu kesatuan formal yang terdiri dari berbagai sumberdaya fisik dan logika yang berkenaan dengan objek-objek yang terdapat di permukaan bumi. Jadi SIG juga merupakan sejenis perangkat lunak yang dapat digunakan untuk pemasukan, penyimpanan, manipulasi, menampilkan, dan keluaran informasi geografis berikut atribut-atributnya.



Gambar 2.2: Komponen SIG

Selain kelima komponen pada Gambar 2.2, ada suatu komponen yang sebenarnya tidak kalah penting, yaitu metode. Sebuah SIG yang baik adalah apabila didukung dengan metode perencanaan desain sistem yang baik se-

suai dengan business rules organisasi yang menggunakan SIG tersebut seperti pada Gambar 2.3.



Gambar 2.3: Organisasi SIG

Perkembangan SIG kesehatan di Indonesia dimulai tahun 1990an, pada saat itu kita kenal dengan SIG PPM (Sistem Informasi Geografis-Pemberantasan Penyakit Menular). Kemudian pada awal tahun 2000an Dapeng Luo, konsultan dari Proyek IPPM (Intensifikasi Pemberantasan Penyakit Menular) mengembangkan SIG (ArcView) kesehatan masyarakat (Malaria, TB, Imunisasi). Khusus untuk lokasi pilot proyek penelitian malaria dilakukan di Kecamatan Banjarmangu Kabupaten Banjarnegara, yang selanjutnya sebagai acuan secara Nasional. Sejak saat itu SIG kesehatan masyarakat banyak dimanfaatkan di Institusi kesehatan dari tingkat puskesmas sampai tingkat pusat. Lingkup kerja SIG kesehatan masyarakat juga telah meluas, tidak hanya pada program pengendalian penyakit, akan tetapi meliputi promosi program kesehatan masyarakat.

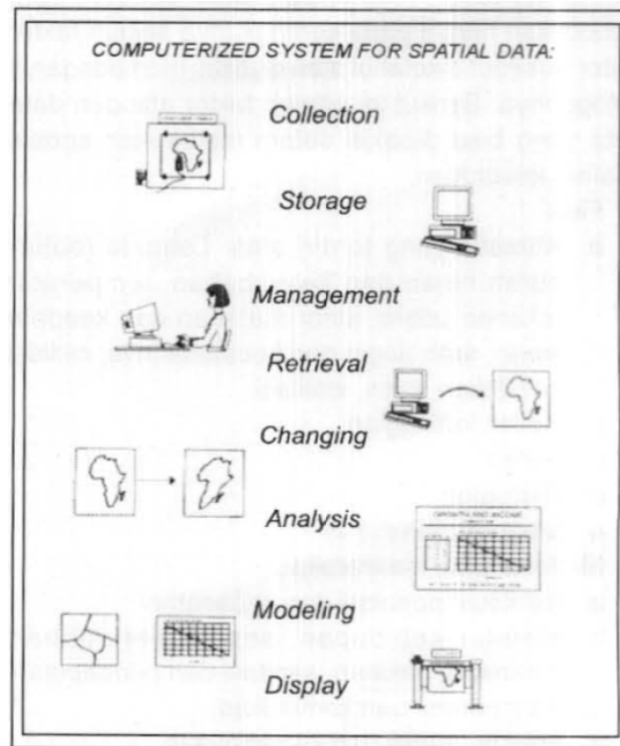
Sistem Informasi Geografis (GIS) berperan penting dalam sektor kesehatan masyarakat, GIS menggabungkan algoritma, analisis spasial, geo-statistik dan pemodelan, menjadikan teknologi SIG alat untuk memprediksi pola pe-

nyakit dan parasite asosiasi ekologi [higgs,2004; Guo dkk 2005; Garcia- Rangel dan Pettorelli,2013]. Munculnya dengan metode GIS baru-baru ini telah meningkatkan pemahaman tentang hubungan spasial antara kesehatan dan tempat, sehingga GIS ini dapat dijadikan alat yang efektif untuk menangani masalah perencanaan kesehatan masyarakat.

Teknologi ini juga memungkinkan analisis hubungan spasial anatara dimensi yang berbeda. Pada kesehatan masyarakat SIG dapat digunakan untuk menggambarkan besar masalah kesehatan dan indentifikasi determinan kesehatan dengan spesifik, sebagai masukan proses pengambilan keputusan, surveilans, intervensi kesehatan dan strategi pencegahan penyakit, serta untuk analisis epidemiologi dan manajemen kesehatan masyarakat.

Pemetaan masalah kesehatan masyarakat secara manual sudah mulai ditinggalkan, sejak perkembangan teknologi pemetaan secara digital banyak digunakan pada institusi kesehatan. Indonesia, pada saat ini SIG dibidang kesehatan telah dikenal luas sebagai alat bantu surveilans, bahkan pada tingkat lanjut SIG dapat digunakan memprediksi suatu kejadian penyakit berdasarkan faktor resiko.

Pada Gambar 2.4, De Mers [1997] SIG adalah suatu alat dengan sistem komputer yang digunakan untuk memasukan (capturing), menyimpan, memeriksa, mengintegrasikan, memanipulasi, menganalisa dan menampilkan data-data yang berhubungan dengan posisi-posisi permukaan bumi.



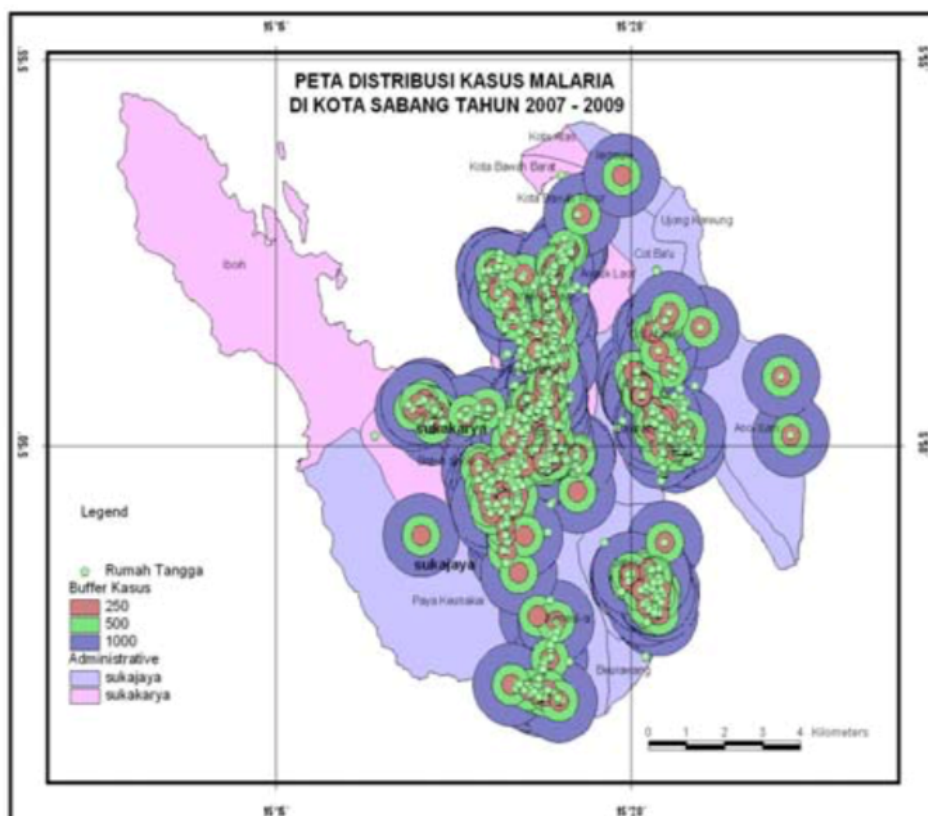
Gambar 2.4: SIG Description (De Mers, 1997)

2.2.1 Geografi di Bidang Kesehatan

Geografis merupakan *"a science concern with rational development and testing of theories that explain and predict the spatial distribution and location of various characteristics on the surface of the earth"* [yates, dalam alfandi, 2001]. Istilah geografi di bidang kesehatan di Indonesia masih terdengar asing. Banyak sekali yang bertanya "apa hubungannya antara geografis dengan kesehatan?". Pertanyaan tersebut adalah mengenai konsep dan defines 'medical geografis' sudah dicatat oleh May[1970], [Learmonth,1975,1978; Boleh, 1983; Philips, 1981]. Dari pertanyaan-pertanyaan yang muncul istilah geografis pelayanan kesehatan, geografi perawatan medik, geografi gizi, geografi epidemiologi, dan lainnya.

Meskipun dalam geografi kesehatan terdapat macam-macam istilah geografi dapat digabungkan menjadi 2 kelompok, yaitu Geografi medis, dan sistem pelayanan masyarakat. Geografi medis diutarakan mengenail eksplorasi, deskripsi dan pemodelan ruang waktu, berkaitan dengan persoalan lingkungan, deteksi dan analisis cluster dan pola penyebran penyakit. Se-

dangkan dengan geografi sistem pelayan masyarakat berkaitan dengan perencanaan, manajemen dan jaminan pelayanan agar sesuai kebutuhan, merumuskan kebutuhan kesehatan masyarakat yang dilayani dan pola wilayah yang dilayani service catchment zones [Boulos,2000]. Penyajian informasi kesehatan dengan menggabungkan antara data dan peta bukanlah hal baru dalam bidang kesehatan. Pada tahun 1854, John Snow secara manual menampilkan informasi wabah kolera dalam bentuk peta sehingga dapat menentukan sumber penularan penyakit, tanpa mengetahui jenis bakteri dan cara penularan wabah [Riner et al., 2004]. Sejak saat itu, penggunaan SIG berkembang lebih luas, tidak hanya terbatas untuk memetakan distribusi penyakit tetapi juga distribusi tenaga dan fasilitas kesehatan, seperti yang ditunjukkan pada Gambar 2.5.



