LAPORAN AKHIR TAHUN

Penelitian Produk Terapan

IDENTIFIKASI WILAYAH SPASIAL KEKUATAN POLITIK CALON WALIKOTA MANADO TAHUN 2015 BERDASARKAN SUARA PEMILIH MENGGUNAKAN ALGORITMA EKSPEKTASI MAKSIMISASI

Tahun ke- 1 dari rencana 2 tahun

TIM PENGUSUL

Ketua Tim : Winsy Ch D Weku, S.Si, M.Cs (0009127607)

Anggota Tim : Altien Rindengan, S.Si, M.Kom (0027047403)

UNIVERSITAS SAM RATULANGI

November 2017

HALAMAN PENGESAHAN
PENELITIAN PRODUK TERAPAN

Judul Penelitian	:	Identifikasi Wilayah Spasial Kekuatan Politik Calon Walikota Manado tahun 2015 Berdasarkan Suara Pemilih Menggunakan Algoritma Ekspetasi Maksimisasi
Kode/Nama Rumpun Ilmu	:	123/Ilmu Komputer
Ketua Peneliti		
a. Nama Lengkap	:	WINSY CHRISTO DEILAN WEKU S.Si
b. NIDN	:	0009127607
c. Jabatan Fungsional	:	Lektor
d. Program Studi	:	Matematika
e. Nomor HP/Surel	:	08114306846/winsyweku@gmail.com
Anggota Peneliti (1)		
a. Nama Lengkap	:	ALTIEN JONATHAN RINDENGAN S.Si,M.Kom
b. NIDN	:	0027047403
c. Perguruan Tinggi	:	Universitas Sam Ratulangi
Lama Penelitian Keseluruhan	:	2 tahun
Usulan Penelitian Tahun ke-	:	1
Biaya Penelitian Keseluruhan	:	Rp 144,900,000.00
Biaya Penelitian		
- diusulkan ke DRPM	:	Rp 74,900,000.00
- dana internal PT	:	Rp 0
- dana institusi lain	:	Rp 0 /in kind tuliskan:

Kota Manado, 27-05-2016

Mengetahui,
Dekan FMIPA UNSRAT

(Prof. Dr. Benny Pinontoan, M.Sc)
NIP/NIK 196606041995121001

Ketua Peneliti

(WINSY CHRISTO DEILAN WEKU S.Si)
NIP/NIK 197612092000121001

Menyetujui,
Ketua LPPM UNSRAT

(Prof. Dr. Ir. Inneke F.M. Rumengan, M.Sc)
NIP/NIK 195711051984032001

Ringkasan

PILWAKO adalah sistem voting khusus dan unik yang berada di Indonesia diikuti oleh semua lapisan masyarakat yang sudah memiliki syarat untuk mengikuti pemilihan. Masalahnya adalah bagaimana menentukan daerah mana yang bisa memberikan hasil yang signifikan dan terukur untuk dianalisis, sehingga bisa digunakan untuk PILWAKO berikutnya. Selain itu terdapat beberapa wilayah yang sulit dijangkau dan diketahui atas ketidakikutsertaan mereka sebagai warganegara yang bersdia memberikan suaranya. Tujuan penelitian ini adalah membuat model peta berdasarkan hasil klastering menggunakan algoritma EM.Dengan kata lain kita diharapkan bisa menentukan daerah atau wilayah yang signifikan yang masih dapat berubah atau kecenderungan yang tidak stabil. Sebuah algoritma ekspektasi ekspektasi clustering probabilistik (EM) untuk mengelompokkan data spasial untuk mengidentifikasi area politik PILWAKO dalam pemilihan Mayor Manado pada tahun 2015.

Penelitian ini menerapkan algoritma clustering hirarkis Gaussian yang dimodifikasi dan algoritma EM untuk model campuran Gaussian yang dimodifikasi dengan kemungkinan penambahan sebuah istilah noise Poisson. Penelitian dilakukan atas beberapa tahap, yaitu: pengumpulan data, pembuatan peta, pengembangan program komputer menggunakan program R, implementasi program, analisis hasil, pembahasan, penulisan penelitian, pemasukan di seminar nasional, pemasukan di jurnal.

Hasilnya adalah bahwa ada tiga daerah bergerombol yang menggunakan algoritma EM. Ketiga wilayah tersebut terdistribusi dengan baik. Di antara daerah berkerumun ada beberapa daerah yang mengalami ketidakpastian. Hal ini mengindikasikan bahwa kawasan ini cenderung berubah dalam memilih kandidat di Pilwako berikutnya.