Gambar 2.5: Contoh Analisis Spasial Kasus Malaria Subang (SIMKES,2009)

2.2.2 Kesalahan yang mudah diketahui pada GIS

- Umur Data

Umur data berbagai data bervariasi, karena tidak mungkin data peta dikumpulkan pada waktu yang sama. Kebanyakan institusi yang berkecimpung dalam bidang lingkungan dan perencanaan cenderung memakai data yang mudah dipublikasi, baik dalam bentuk peta maupun laporan dan dilengkapi dengan citra penginderaan jauh studi lapangan. Berbagai obyek yang mempunyai sifat dinamik maka pendekatan ini tidak tepat, tetapi untuk berbagai data dasar maka hal ini tidak menjadi persoalan. Data spasial yang bersifat dinamik antara lain penggunaan lahan, sedangkan data spasial yang tidak bersifat dinamik antara lain data topografi, data tanah atau yang paling lama dalam bidang ilmu kebumihantaran adalah data geologi. Dalam pengertian yang luas sifat geologi berubah lebih lambat dibanding tanah, rejim air, vegetasi atau penggunaan lahan. Tetapi ada juga kemungkinan bahwa data yang tua tidak sesuai karena pada waktu pembuatannya tergantung ke sistem standar yang tidak sesuai lagi dengan keadaan sekarang.

- Cakupan Areal Studi

Idealnya untuk seluruh daerah studi kualitas informasi seragam. Secara umum data yang tersedia di suatu daerah baik untuk informasi individu maupun untuk berbagai tingkat informasi, di negara-negara yang sudah berkembang sekalipun, dapat dikatakan bahwa tidak ada cakupan informasi sumberdaya lahan yang lengkap pada seluruh daerah studi, kecuali untuk wilayah yang kecil yang tidak sesuai untuk tujuan tertentu. Dalam hal peta tanah misalnya, banyak Negara yang cakupan peta tanahnya terpecah-pecah pada skala antara 1:25,000 – 1:50,000. Selain itu selama 30–40 tahun berlalunya fungsi survei tanah, konsep-konsep dan definisi-definisi tanah, serta cara-cara tanah dipetakan, bahkan termasuk surveyornya juga telah berganti. Sistem administrasi peta untuk berbagai instansi juga dapat berbeda sehingga penyambungan antara lembar peta tidak dapat dilakukan. Jika peta cakupannya tidak lengkap, maka keputusan memperbaiki harus dibuat sehingga keseragaman yang diinginkan dapat tercapai. Pilihan dapat dibuat dengan menambah data lebih banyak pada bagian yang kurang, atau menggeneralisasikan data yang lebih detail disesuaikan dengan data yang kurang detail. Perlu dicatat bahwa tidak diijinkan memperbesar skala peta yang lebih kecil untuk memenuhi cakupan wilayah, karena selain kedetilannya tidak bertambah penyajian skala baru tersebut melanggar kode etik kartografi.

- Kerapatan Pengamatan

Kerapatan pengamatan merupakan ukuran keakuratan suatu peta, akan tetapi sering tidak dicantumkan dalam berbagai peta tematik khususnya pada peta tanah. Di Indonesia pada berbagai proyek tertentu seperti Land Resources Evaluation and Planning Project II, titik lokasi sampel pengamatan harus diletakkan pada peta khusus. Pada masa lalu hal ini telah dilakukan pada beberapa macam survei, tetapi tidak semua peta tanah mempunyai titik pengamatan ini. Juga dalam proyek pengembangan wilayah transmigrasi, sampel pengamatan ini wajib dicantumkan baik untuk pemetaan tanah ataupun untuk lokasi pengamatan air bersih atau hidrologi. Data mengetahui kerapatan ini dapat digunakan untuk menilai keakuratan data peta tematik yang dianalisis. Dengan adanya data ini maka berbagai sistem perhitungan secara statistik dapat dilakukan. Untuk kepentingan pengujian keakuratan data, pemanfaatan statistika spasial akhir-akhir ini berkembang sangat pesat. Umumnya peta tematik baik di Indonesia maupun dunia, tidak menyajikan data lokasi maupun data mental pada peta, yang umum adalah informasi hasil generalisasinya. Padahal dari sudut pengembangan bank data, hal ini sangat diperlukan. Mungkin masih dibutuhkan waktu sampai hal ini terlaksana.

- Relevansi Data

Tidak semua data yang dipakai untuk pengolahan data geografis relevan dengan tujuan yang ingin dicapai, tetapi sering dipilih sebagai pengganti. Data yang diturunkan melalui pengindraan jauh dapat dipakai untuk menduga penggunaan lahan, biomassa, kelembaban atau pengamatan landform yang selanjutnya dipergunakan untuk menduga jenis tanah yang diikuti dengan pengamatan lapangan. Jika hubungan data pengganti dengan variable yang diinginkan telah diketahui dengan baik maka pengganti ini juga menjadi sumber informasi yang baik. Peta-peta yang diinginkan adakalanya dapat dikembangkan dari data lain dengan cara korelasi secara statistik yang lebih murah untuk mendeteksi sifat-sifat yang ada. Misalnya menggunakan analisis regresi, analisis co-spektral atau co-kriging. Dalam semua kasus data yang memakai cara pendugaan biasanya berkualitas lebih rendah dari data yang diinginkan. Pemakai harus waspada terhadap tampilan data yang indah. Bagaimanapun juga, adakalanya lebih efisien memakai data pengganti, karena biaya dan kesulitan mengukur sifat-sifat asli.

- Format

Menurut Burrough [2010] ada 2 macam bentuk data yang penting dalam SIG, yang pertama, kemudahan data disajikan dalam media magnetic dan ditransfer ke suatu sistem computer ke sistem lain. Dalam hal ini termasuk berbagai pertimbangan seperti: media magnetik (disket, data-liune), kenampakan informasi ditulis panjang balok perekam, jumlah jalur (biasanya 9), kecepatan bit per inci (biasanya 800, 1600 atau 6250 BPI), tipe karakter (ASCII, Binary) dan panjang rekaman. Pertimbangan kedua adalah cara data disusun atau struktur data tersebut apakah dinyatakan dengan kode titik-titik, vektor atau raster ? Jika daerah pengamatan tersebut dibuat dalam format raster, berapa ukuran selnya ? Apakah format data yang ada terikat ke sistem data tertentu dan dapat dikonversi ?. banyak sistem SIG sekarang, datanya dapat dikonversi dari suatu format ke format yang lain. Pemahaman format data yang tersedia akan memudahkan pemahaman kemungkinan hasil yang dapat diperoleh

2.3 Semantic Geospasial

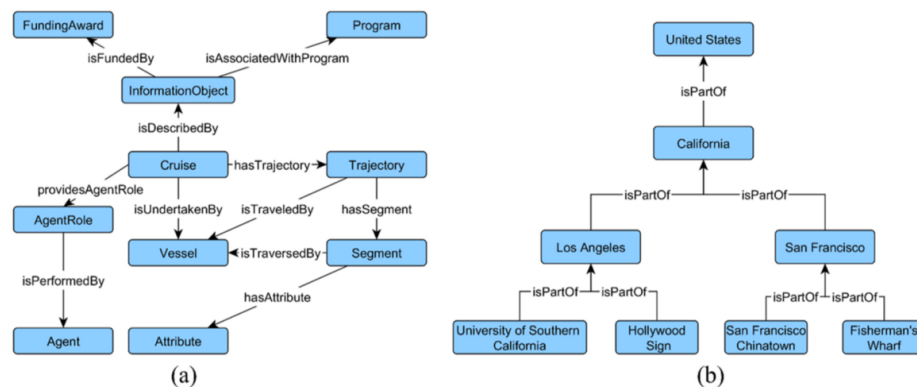
Semantic geospasial adalah subbidang yang diakui dalam GIScience [Agarwal (2005); D.Mark, Egenhofer, Hirle,& Smith 2000], juga melibatkan berbagai bidang penelitian terkait. Kuhn [2005] mendefinisikan geospasial semantic sebagai “memahami GIS, dan menangkap pemahaman ini dalam teori formal”. Kemajuan teknologi komputer dan informasi, terutama Web, telah sangat memudahkan penelitian semantic geospasial, dengan semantic web awalnya diusulkan oleh Berners-Lee, Hendler, dan Lassila [2001], Egenhofer [2002] membayangkan web semantic geospasial yang mampu memahami semantic permintaan geospasial pengguna dan secara otomatis mendapatkan hasil yang relevan.

Interoperabilitas antara Building Information Model (BIM) dan Sistem Informasi Geografis (SIG) adalah salah satunya masalah utama yang dihadapi membangun sistem informasi dan praktisi saat ini. Integrasi ini bertujuan untuk memenuhi peningkatan permintaan untuk analisis konstruksi, aplikasi perencanaan kota, manajemen bencana, kadaster dan tanah air keamanan dan aplikasi lainnya [Bonn,2005]. Aplikasi ini tidak hanya membutuhkan geometri 3D dan informasi tampilan. Sebaliknya, mereka membutuhkan informasi semantik yang kompleks. Model konseptual telah dikembangkan di

dua dunia, BIM dan GIS, berupa model geometris dan semantik. Untuk tujuan visualisasi saja, geometris bersama dengan tampilan dan informasi tekstur cukup untuk merepresentasikan objek spasial 3D. Namun, dibutuhkan model semantik yang berbeda teknik dan aplikasi perencanaan yang membutuhkan query dan analisis yang kompleks [Stadler A dan Kolbe TH,2007]

Semua bahasa di domain berbeda dicirikan oleh sintaksis dan semantiknya. Bagian sintaks, di satu sisi, menjelaskan bagaimana simbol dan kata dikenali dalam bahasa. Ini juga berisi aturan bagaimana bentuk dirumuskan dengan baik kalimat menggunakan simbol yang dikenali. Tidak hanya simbol yang harus disepakati saat berkomunikasi. Namun, aturan sintaks harus didefinisikan dan dipahami dengan baik [Clarendon Press,Oxford,1975]. Dalam hal informasi spasial, kesepakatan antara IFC dan CityGML berarti bahwa masing-masing dari mereka menggunakan bahasa alami yang benar secara tata bahasa dalam arah dan pemetaan spasial verbal yang sesuai dengan prosedur yang diterima geometris di tempat lain. Bagian semantik, di sisi lain, menyangkut tentang makna ekspresi bahasa yang mencerminkan interpretasi objek dan bagian dalam bahasa spasial

Pendekatan utama untuk memungkinkan interoperabilitas semantic adalah mengembangkan ontology, saat belajar di bidang filsafat sebagai hakikat makhluk, ontology dalam semantic geospasial lebih dekat dengan yang ada di ilmu komputer dan bioinformatika, yang berfungsi memformalkan makna konsep dalam cara yang dapat dimengerti mesin [Bittner, Donnelly, & Winter, 2005; Couclelis, 2009; Gruber, 1993;Guarino, 1998; Stevens, Goble, & Bechhofer, 2000]. Dari perspektif struktur data, seorang ontology bisa dianggap grafik dengan konsep sebagai node dan relasi tepi seperti paada Gambar 2.6.



Gambar 2.6: (a) Fragmen Ontology Proyek NSF GeoLink (Adila Krisnadhi, et al., 2015); (b) Bagian dari Gazetteer

Ontologi adalah alat yang sangat penting dalam domain interoperabilitas. Mereka digunakan untuk mendefinisikan dan semantic ambigu dari sistem terminology [Gruber T,1993; Guarino N,1998]. Ontologi terutama digunakan untuk Bahasa komunikasi antara manusia atau komputer.baru-baru ini digunakan untuk informasi spasial dengan menentukan semantic objek spasial dan keterkaitannya.

Di satu sisi, ontology berbasis logika didasarkan pada teori logis yang dikemukakan oleh Copi [1979]. Jenis ontology ini, dari namanya, ditentukan oleh aksioma logis dan definisi untuk mengungkapkan hubungan antara entitas dan kelas.

Ontologi harus dikembangkan sebelum dapat digunakan dalam GIS. Tiga jenis onotologi bisa jadi diidentifikasi dari literatur: ontology tingkat atas, ontology domain, dan ontology design pattern (ODP). Ontology tingkat atas berisi istilah umum (misalnya, isPartOf, edurant,dan perdurant) yang dapat digunakan di seluruh domain, sementara ontology domain memformalkan konsep untuk disiplin tertentu [Ashburner,et al.,2000;Guarino,1997; Rogers & Rektor, 1996]. Ontologi yang digunakan dalam GIScience umumnya dianggap sebagai ontology domain, dan sering disebut onotlogi geografis atau geo-ontologi [F.Fonseca,Camara,& Miguel Monteiro, 2006; Tonai & Kavouras,2004] pola desain ontology dikembangkan berdasarkan aplikasi. Alih-alih mencari kesepakatan di dalam atau di seluruh domain, mereka menganggap kebutuhan umum digunakan Bersama oleh beberapa aplikasi [Gangemi, 2005; Gangemi&Presutti, 2009].

2.4 Artificial Intelligence (AI)

Artificial Intelligence (AI) teknologi dasar untuk membangun sistem pakar. Dalam kecerdasan buatan, para ilmuwan mempelajari penalaran dan mesin, dan pertimbangankan pertanyaan-pertanyaan seperti apakah sebuah mesin dapat dengan tepat dianggapan berasalan atau berpikir, dan apa yang akan dihitung sebagai ujian untuk penalaran.

AI memiliki banyak sub-bidang dan teknologi canggih, untuk contoh teknologi baru sedang dikembangkan secara konstan:

1. Jaringan saraf tiruan: mensimulasikan cara kerja neuron di otak,
2. Pemrosesan Bahasa alami: untuk menghasilkan komputer sistam yang dapat memahami, menerjemahkan, dan berkomunikasi dengan manusia
3. Algoritma genetik: memecahkan masalah dengan analogi dengan evolusi biologis melalui seleksi alam, dll.

Kecerdasan adalah kemampuan untuk memperoleh dan menerapkan pengetahuan, berasalan secara deduktif dan menunjukan kreativitas dan AI didefinisikan sebagai kemampuan untuk melakukan fungsi yang biasanya terkait dengan kecerdasan manusia.

Saat ini, AI diterima secara luas sebagai kumpulan dari Teknik pemrograman canggih mempelajari sifat kecerdasan dengan membangun sistem komputer, dan aplikasi dari wawasan ini dalam memecahkan masalah dunia nyata. Dari ini perspektif, AI berusaha untuk mengembangkan sistem yang mencoba meniru kecerdasan manusia tanpa mengklaim pemahaman tentang proses yang mendasari terlibat.

Kecerdasan buatan (AI) adalah cabang ilmu komputer yang bertujuan untuk mensimulasikan pemikiran proses otak manusia, biasanya melalui penggunaan perangkat lunak [LM Cowardin, PM Arnold, TL Shaffer, 1988]. AI adalah teknologi dasar untuk membangun sistem pakar. Dalam kecerdasan buatan, para ilmuwan mempelajari penalaran dan mesin, dan pertimbangan pertanyaan-pertanyaan seperti apakah sebuah mesin dapat dengan tepat dianggap beralasan atau berpikir, dan apa yang akan dihitung sebagai ujian untuk penalaran.


Dari beberapa metode AI yang terutama digunakan pada GIS tergolong jaringan saraf tiruan (JST) dan logika fuzzy. Jaringan saraf tiruan adalah program komputer yang dirancang untuk memodelkan otak manusia dan kemampuannya untuk mempelajari tugas-tugas penggunaan model matematika.

AI menyediakan Teknik yang cukup canggih untuk GIS sedangkan GIS adalah teknologi yang kuat dengan kumpulan data yang luas dan cakupan yang luas untuk AI. Misalnya logika fuzzy telah berhasil diterapkan untuk masalah spasial yang tidak dapat seperti data pengumpulan, representasi, dan analisis serta klasifikasi tanah, dan citra pengideraan jauh.

2.5 Program Indonesia Sehat dengan Pendekatan Keluarga (PIS-PK)

Pengembangan Program Indonesia Sehat merupakan salah satu program unggulan Pemerintah Indonesia dalam bidang Kesehatan. Salah satu terobosan dari Kementerian untuk mewujudkan Indonesia Sehat adalah Program Indonesia Sehat melalui Pendekatan Keluarga yang dikenal dengan PIS-PK. Saat ini program PIS-PK telah berjalan diseluruh wilayah Indonesia melalui pendataan yang dilakukan oleh setiap Puskesmas di seluruh wilayah Indonesia melalui orientasi keluarga.

Salah satu sarana pendukung PIS-PK adalah system informasi yang berkualitas sehingga proses pendataan maupun intervensi terkait pelaksanaan program menjadi terorganisir dengan baik. Aplikasi PIS-PK telah dikembangkan sejak tahun 2016 dengan berpusat kepada pendataan 12 Indikator PIS-PK di seluruh Indonesia yang diuraikan pada Gambar 2.7.



Indikator Keluarga Sehat	
1	Keluarga mengikuti program KB (keluarga berencana)
2	Ibu hamil memeriksakan kehamilannya (ANC) sesuai standar
3	Bayi mendapatkan Imunisasi lengkap
4	Pemberian ASI eksklusif bayi 0-6 bulan
5	Pemantuan pertumbuhan balita
6	Penderita TB Paru yang berobat sesuai standar
7	Penderita hipertensi yang berobat teratur
8	Penderita gangguan jiwa berat yang diobati
9	Tidak ada anggota keluarga yang merokok
10	Sekeluarga sudah menjadi anggota JKN
11	Mempunyai sarana air bersih
12	Menggunakan jamban keluarga

Gambar 2.7: Indikator Keluarga Sehat (IKS)

Namun saat ini aplikasi PIS-PK dinilai kurang dapat memenuhi seluruh kebutuhan baik dari sisi internal Kementerian Kesehatan maupun Eksternal seperti Pemerintah Daerah. Salah satu yang menjadi kendala dalam aplikasi PIS-PK adalah kesulitan untuk melakukan pengunduhan data laporan terkait 12 indikator secara realtime; maupun proses integrasi dengan layanan aplikasi lain; serta kesulitan dalam melakukan manajemen data dan analisis indikator terutama bila terjadi perubahan seperti pergantian tahun survey maupun perubahan dan penambahan indikator lain seperti pada Gambar 2.8.



Gambar 2.8: Penjabaran Visi & Misi Presiden Menjadi Program Indonesia Sehat

2.5.1 Indeks Keluarga Sehat (IKS)

IKS adalah perhitungan kedua belas indikator keluarga sehat yang perhitungannya diambil dari rekapitulasi data dari 12 indikator yang diambil data dan hasilnya dibagi menjadi 3, yaitu: 1) Keluarga sehat bila IKS > 0,800; 2) Keluarga Pra Sehat, bila IKS = 0,500 - 0,800 3) Keluarga Tidak Sehat, bila IKS < 0,500. Pendataan keluarga dilakukan terhadap seluruh keluarga dalam wilayah kerja Puskesmas dengan menggunakan formulis pengumpulan data untuk Prokesga. Prokesga berisi data tingkat keluarga dan data dari semua individu anggota keluarga tersebut (sebagaimana tercantum dalam Kartu Keluarga). Data yang dicatat minimal data 12 indikator tersebut. Sesuai kondisi daerah, Prokesga dapat dikembangkan sehingga mencakup indikator-indikator lain yang dianggap penting di daerah tersebut [Dinas Kesehatan,2021]. Gambar 2.9 merupakan contoh hasil indeks keluarga sehat pada Kabupaten Ngawi.