PRAKATA

Laporan penelitian akhir tahun ini dibuat dan dilaporkan sebagai bagian pertanggung jawaban atas tuntutan sebagai seorang peneliti. Penelitian ini dilaksanakan di Laboratorium Komputer, jurusan Matematika, FMIPA, Universitas Sam Ratulangi, Manado, dengan judul Identifikasi Wilayah Spasial Kekuatan Politik Calon Walikota Manado Tahun 2015 Berdasarkan Suara Pemilih Menggunakan Algoritma Ekspektasi Maksimisasi. Penelitian dilakukan selama 5 bulan, dimana sudah dimulai dari bulan Mei 2107 dan berakhir dbulan September 2017.

Penelitian pemetakan dan klastering ini dilakukan berdasarkan prinsip simulasi dan visualisasi, dimana pemetakan menggunakan peta model hasil klasterisasi sedangkan klastering menghasilkan model klaster yang digunakan dalam memetakan.

Pada kesempatan ini, peneliti ingin mengucapkan terima kasihbsedalam-dalamnya kepada beberapa pihak yang membantu terwujudnya penelitian ini:

1. DIKTI yang telah menyetujui dan memberikan bantuan dana kepada peneliti untuk melaksanakan penelitian ini.
2. Prof. Dr. Ir. Inneke F.M. Rumengan, M.Sc sebagai Ketua Lembaga Penelitian, Pengabdian dan Pemberdayaan Masyarakat UNSRAT atas dukungan yang diberikan.
3. Prof. Dr. Benny Pinontoan, M.Sc sebagai Dekan FMIPA UNSRAT yang telah menyetujui dan membantu terlaksananya penelitian ini.
4. Anggota peneliti yang telah memberikan waktu, tenaga dan pikiran dalam melakukan penelitian dan melakukan pelaporan.
5. Berbagai pihak yang telah membantu mewujudkan penelitian ini.

Penelitian ini masih jauh dari sempurna, oleh karena itu kritik dan saran sangat diharapkan gun memperbaiki kualitas penelitian dan laporan ini. Semoga laporan penelitian ini dapat diterima dengan baik.

Manado, November 2017

Peneliti

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

PEMILIHAN UMUM DI INDONESIA

Proses PILWAKO di Indonesia adalah perjalanan panjang yang harus mengakomodasi kepentingan politik dan keinginan masyarakat, kepentingan pusat dan daerah, serta kepentingan nasional dan internasional (lihat Hutapea [8], Arfa'i [9]). Implementasi PILWAKO dilakukan serentak untuk menghindari praktik penipuan dan kandidat harus mengikuti masyarakat. Tidak seperti eksekusi sebelumnya yang dilakukan melalui mekanisme pengangkatan oleh pihak berwenang, PILWAKO dilakukan untuk mengakomodasi keinginan masyarakat untuk memilih pemimpin yang lebih transparan dan lebih dikenal. Hal ini dimaksudkan untuk menghindari penolakan masyarakat luas yang tidak mengenal sistem hirarkis.

Pemilihan langsung adalah proses pemilihan Gubernur Kepala Daerah yaitu pemilihan Gubernur dan Wakil Gubernur, tingkat Kabupaten yaitu pemilihan Bupati dan Wakil Bupati dan Tingkat Kota yaitu pemilihan Walikota dan Wakil Walikota yang dilaksanakan secara bersamaan atau bersamaan secara bersamaan secara langsung dan demokratis oleh orang, yaitu menjalankan kedaulatan rakyat. Pemerintah dan DPR sepakat bahwa jadwal pilwako / pilkada akan dibagi menjadi tiga gelombang, yaitu pada tahun 2015, 2017, dan 2018 dan seterusnya sampai Pemilukada secara bersamaan pada tahun 2027. Secara umum, pilwako / pilkada secara bersamaan hadir sebagai sarana penguatan di Indonesia. Jauh dari paling sedikit ada tiga hal yang harus dijawab dari kedatangan pilkada secara bersamaan; Pertama, untuk menciptakan pemilihan yang efisien dan efektif; Kedua, memperkuat keterwakilan antara masyarakat dan kepala daerah; Ketiga, menciptakan pemerintahan daerah yang efektif dan efisien.

Dalam penelitian ini akan dilakukan penelitian klasifikasi daerah pada pemilihan kota Manado sebagai calon walikota pada tahun 2015 dengan menggunakan Gaussian Mixture Model (algoritma EM) untuk mengidentifikasi area clustering. Setiap daerah diharapkan memiliki probabilitas untuk berkumpul dalam satu kelompok tertentu, yang diperkirakan dengan memaksimalkan vektor probabilitas parameter distribusi.