Gambar 2.9: Hasil Indeks Keluarga Sehat Kab. Ngawi

2.6 Machine Learning

Menurut [Samuel1959], *Machine learning* atau disebut pembelajaran mesin merupakan bidang studi komputer yang memberikan kemampuan kepada mesin komputer untuk belajar tanpa harus secara eksplisit. Terdapat beberapa kategori dari Machine learning, yaitu:

- Berdasarkan bagaimana cara pelatihan dengan menggunakan supervisi (pelabelan) dari manusia atau tanpa supervisi manusia. Terdiri dari empat kategori yaitu *supervised*, *unsupervised*, *semisupervised*, dan *reinforcement learning*.
- Berdasarkan bagaimana cara metode pembelajaran. Terdiri dari dua kategori, yaitu: *online learning* dan *batch learning*.
- Berdasarkan bagaimana cara perbandingan data. Terdiri dari dua kategori, yaitu: *instance based* dan *model-based learning*.

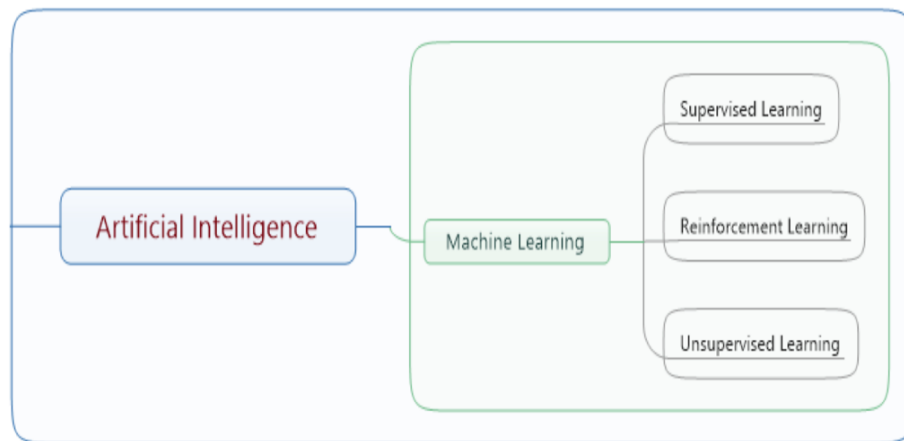
Umumnya machine learning dikategorikan berdasarkan jenis pelatihan yaitu berdasarkan pelabelan supervisi manusia atau tanpa supervisi manusia.

1. ***Supervised Learning***, pada metode ini data yang di input ke dalam suatu algoritma mencakup kategori yang diinginkan atau sering disebut diberikan pelabelan. Metode klasifikasi umumnya menggunakan

jenis pelatihan supervisi atau memprediksi nilai dari sebuah angka. Sebagai contoh adalah penentuan harga mobil berdasarkan fitur dari mobil tersebut misalnya jarak tempuh, umur, merk, dan lainnya yang dikenal dengan predictors. Beberapa algoritma yang menggunakan supervised learning adalah K-Nearest Neighbors, Linear Regression, Logistic Regression, Support Vector Machines (SVMs), Decision Trees and Random Forests, dan Neural networks.

2. ***Unsupervised Learning***, pada metode ini data yang akan diinput ke dalam suatu algoritma tidak melalui kategorisasi atau pelabelan dari manusia, karena seluruh kategorisasi akan dilakukan berdasarkan data yang dimasukan dan akan terbentuk berdasarkan ciri yang bersesuaian. Terdapat beberapa algoritma unsupervised learning di antaranya:
 - (a) *Clustering*, seperti: k-Means, Hierarchical Cluster Analysis (HCA), dan Expectation Maximization
 - (b) *Association rule learning*, seperti: Apriori, dan Eclat
 - (c) Reduksi Dimensi dan Visualisasi, seperti: Principal Component Analysis (PCA), Kernel PCA, Locally-Linear Embedding (LLE), dan t-distributed Stochastic Neighbor Embedding (t-SNE).
3. ***Semisupervised Learning***, pada metode ini jenis pelatihan yang digunakan adalah gabungan metode supervised dan unsupervised learning, sehingga terdapat data yang telah diketahui kategorisasi atau dilabelkan, dan data yang belum dilabelkan. Salah satu algoritma yang menggunakan metode ini adalah Deep Belief Networks (DBNs) dengan komponen yang dikenal dengan Restricted Boltzmann Machines (RBMs). Metode RBM akan dilatih menggunakan data unsupervised, kemudian sistem tersebut akan dioptimalisasi menggunakan teknik supervised.
4. ***Reinforcement Learning***, pada metode ini dikenal istilah agent yang harus mempertimbangkan lingkungan dalam sistem pembelajaran. Sistem akan memilih, melakukan tindakan dan mengukur akibat (rewards) agar terhindar akibat yang negatif (penalties), sehingga diperlukan suatu strategi yang dikenal dengan policy untuk mendapatkan dampak yang positif.

Berdasarkan penjelasan diatas bahwa *machine learning* terbagi menjadi tiga kategori; Supervised Learning, Unsupervised Learning, Reinforcement Learning berkaitan dengan artificial intelligence, berikut dijelaskan dalam Gambar 2.10



Gambar 2.10: Skema Artificial Intelligence dan Machine Learning

2.7 Klasifikasi Teks

Proses klasifikasi teks merupakan suatu cara dalam menemukan sebuah model yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui sebelumnya dari suatu objek. Dua proses yang dilakukan pada pengklasifikasian data, yaitu:

- **Proses *training*.** Training set yang digunakan pada proses ini merupakan data training yang telah memiliki label untuk membangun sebuah mode.
- **Proses *testing*.** Testing set digunakan untuk melakukan prediksi label dan melihat keakuratan model atau fungsi yang akan dibangun pada proses training.

Klasifikasi dokumen adalah pemberian kategori yang telah didefinisikan kepada dokumen yang belum memiliki kategori (Goller, Loning, Will, dan Wolff, 2000). Mengklasifikasi dokumen merupakan salah satu cara untuk mengorganisasikan dokumen. Pengelompokan dengan kategori yang sama dilakukan pada dokumen-dokumen yang memiliki isi yang sama. Hal ini mennja-

dikan sebuah kemudahan bagi orang yang akan melakukan pencarian informasi, karena kategori yang dianggap tidak relevan akan dilewati.

Metode klasifikasi secara otomatis berkembang seiring dengan adanya jumlah dokumen yang meningkat dengan pesat. Dokumen-dokumen yang telah dilakukan klasifikasi sebelumnya dijadikan acuan dalam proses *learning* pada implementasi metode klasifikasi. Kondisi ini memberikan manfaat dalam hal perolehan hasil yang lebih baik dan peghematan waktu kerja (Hadiyono, 2018).

2.8 Metode Klasifikasi

Klasifikasi bisa dikatakan sebagai pembelajaran mesin karena memiliki kemampuan untuk menggunakan pengetahuan yang telah ada sebelumnya untuk menghasilkan penentuan objek baru, keseluruhan klasifikasi terletak pada kemampuan sistem untuk memberi label terhadap objek sesuai dengan kasus yang telah ada tanpa mengubah sistem jika dihadapkan objek baru.

2.8.1 K-nearest neighbors (KNN)

K-Nearest Neighbors Algorithm (KNN) : merupakan sebuah metode yang digunakan dalam penyelesaian masalah klasifikasi sebuah objek dengan melakukan implementasi feature space dimana sebuah objek yang menjadi data training dan dijadikan sebuah model data diberikan pembobotan nilai dan direpresentasikan ke dalam n-dimensional vektor. Kemudian penyelesaian masalah dilakukan dengan mengukur jarak terdekat objek baru dengan model data yang ada di dalam n-dimensional vektor lalu dilakukan proses pemberian sebuah kategori pada objek baru tersebut.

2.8.2 Naïve bayes (NB)

Naïve Bayes merupakan salah metode dari Machine Learning yang memanfaatkan konsep probabilitas dan konsep statistika untuk menghasilkan sebuah prediksi atau variabel baru dalam suatu permasalahan. Algoritma Naïve Bayes pertama kali dikemukakan oleh Thomas Bayes pada tahun yang memiliki tujuan agar bisa memberikan suatu prediksi probabilitas di masa depan

berdasarkan pola-pola serta pengalaman yang terbentuk sebelumnya. Naïve Bayes juga merupakan sebuah algoritma klasifikasi yang berakar pada Teorema Bayes.

2.9 Rangkuman Hasil Penelitian Terkait

Penelitian-penelitian terkait yang berhubungan dengan pengembangan SIG akan disajikan pada Tabel 2.1

Tabel 2.1: Rangkuman Hasil Penelitian Terkait

No	Nama Peneliti, Tahun	Topik	Kelebihan	Keterbatasan
1	V. Voženílek, 2009	Konvergensi antara Ai dan GIS	Jaringan saraf tiruan telah terbukti bermanfaat dalam interpretasi informasi sumber daya spasial	Klasifikasi GIS dalam dunia nyata itu tidak realistis mengharapkan akurasi klasifikasi 100% pada semua kumpulan data
2	Royal institue of technology (KTH), 2010	Building information mdels (BIM)	CityGML (3D objekperkotaan) dan IFC (mewakili model semantik yang sangat detail untuk bangunan) dianggap semantik yang paling menonjol model diGIS dan BIM	Salah satu masalah utama yang dihadapi aplikasi pemodelan kota 3D adalah kurangnya interoperabilitas diantara berbagai BIM dan model GIS.
3	Anggoro Sigitt, A, 2011	Mengembangkan SIG berbasis web untk monitoring banjir	SIG web dapat memberika informasi tentang kerawanan banjir di wilayah DAS Bengawan Solo Hulu.	Pengumpulan data spasial yang cukup sulit untuk mengolahnya.

No	Nama Peneliti, Tahun	Topik	Kelebihan	Keterbatasan
4	Elelma Cerutti, Francesca Noardo, dan Antonia Spano, 2015	GIS dapat menjadi instrumen yang efektif untuk mengelola data architectural heritage, untuk meminta data, bertujuan pelestarian.	Dalam karya ini, beberapa struktur memungkinkan penggunaan informasi semantik CH (seperti ontologi dan bahasa formal) telah digunakan bersama alat yang interoperabilitas data.	Seluruh alur kerja memiliki batasan karena tidak adanya model data standar terintegrasi resmi, khususnya kurangnya standar di akuisisi dan plot fase sejarah dokumentasi architectural heritage
5	Stephanie M. Flectcher-Lartey, 2015	Teknologi GIS di sektor kesehatan masyarakat, dengan aplikasi yang berkaitan dengan pemetaan dan pemahaman penyakit parasit	Teknologi GIS berkonstrinbusi signifikan terhadap pemahaman tresebut ekologi parasit dan asosiasinya dengan penyebaran penyakit, dan intervensi pencegahan, terutama di daerah berkembang	GIS kurang dimanfaatkan beberapa negara di bidang kesehatan masyarakat, sementara keterbatasan sistemik (kurangnya infrastruktur, pelatih-an, pemeliharaan database jangka panjang), dan biayanya terlalu tinggi

No	Nama Peneliti, Tahun	Topik	Kelebihan	Keterbatasan
6	Yingjie HU, 2017	Semantik Geospasial dapat memfasilitasi perencanaan sistem informasi geografis (GIS)	Dapat membangun ontologi, gazetteer, dan data teraut dapat membantu mesin memproses geografs informasi lebih efektif secara otomatis mengekstrak pengetahuan dengan lebih efisien, sehingga gazeteers banyak digunakan di GIR dan semantik tepat untuk mengekstrak dan membedakan nama tempat (<i>geograp infom retrieval</i>)	Data geografi yang besar, teks tidak terstruktur dalam jumlah besar di web, dan cepat pengembangan metode pemrosesan bahasa alami memungkinkan arah penelitain baru di geospasial semantik
7	Kamel Boulos, 2019	<i>Artificial intelligence (AI)</i> dan <i>geografic information system (GIS)</i> , berperan dalam kesehatan dan perawatan kesehatan, karena suatu daerah merupakan bagian integral dari populasi maupun kesehatan individu.	Aplikasi GeoAI ini mempunyai potensi dalam beberapa ilmu dalam kesehatan masyarakat, pengobatan presisi, dan menjadi kota sehat pintar yang didukung oleh <i>Internet of Things</i>	Kurangnya label data penelitian untuk algoritma AI, secara khusus diawasi pembelajaran melibatkan memprediksi label atau respon masing-masing titik data

No	Nama Peneliti, Tahun	Topik	Kelebihan	Keterbatasan
8	Nasim Tohidi dan Rustam B. Rustamov, 2020	Aplikasi algoritma ML untuk proses citra satelit atau GIS	Teknologi ML dan GI berhasil diterapkan untuk pemantauan dan observasi konsekuensi pembangunan megacity . Penggabungan ML dan GIS menawarkan mekanisme potensial untuk mengurangi biaya analisis informasi spasial dengan mengurangi jumlah waktu yang dihabiskan untuk interpretasi data	Konsep tidak tepat yang digunakan dengan besar volume data geofafis, harganya mahal
9	Issues of Healthcare Planning and GIS: A Review, 2020	Mengidentifikasi terkait geomatika kesehatan berdasarkan pemahaman hubungan spasial	GIS alat yang tepat untuk pengambilan keputusan spasial dalam kesehatan masyarakat	Data sensus yang digunakan untuk pemodelan epidemi harus berupa kualitas dan akurasi tinggi agar sejalan dengan persyaratan teknik GIS

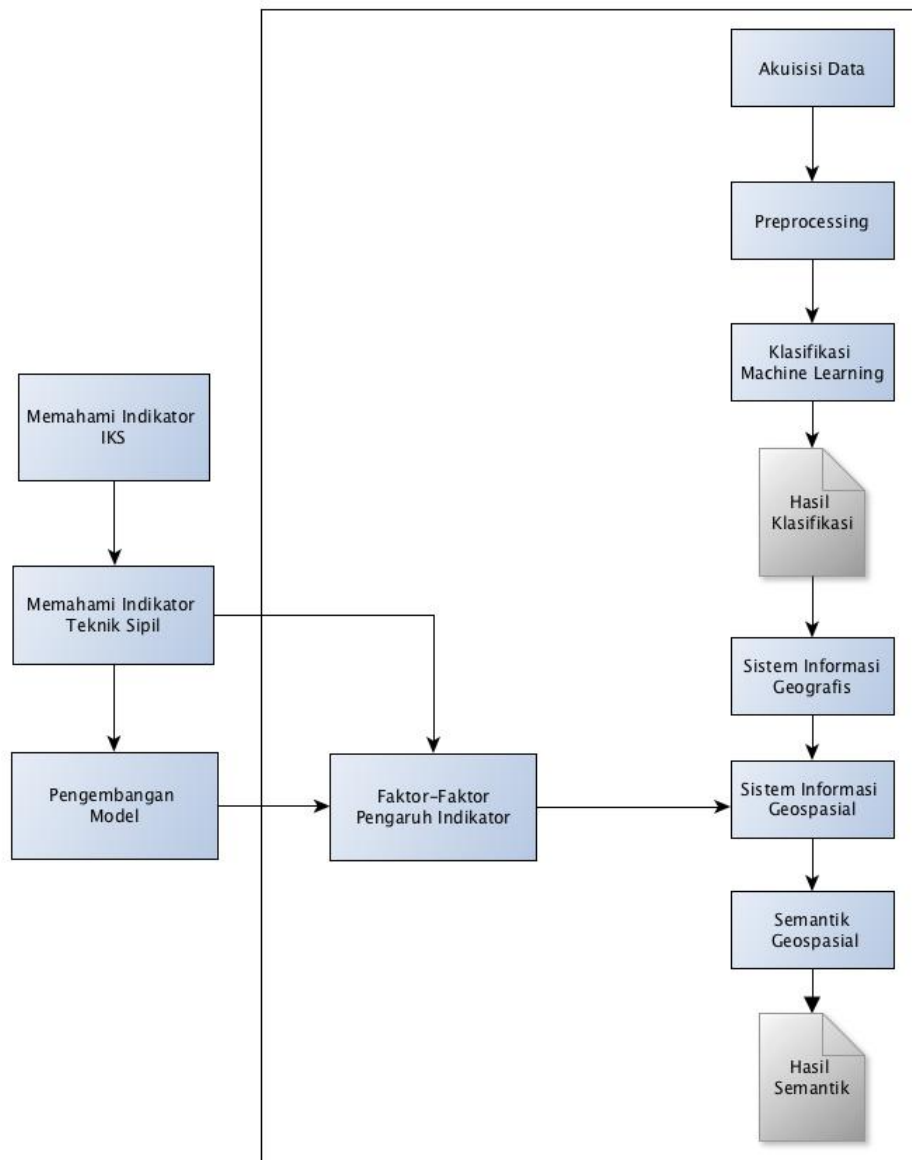
Bab 3

Metode Penelitian

3.1 Tahapan Penelitian

Penelitian ini berusaha mengembangkan model SIG tentang data spasial kesehatan masyarakat menggunakan semantik AI. Pada gambar 3.1 terlihat tahapan penelitian untuk mengembangkan SIG data spasial kesehatan masyarakat terdiri dari 3 tahapan, yaitu:

1. Memahami Indikator IKS
2. Memahami Indikator IKS yang berhubungan dengan Teknik Sipil
3. Pengembangan Model, yang terdiri dari tahapan:
 - (a) Akuisisi Data
 - (b) *Preprocessing*
 - (c) Klasifikasi *Machine Learning*
 - (d) Sistem Informasi Geografis
 - (e) Sistem Informasi Geospasial
 - (f) Semantik Geospasial



Gambar 3.1: Tahapan Penelitian

3.2 Memahami Indikator IKS

Penelitian ini melakukan analisis pada 12 indikator IKS seperti diuraikan pada Tabel 3.1.

Tabel 3.1: Indikator Keluarga Sehat

No	Indikator
1	Keluarga mengikuti program KB (keluarga berencana)
2	Ibu hamil memeriksakan kehamilannya (ANC) sesuai standar
3	Bayi mendapatkan imunisasi lengkap
4	Pemberian ASI eksklusif lengkap
5	Pemantauan pertumbuhan balita
6	Penderita TB paru yang berobat sesuai standar
7	Penderita hipertensi yang berobat teratur
8	Penderita gangguan jiwa berat yang diobati
9	Tidak ada anggota keluarga yang merokok
10	Sekeluarga sudah menjadi anggota KJKN
11	Mempunyai sarana air bersih
12	Menggunakan jamban keluarga

3.3 Memahami Indikator IKS yang Berhubungan dengan Teknik Sipil

Penelitian ini melakukan analisis pada 12 indikator IKS yang berkaitan dengan ilmu Teknik sipil, diperoleh terdapat 2 indikator yang berkaitan dengan teknik sipil yaitu mempunyai sarana air bersih dan menggunakan jamban keluarga. Berikut adalah Tabel 3.2 pada beberapa wilayah berkategori keluarga prasehat dan keluarga tidak sehat.