BAB II

TINJAUAN PUSTAKA

Klasterisasi

Di beberapa bidang sains, terutama statistik dan optimasi atau penelitian operasi dan bidang terapan lainnya, masalah utamanya adalah bagaimana mengoptimalkan dengan memaksimalkan atau meminimalkan fungsi beserta variasinya. Karena masalah optimasi ini mungkin mengasumsikan beberapa tipe yang berbeda, masing-masing memiliki karakteristik tersendiri, banyak teknik telah dikembangkan untuk menyelesaikannya. Teknik ini sangat penting dalam data mining dan area penemuan pengetahuan karena dapat digunakan sebagai dasar untuk metode yang paling kompleks dan hebat.

Salah satu tekniknya adalah Maximum Likelihood dan tujuan utamanya adalah menyesuaikan model statistik dengan kumpulan data yang spesifik, memperkirakan parameter yang tidak diketahui sehingga fungsi yang dapat menggambarkan semua parameter dalam dataset. Dengan kata lain, metode ini akan menyesuaikan beberapa variabel model statistik dari dataset atau distribusi yang diketahui, sehingga model dapat "menggambarkan" setiap sampel data dan memperkirakan yang lain.

Disadari bahwa pengelompokan dapat didasarkan pada model probabilitas untuk menutupi nilai yang hilang. Ini memberikan wawasan tentang kapan data harus sesuai dengan model dan telah menghasilkan pengembangan metode pengelompokan baru seperti Expectation Maximization (EM) yang didasarkan pada prinsip Maximum Likelihood dari variabel yang tidak diamati dalam model campuran hingga.

Dalam makalah ini kami mengusulkan menggunakan teknik klasifikasi berbasis probabilistik untuk dua dataset untuk mengidentifikasi area clustering. Kami melakukan perbandingan yang luas dari indeks yang disebutkan di atas bersamaan dengan algoritma clustering Expectation-Maximization pada sejumlah kumpulan data yang banyak digunakan, dan membuat analisis sederhana dari hasil eksperimen.

Algoritma EM dapat digunakan untuk mencocokkan dan mengelompokkan kumpulan data berdasarkan kesamaan objek yang diamati. Namun, karena kita menggunakan Gaussian bola, skalar varian digunakan di tempat matriks kovariansi. Probabilitas sebelumnya untuk setiap Gaussian adalah fraksi titik dalam cluster yang didefinisikan oleh Gaussian. Parameter ini dapat diinisialisasi dengan memilih secara acak sarana Gaussian, atau dengan menggunakan keluaran algoritma EM untuk pusat awal. Algoritma konvergen pada solusi optimal lokal dengan iteratif memperbarui nilai untuk sarana dan varians.

Gaussian Finite Mixture Models

Misalkan $x = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ merupakan sampel n pengamatan independen yang terdistribusi secara independen. Distribusi setiap pengamatan ditentukan oleh fungsi kepadatan probabilitas melalui model campuran model G yang terbatas, yang mengambil bentuk berikut (lihat Scrucca dan Fop [4]).

$$f(x_i;\Psi)=\sum_{k=1}^{G} \pi_k f_k(x_i;\theta_k) \quad (1)$$

dimana $\boldsymbol{\Psi} = \{\pi_1, \ldots, \pi_{G-1}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_G\}$ merupakan parameter dari mixture model, $f_k(\boldsymbol{x}_i; \boldsymbol{\theta}_k)$ adalah komponen ke k komponen kepadatan $\boldsymbol{x}_i$ dengan vektor $\boldsymbol{\theta}_k$, $(\pi_1, \ldots, \pi_{G-1})$ adalah bobot campuran atau probability (sedemikian hingga $\pi_k > 0$, $\sum_{k=1}^{G} \pi k = 1$), dan G adalah banyaknya komponen mixture.

Dengan asumsi bahwa G adalah tetap, parameter model campuran Ψ biasanya tidak diketahui dan harus diperkirakan. Fungsi log-likelihood yang sesuai dengan persamaan (1) diberikan oleh $\ell(\Psi;x1,\ldots,xn)=\sum_{i=1}^{n} log\ (f(x_i;\Psi))$. Maksimalisasi fungsi log-likelihood secara langsung sangat rumit, sehingga estimator likelihood maksimum (MLE) dari model campuran hingga biasanya diperoleh melalui algoritma EM (lihat Dempster, Laird dan Rubin [2], MacLahlan and Peel [5].