Tabel 3.2: Indikator IKS Teknik Sipil

Indikator	Hidrologi	Geotek	Transportasi	Struktur
Mempunyai sarana air bersih	x	x		x
Mempunyai sarana air bersih	Desa Payaman, 2016. 79% atau 196 KK yang tidak memiliki sarana pembuangan air limbah yang sesuai dengan persyaratan kesehatan. Menyalurkan limbah cair pada kawasan rumah tangga diperlukan sarana berupa saluran pembuangan air limbah ataupun sumur resapan. sehingga tidak mencemari air tanah, tidak menimbulkan sarang nyamuk dan tikus, tidak menimbulkan bau	Daerah Cipanas, tidak memiliki jamban dikarenakan daerahnya dikelilingi sungai besar, dan letak geografis yang kurang yang memungkinkan		Permukiman kumuh di bantaran sungai Deli dan faktor manusia yang menggunakan bantaran sungai untuk mendirikan rumah tinggal dan tempat pembuangan sampah telah menyempitan sungai

Indikator	Hidrologi	Geotek	Transportasi	Struktur
Mempunyai sarana air bersih	Berdasarkan data dari Puskesmas Kampung Baru Kec. Medan Maimun, 2011 diare merupakan penyakit terbesar (879 kasus), berdasarkan survei sungai Deli di kelurahan Sukaraja terlihat keruh, berwarna coklat kekuningan dan terlihat adanya bangkai limbah industri domestik, sementara di hilir sungai sebagian besar masyarakat menggunakan air sungai Deli utk MCK			
Mempunyai sarana air bersih	Kab Purwantoro, tidak memiliki sumber air yang bersih, sehingga nilai yg memiliki air bersih 44,8% dan yang tidak memiliki sebanyak 55,2%. Sumur milik salah satu desa di Kecamatan Purwantoro digunakan dibandingkan dengan 1 sumur digunakan atau diambil oleh air pada 10 rumah tangga. Akses ke PDAM belum masuk ke wilayah tersebut.			

Indikator	Hidrologi	Geotek	Transportasi	Struktur
Mempunyai sarana air bersih	Desa Sungai Terap Kec Kumpeh Ulu Kab Muaro Jambi. Kurangnya ketersediaan air karena 1 sumur dapat dipakai 4-8 KK ini menyebabkan sumur cepat kering. Makanya solusinya pemerintah setempat dapat menyediakan sarana air bersih berupa PDAM dan memperbanyak sumur gali gratis dengan kedalaman >10m agar musim kemarau tidak ada lagi sumur yang kering			
Mempunyai sarana air bersih	Kab. Sampang JaTim. Dari hasil survei bahwa daerah sampang kondisi secara geografis sebagian besar kering sehingga utk ketersediaan air masih sangat memperhatikan, sehingga masyarakat mengandalkan air hujan			Infrastrukturnya sangat minim, diharapkan kedepannya PUPR dapat membuat- an jamban, PDAM, dan jamban

Indikator	Hidrologi	Geotek	Transportasi	Struktur
Mempunyai sarana air bersih	Aliran sungai Winogo (Jogja). Memiliki pengaruh yang buruk terhadap sanitasi dimana sungai Winogo, dimana di daerah tersebut adanya peternakan babi dan tumpukan sampah yang ada disisi sungai.			
Menggunakan jamban keluarga		x	x	x

Indikator	Hidrologi	Geotek	Transportasi	Struktur
Menggunakan jamban keluarga	Desa Sungai Terap Kec Kumpeh Ulu Kab Muaro Jambi. Masih ada 22 responden yang tidak memiliki jamban, dikarenakan pengaruhnya ketersediaan air yang sangat penting, sehingga masyarakat BABS di tepi sungai.			
Menggunakan jamban keluarga	Desa Modelomo memiliki data menggunakan jamban 6,44%, Rumah Tangga yang belum memakai fasilitas BAB 90,30%, BABS 93,36%, dan diare 26,81%. Diharapkan kedepannya perlunya pemahaman tentang pemakaian jamban utk mendukung PIS-PK			

Indikator	Hidrologi	Geotek	Transportasi	Struktur
Menggunakan jamban keluarga	Desa cikunir merupakan desa dengan cakupan akses sanitasi (jamban sehat) yang rendah.yang tidak memiliki jamban sebesar 17,6% sehingga dapat menyebabkan atau menyebarkan penyakit dan mengotori lingkungan pemukiman			
Menggunakan jamban keluarga	Faktor lingkungan di Desa Cintaraja pada tahun 2018 mengalami kenaikan kasus yaitu penyakit diaere yang disebabkan dari kurang layaknya jamban, saluran limbah umum dan sumur resapan			

Indikator	Hidrologi	Geotek	Transportasi	Struktur
Menggunakan jamban keluarga	Angka diare lebih tinggi sebesar 66% pada anak-anak dari keluarga yang melakukan buang air besar di sungai atau selokan dibandingkan mereka pada rumah tangga dengan fasilitas toilet pribadi dan septik tank. (Kemenkes RI, 2016). Dinas Kesehatan Kabupaten Bantul, penderita diare di Kabupaten Bantul tergolong tinggi. (Dinkes Bantul, 2017			
Menggunakan jamban keluarga	Desa Bandung Kecamatan Kebumen menunjukkan bahwa indikator akses terhadap jamban sehat masih rendah, karena masih ada warga yang buang air besar di sungai . Kurangnya pengetahuan dan sikap masyarakat yang masih Buang Air Besar (BAB) di sungai.			
Menggunakan jamban keluarga	Desa Pelakar Jaya, tidak menggunakan jamban ketika BAB dan BAK, karena kondisi tempat tinggal yang kurang layak			

Pada Tabel 3.2 menjelaskan bahwa dalam indikator keluarga sehat dapat

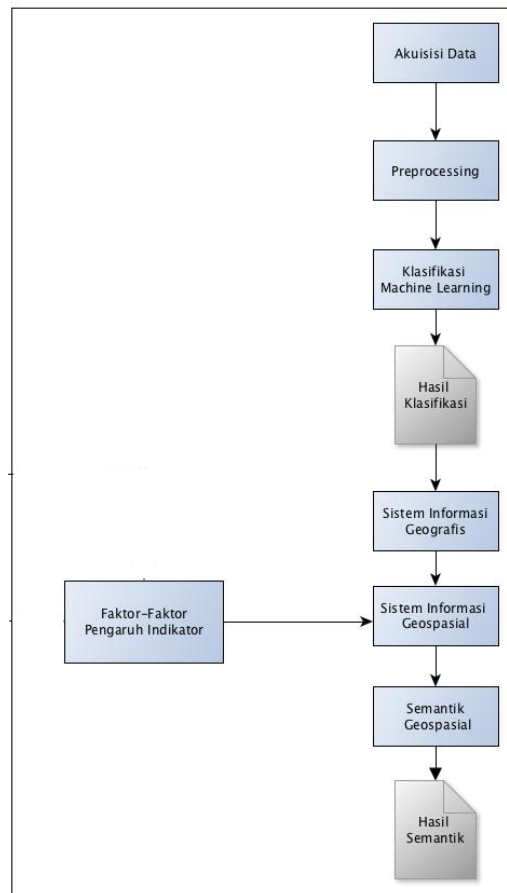
dihubungkan pada ilmu Teknik sipil, yang dimana dilihat dari beberapa kondisi yaitu:

- Kurangnya sarana air bersih, kurangnya sumber air (PDAM)
- Adanya pencemaran aliran sungai, dimana tidak adanya pembuangan air limbah
- Pembangunan jamban yang kurang layak

Sehingga dapat mengakibatkan wilayah tersebut termasuk dalam kriteria keluarga prasehat dan keluarga tidak sehat.

3.4 Pengembangan Model

Pada tahapan pengembangan model, terdiri dari 6 proses di dalamnya, yaitu akuisisi data, *preprocessing*, klasifikasi *Machine Learning*, Sistem Informasi Geografis, Sistem Informasi Geospasial dan Semantik Geospasial seperti pada Gambar 3.2.



Gambar 3.2: Tahap Pengembangan Model

3.4.1 Akuisisi Data

Proses akuisisi data adalah tahap pengumpulan data yang merupakan tahap awal dari penelitian ini. Data yang dibutuhkan yaitu dokumen IKS, dokumen IKS diperoleh dari Kementrian Kesehatan. Proses pengumpulan data IKS oleh Kementrian Kesehatan

3.4.2 Preprocessing Data

Tahapan kedua dalam penelitian ini *preprocessing* data tahap melakukan persiapan data sebelum melakukan proses klasifikasi dengan *machine learning*. Tahap *preprocessing* pada penelitian ini adalah:

1. Tahap pemeriksaan data. Pada tahap ini dilakukan pemeriksaan terhadap data yang akan digunakan, data -data yang tidak dibutuhkan

akan dihapus, dan data yang tidak sesuai akan diperbaiki sesuai dengan format data.

2. Penghapusan data. Data yang tidak digunakan akan dihapus, dan hanya data yang sesuai dengan fitur akan masuk pada proses klasifikasi
3. Pelabelan data. Data diberikan label berdasarkan jenis klasifikasi,
 - Label 1 menyatakan kategori keluarga sehat,
 - Label 2 menyatakan kategori Keluarga Pra Sehat,
 - Label 3 menyatakan kategori keluarga Tidak Sehat

3.4.3 Klasifikasi Machine Learning

Proses klasifikasi dilakukan untuk memprediksi data IKS menjadi tiga kriteria, yaitu:

- Keluarga sehat
- Keluarga Pra Sehat
- Keluarga Tidak Sehat

Metode klasifikasi yang digunakan adalah *support vector machine* berbasis *machine learning*.

3.4.4 Sistem Informasi Geografis

Pada tahap ini akan dilakukan pembuatan peta sistem informasi geografis berdasarkan klasifikasi tahap sebelumnya. Pada peta ini akan terlihat wilayah-wilayah mana yang masuk ke dalam keluarga sehat, keluarga pra sehat, dan keluarga tidak sehat.

3.4.5 Sistem Informasi Geospasial

Pada tahap ini diambil dari indikator yang berhubungan dengan Teknik sipil, dan melihat sistem informasi geografis pada wilayah berkategori, keluarga sehat, keluarga prasehat, keluarga tidak sehat.

Dari hasil sistem informasi geospasial akan diartikan atau diterjemahkan lebih detail dengan semantik geospasial wilayah mana saja yang termasuk kategori keluarga sehat, keluarga prasehat, keluarga tidak sehat sehingga diperoleh hasil yang dapat memberikan solusi agar kedepannya seluruh masyarakat Indonesia memahami pentingnya kesehatan.

Penelitian ini dilakukan selama dua tahap dengan target penyelesaian di akhir 2021, Tabel 3.3 menggambarkan rencana penelitian yang akan dilakukan dalam menyusun disertasi ini.

Kegiatan	Bulan ke-																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Penyusunan Proposal																		
Pengajuan Proposal																		
Pengerjaan Penelitian																		
Publikasi Ilmiah / Seminar																		
Pengembang an Atas Saran Seminar																		
Penyusunan Desertasi																		
Sidang Desertasi																		

Bab 4

Hasil dan Pembahasan

4.1 Akuisisi Data

Tahapan pertama pada penelitian ini adalah akuisisi data. Akuisisi data adalah proses pengumpulan data-data yang diambil dari Kementerian Kesehatan Republik Indonesia. Data yang diperoleh berupa laporan data IKS sampai dengan Januari tahun 2020.

Tabel 4.1 adalah deskripsi data laporan IKS sampai dengan Januari 2020

Tabel 4.1: Data Laporan IKS Januari 2020

Jumlah Data	81.900
Jumlah Provinsi	34
Jumlah Kecamatan	6.390
Jumlah Kelurahan	55.344

Data IKS yang diperoleh seperti pada Gambar 4.1 terdiri dari field berikut:

- Kode
- Nasional
- nama_provinsi
- provinsi
- nama_kota

- kota
- nama_kecamatan
- kecamatan
- kelurahan
- nama_kelurahan
- jumlah_kk
- jumlah_kk_sehat
- jmlh_pra_sehat
- jmlh_tdk_sehat
- IKS Inti
- Indikator 1 (KB)
- Indikator 2 (Ibu Bersalin di faskes)
- Indikator 3 (Imunisasi)
- Indikator 4 (ASI)
- Indikator 5 (Pertumbuhan Balita)
- Indikator 6 (TB)
- Indikator 7 (Hipertensi)
- Indikator 8 (Gangguan Jiwa)
- Indikator 9 (Tidak ada anggota keluarga yang merokok)
- Indikator 10 (Sekeluarga sudah menjadi anggota JKN)
- Indikator 11 (Mempunyai Sarana air bersih)
- Indikator 12 (menggunakan jamban keluarga)

49

kode	nasional	nama_provinsi	provinsi	nama_kota	kota	nama_kecamatan	kecamatan	kelurahan	nama_kelurahan	jumlah_kk	jumlah_kk_sah	jumlah_rhgs	jumlah_rhgs_tdk_sah	ltsa	ltsa_tdk	
		INDONESIA	ACEH	*1						955452	239598	599920	115994	0	25	
		INDONESIA	ACEH	*1						28551	6932	18926	2693	0	24	
01		INDONESIA	ACEH	*1		ACEH SELATAN	*01	-	-	748	104	579	65	0.14	0.14	
01	01.2001	INDONESIA	ACEH	*1		ACEH SELATAN	*01	BAKONGAN	*01	2001	62	26	60	0.26	0.26	
01	01.2002	INDONESIA	ACEH	*1		ACEH SELATAN	*01	BAKONGAN	*01	2002	130	17	107	0.8	0.13	
01	01.2003	INDONESIA	ACEH	*1		ACEH SELATAN	*01	BAKONGAN	*01	2003	142	15	89	40	0.71	
01	01.2004	INDONESIA	ACEH	*1		ACEH SELATAN	*01	BAKONGAN	*01	2004	104	69	10	0.28	0.28	
01	01.2015	INDONESIA	ACEH	*1		ACEH SELATAN	*01	BAKONGAN	*01	2015	1	0	0	0	0	0
01	01.2016	INDONESIA	ACEH	*1		ACEH SELATAN	*01	BAKONGAN	*01	2016	209	19	17	0.22	0.22	
01	01.2017	INDONESIA	ACEH	*1		ACEH SELATAN	*01	BAKONGAN	*01	2017	72	10	59	3	0.14	0.14
02		INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		2477	548	1860	271	2.22	0.19	
02	02.2001	INDONESIA	ACEH	*1		ACEH SELATAN	*01	TAJAR HARAPAN	*01	15	15	0	0	0.06	0.06	
02	02.2002	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KRUIKUN BATTEE	*01	14	4	9	1	0.29	0.1	
02	02.2003	INDONESIA	ACEH	*1		ACEH SELATAN	*01	PASIE ASAHAN	*01	11	0	9	0	0.29	0.1	
02	02.2004	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		11	0	9	0	0.29	0.1	
02	02.2005	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		228	60	138	30	0.26	0.3	
02	02.2006	INDONESIA	ACEH	*1		ACEH SELATAN	*01	JAMBO MANYANG	*01	11	1	10	0	0.26	0.1	
02	02.2007	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		101	39	60	2	0.30	0.1	
02	02.2008	INDONESIA	ACEH	*1		ACEH SELATAN	*01	LUMU PURUT	*01	212	55	146	11	0.26	0.3	
02	02.2009	INDONESIA	ACEH	*1		ACEH SELATAN	*01	PULO KAMBERO	*01	218	28	168	2	0.26	0.3	
02	02.2010	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		216	51	149	16	0.24	0.3	
02	02.2012	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KAMPUNG PATA	*01	286	30	172	64	0.11	0.1	
02	02.2013	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		1	0	1	0	0.24	0.1	
02	02.2016	INDONESIA	ACEH	*1		ACEH SELATAN	*01	LEMBA SAMPANG LUE	*01	85	28	49	8	0.33	0.1	
02	02.2017	INDONESIA	ACEH	*1		ACEH SELATAN	*01	SIANG GEERINGENG	*01	75	19	53	3	0.26	0.1	
02	02.2018	INDONESIA	ACEH	*1		ACEH SELATAN	*01	PASIE KUALA	*01	27	0	27	0	0.26	0.1	
02	02.2019	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KEDUA PADANG	*01	124	17	83	24	0.14	0.1	
02	02.2020	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KUTAPALAN	*01	1	1	0	0	0.26	0.1	
02	02.2021	INDONESIA	ACEH	*1		ACEH SELATAN	*01	GUNUNG PUDUNG	*01	146	23	99	24	0.16	0.1	
02	02.2022	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		143	33	99	11	0.23	0.1	
02	02.2023	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET UTARA*02		21	25	14	0	0.26	0.1	
03		INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		2678	693	1751	232	2.66	0.2	
03	03.2001	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		325	101	217	7	0.31	0.1	
03	03.2002	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		2002	57	11	46	0.23	0.1	
03	03.2003	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		165	43	107	15	0.26	0.1	
03	03.2004	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		179	54	119	15	0.26	0.1	
03	03.2005	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		112	31	77	7	0.26	0.1	
03	03.2006	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		143	42	97	4	0.26	0.1	
03	03.2007	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		303	175	133	33	0.37	0.1	
03	03.2008	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		134	23	95	16	0.17	0.1	
03	03.2009	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KEDUA KANDANG	*01	40	5	27	8	0.12	0.1	
03	03.2010	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		165	39	126	15	0.26	0.1	
03	03.2011	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		17	5	12	0	0.26	0.1	
03	03.2012	INDONESIA	ACEH	*1		ACEH SELATAN	*01	PASIE MERAPAS	*01	58	13	38	7	0.22	0.1	
03	03.2013	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		16	13	15	0	0.26	0.1	
03	03.2014	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		309	65	208	36	0.21	0.1	
03	03.2015	INDONESIA	ACEH	*1		ACEH SELATAN	*01	KULET SELAT*03		21	21	0	0	0.26	0.1	
03	03.2016	INDONESIA	ACEH	*1		ACEH SELATAN	*01	PASIE LEMBANG	*01	107	21	72	14	0.14	0.1	
03	03.2017	INDONESIA	ACEH	*1		ACEH SELATAN	*01	INDRA DAMAI	*01	150	28	107	15	0.19	0.1	
04		INDONESIA	ACEH	*1		ACEH SELATAN	*04	LABUHAN HA'DA		2386	612	1586	163	2.46	0.2	
04	04.2001	INDONESIA	ACEH	*1		ACEH SELATAN	*04	LABUHAN HA'DA		211	48	141	24	0.22	0.2	

Gambar 4.1: Data IKS Setiap Wilayah (a)

[illegible]

Gambar 4.2: Data IKS Setiap Wilayah (b)

Indikator 6 (TB)				Indikator 7 (Hipertensi)				Indikator 8 (ODGJ)				Indikator 9 (SAB)				Indikator 10 (SAB)			
prosentase	sigma	y	z	sigma	prosentase	sigma	y	z	sigma	prosentase	sigma	y	z	sigma	prosentase	sigma	y	z	sigma
32.85	7803	1592	37.66	16807	13018	16487	38.42	4811	1391	94272	42.45	30661	50896	31	9235	80322	9224	17	
37.11	400	678	27473	41.85	2287	3165	23089	44.65	192	238	28121	40.21	11008	11043	0	84.91	27097	1454	0
46.15	0	0	0	92	100	18	0	74	100	1	0	91	36.96	34	58	0	88.91	91	0
0	0	0	0	130	90	9	1	1207	0	0	4	128	25.38	33	877	0	100	130	0
66.67	4	2	138	73.33	11	4	129	0	0	0	144	29.17	42	102	0	100	144	0	0
50	1	1	102	100	16	0	88	0	0	0	104	17.31	18	86	0	89.04	103	1	0
0	0	0	1	0	0	0	1	0	0	0	1	100	1	0	0	100	1	0	0
80	1	1	293	66.67	4	2	198	0	0	0	205	12.68	26	179	0	86.81	204	1	0
0	0	3	89	21.05	4	15	53	0	0	0	72	30.56	22	50	0	83.06	67	5	0
33.33	20	40	2417	58.04	166	120	2191	52	13	12	2482	34.29	849	1028	0	85.53	2394	123	0
100	2	0	83	90	1	1	83	0	0	0	85	30.59	28	59	0	82.84	79	6	0
0	0	1	13	0	0	0	14	100	1	0	13	28.57	4	10	0	82.86	13	1	0
0	0	0	11	0	0	0	11	0	0	0	11	8.09	1	10	0	81.82	9	2	0
20	1	4	145	18.18	4	18	128	0	0	1	149	26.67	55	85	0	82.67	139	11	0
80	3	3	222	62.5	5	3	220	33.33	1	2	225	35.96	82	146	0	87.57	222	6	0
0	0	0	59	100	3	0	58	0	0	0	61	26.07	22	39	0	70.49	43	18	0
80	1	1	99	37.88	25	41	35	66.67	2	1	88	43.56	44	57	0	87.03	88	3	0
80	2	2	208	55.56	5	4	203	100	1	0	211	31.13	66	146	0	88.58	209	3	0
0	0	3	115	66.7	23	10	85	0	0	4	114	27.12	32	86	0	88.31	116	2	0
40	4	6	206	83.33	15	3	198	0	0	1	215	26.85	58	158	0	88.54	215	1	0
0	0	3	253	100	30	0	236	100	1	0	285	27.62	74	192	0	86.5	262	4	0
0	0	0	176	90	9	1	166	0	0	1	175	29.55	52	124	0	87.73	172	4	0
0	0	2	83	66.67	2	1	82	0	0	0	85	82.84	45	40	0	82.84	79	6	0
25	1	3	71	50	5	5	65	100	1	0	74	37.33	28	47	0	87.33	73	2	0
0	0	1	88	50	1	1	85	0	0	0	87	61.19	41	28	0	84.03	63	4	0
87.14	4	3	117	51.02	25	24	75	100	2	0	122	33.87	42	82	0	85.32	81	43	0
0	0	0	143	100	3	0	143	0	0	1	142	39.18	56	87	0	100	143	0	0
40	2	3	141	100	3	0	143	66.67	2	1	143	43.15	63	83	0	86.58	141	9	0
0	0	0	143	100	3	0	143	0	0	1	142	39.18	56	87	0	100	143	0	0
0	0	3	178	100	7	0	174	100	2	0	179	20.59	55	126	0	86.9	179	2	0
44.38	55	89	2552	23.94	266	845	1565	32.65	16	33	2827	35.5	950	1726	0	83.57	2504	172	0
41.18	7	10	308	32.28	61	128	136	50	2	2	321	62	169	156	0	89.69	324	1	0
75	6	2	173	22.58	14	48	119	20	1	4	176	27.07	49	132	0	86.13	174	7	0
40	2	3	180	27.69	18	47	100	28.57	2	5	158	47.27	78	87	0	85.15	157	8	0
30.77	4	9	186	16.67	14	70	95	25	1	3	175	37.99	68	111	0	82.74	186	13	0
57.14	4	3	105	17.34	5	24	83	0	0	0	112	38.11	40	72	0	84.64	106	6	0
80	1	1	141	29.73	11	28	106	80	1	1	141	25.57	38	105	0	100	143	0	0
35	7	13	913	27.4	62	161	310	87.14	4	3	505	37.15	108	305	0	81.74	489	44	0
63.84	7	4	123	26.42	14	39	81	0	0	2	132	34.33	46	88	0	81.79	123	11	0
20	1	4	35	21.43	3	11	26	0	0	1	39	45	18	22	0	77.5	31	9	0
80	3	2	90	25.93	7	20	28	0	0	1	54	30.91	17	38	0	84.55	52	3	0
0	0	3	14	50	5	5	7	100	1	0	16	41.18	7	10	0	100	17	0	0
80	4	1	53	20.69	6	23	29	0	0	0	58	46.55	27	31	0	84.63	55	3	0
100	4	0	57	46.43	17	15	33	0	0	0	61	26.07	22	39	0	85.25	52	9	0
14.29	1	6	302	6.09	17	108	194	42.86	3	4	302	25.87	79	220	0	82.86	286	23	0
60	3	2	102	29.73	11	26	70	0	0	1	106	24.3	26	81	0	83.16	89	18	0
0	0	2	105	5.86	3	50	54	25	1	3	103	33.64	36	71	0	86.39	101	6	0
20	1	4	145	21.43	12	84	94	0	0	3	147	21.33	32	118	0	82.67	139	11	0
24.1	30	63	2303	60.52	187	122	2077	75	30	10	2446	41.41	868	1388	0	84.7	2021	365	0
12.5	1	7	203	54.55	12	10	189	100	1	0	210	44.08	93	118	0	78.67	166	45	0

Gambar 4.3: Data IKS Setiap Wilayah (c)

Indikator 11 (Jamban Sehat)				Indikator 12 (JKN)			
prosentase_id	sigma_y_id_11	sigma_t_id_11	sigma_n_id_11	prosentase_id	sigma_y_id_12	sigma_t_id_12	sigma_n_id_12
84.2	804147	150864	471	80.99	772988	181467	1027
90.4	25794	2740	17	75.51	21517	6977	57
92.51	692	56	0	89.41	667	79	2
100	92	0	0	100	91	0	1
97.69	127	3	0	93.08	121	9	0
70.83	102	42	0	69.44	100	44	0
94.23	98	6	0	82.52	85	18	1
100	1	0	0	100	1	0	0
98.05	201	4	0	96.59	198	7	0
98.61	71	1	0	98.61	71	1	0
89.99	2229	248	0	74.31	1840	636	1
98.82	84	1	0	88.24	75	10	0
100	14	0	0	100	14	0	0
100	11	0	0	63.64	7	4	0
84	126	24	0	64.67	97	53	0
92.98	212	16	0	76.75	175	53	0
98.36	60	1	0	96.72	59	2	0
100	101	0	0	97.03	98	3	0
97.17	206	6	0	88.15	186	25	1
87.29	103	15	0	90.68	107	11	0
89.35	193	23	0	91.67	198	18	0
80.45	214	52	0	37.59	100	166	0
95.45	168	8	0	70.45	124	52	0
83.53	71	14	0	89.41	76	9	0
98.67	74	1	0	98.67	74	1	0
100	67	0	0	97.01	65	2	0
87.1	108	16	0	98.39	122	2	0
94.44	17	1	0	72.22	13	5	0
77.4	113	33	0	45.89	67	79	0
95.1	136	7	0	60.14	86	57	0
83.43	151	30	0	53.59	97	84	0
89.01	2381	294	1	89.31	2390	286	0
97.23	316	9	0	94.77	308	17	0
95.58	173	8	0	96.13	174	7	0
76.97	127	38	0	88.48	146	19	0
91.06	163	16	0	92.74	166	13	0
90.18	101	11	0	95.54	107	5	0
86.01	123	20	0	93.01	133	10	0
88.56	472	61	0	94.18	502	31	0
96.27	129	5	0	80.6	108	26	0
80	32	8	0	72.5	29	11	0
89.09	49	6	0	87.27	48	7	0
94.12	16	1	0	94.12	16	1	0
89.47	51	6	1	79.31	46	12	0
81.97	50	11	0	65.57	40	21	0
84.14	260	49	0	85.76	265	44	0
80.37	86	21	0	77.57	83	24	0
85.05	91	16	0	85.98	92	15	0
94.67	142	8	0	84.67	127	23	0
92.82	2211	171	4	85.99	2031	331	24
95.73	202	9	0	70.95	149	61	1

Gambar 4.4: Data IKS Setiap Wilayah (d)

4.2 Preprocessing

Tahapan selanjutnya setelah proses akuisisi data kedua dalam penelitian ini yaitu *preprocessing* data. Tahap ini melakukan persiapan data sebelum melakukan proses klasifikasi dengan *machine learning*.

4.2.1 Tahap pemeriksaan data

Pada tahap ini dilakukan pemeriksaan terhadap data yang akan digunakan, dilakukan kesesuaian fitur data dan filter data yang dibutuhkan.

4.2.2 Penghapusan data

Setelah dilakukan pemeriksaan data, maka diperoleh data-data yang dianggap tidak sesuai dengan fitur dan dilakukan penghapusan data tersebut. Dari hasil filter juga diperoleh data yang dianggap tidak dibutuhkan sehingga data tersebut dihapus.

4.2.3 Pelabelan data

Setelah data yang dihasilkan dari tahap sebelumnya dianggap sudah memenuhi kesesuaian dengan fitur, maka selanjutnya dilakukan pelabelan data. Data diberikan label berdasarkan jenis klasifikasi seperti pada Tabel 4.2.

Tabel 4.2: Label Data

Label	Kategori
1	Keluarga sehat
2	Keluarga pra sehat
3	Keluarga tidak sehat

Daftar Pustaka

- Agarwal, P. (2005). Ontological considerations in giscience. *International Journal of Geographical Information Science*, 19(5):501–536.
- Boulos, M. N. K., Peng, G., and VoPham, T. (2019). An overview of geoi applications in health and healthcare. *International Journal of Health Geographics*, 18(1):1–9.
- Goller, C., Löning, J., Will, T., and Wolff, W. (2000). Automatic document classification. *Informationskompetenz-Basiskompetenz in der Informationsgesellschaft*, page 145.
- Hadiyono, A. (2018). *Metode Analisis Berbasis NLP dan DBN untuk Pengukuran Dampak Media Online*. PhD thesis, Universitas Gunadarma.
- Sigit, A. A., Priyono, P. P., and Andriyani, A. A. (2011). Aplikasi sistem informasi geografis (sig) berbasis web untuk monitoring banjir di wilayah das bengawan solo hulu. *Semantik*, 1(1).