Dalam pendekatan berbasis model untuk pengelompokan, masing-masing komponen dari kepadatan campuran yang terbatas biasanya dikaitkan dengan kelompok atau cluster. Sebagian besar aplikasi berasumsi bahwa semua kepadatan komponen timbul dari keluarga distribusi parametrik yang sama, walaupun hal ini tidak menjadi masalah pada umumnya. Model yang populer adalah Gaussian Mixture Model (GMM), yang mengasumsikan distribusi Gaussian (multivariat) untuk setiap komponen, yaitu $fk\ (x;\ \theta k) \sim N(\mu k, \boldsymbol{\Sigma} k)$. Dengan demikian, cluster adalah ellipsoidal, berpusat pada vektor mean μk, dan dengan fitur geometris lainnya, seperti volume, bentuk dan orientasi, ditentukan oleh matriks kovarians Σk. Parameter parameter dari matriks kovarians dapat diperoleh dengan cara dekomposisi eigen dari bentuk $\sum k=\lambda k D k A k D^{\top} k$, di mana λk adalah skalar yang mengendalikan volume ellipsoid, Ak adalah matriks diagonal yang menentukan bentuk kontur kepadatan. dengan det (Ak) = 1, dan Dk adalah matriks ortogonal yang menentukan orientasi ellipsoid yang sesuai (lihat Banfield and Raftery [6], Celeux dan Govaert [7].

Dalam satu dimensi, hanya ada dua model: E untuk varians yang sama dan V untuk varians yang bervariasi. Dalam pengaturan multivariat, volume, bentuk, dan orientasi kovarian dapat dibatasi sama atau bervariasi antar kelompok.

Model Terbaik

Pilihan model yang tersedia dalam paket R mclust dirangkum dalam Tabel 1. Dalam satu dimensi, hanya ada dua model: E untuk varians yang sama dan V untuk variasi varians. Di lebih dari satu dimensi, pengidentifikasi model mengkodekan karakteristik geometris model. Sebagai contoh, EVI menunjukkan sebuah model di mana volume dari semua kelompok sama (E), bentuk gugus dapat bervariasi (V), dan orientasinya adalah identitas (I). Cluster pada model ini memiliki kovarian diagonal dengan orientasi

sejajar dengan sumbu koordinat. Parameter yang terkait dengan karakteristik yang ditunjuk oleh E atau V ditentukan dari data. Model 'terbaik' dapat diperkirakan dengan model pas dengan parameterisasi yang berbeda.

Tabel 1: Parameterisasi model campuran Gaussian multivariat tersedia dalam mclust. [3]

Identifier	Model	#Covariance parameters	Distribution
EII	λI	1	Spherical
VII	$\lambda_k I$	G	Spherical
EEI	λA	D	Diagonal
VEI	$\lambda_k A$	G + (d-1)	Diagonal
EVI	λA_k	1 + G(d-1)	Diagonal
VVI	$\lambda_k A_k$	Gd	Diagonal
EEE	λDAD^T	d(d+1)/2	Ellipsoidal
EEV	$\lambda D_k A D_k^T$	1+(d-1)+G[d(d-1)/2]	Ellipsoidal
VEV	$\lambda_k D_k A D_k^T$	G+(d-1)+G[d(d-1)/2]	Ellipsoidal
VVV	$\lambda_k D_k A_k D_k^T$	G[d(d+1)/2]	Ellipsoidal

Pengenal model menggunakan tiga huruf untuk mengkodekan karakteristik geometris kode: volume, bentuk, dan orientasi. E berarti sama dan V berarti bervariasi di seluruh komponen atau kelompok; Saya mengacu pada matriks identitas dalam menentukan bentuk atau orientasi dan merupakan kasus khusus E. Di kolom yang diberi label '# Covariance parameters', d menunjukkan dimensi data, dan G menunjukkan jumlah komponen campuran. Jumlah parameter untuk masing-masing model dapat diperoleh dengan menambahkan parameter Gd untuk parameter mean dan G-1 untuk proporsi pencampuran.

Algoritma EM

Algoritma EM adalah metode clustering yang tidak diawasi, yaitu, tidak memerlukan fase pelatihan, berdasarkan model campuran. Ini mengikuti pendekatan iteratif, suboptimal, yang mencoba untuk menemukan parameter dari distribusi probabilitas yang memiliki kemungkinan atribut maksimum.

Secara umum, input algoritma adalah kumpulan data (x), jumlah total cluster (M), kesalahan yang diterima untuk berkumpul (e) dan jumlah iterasi maksimum. Untuk setiap iterasi, pertama-tama dijalankan E-Step (E-xpectation), yang memperkirakan probabilitas masing-masing titik dimiliki oleh masing-masing cluster, diikuti oleh M-step (M-aximization), yang mengestimasi ulang vektor parameter dari distribusi probabilitas masing-masing kelas. Algoritma selesai ketika parameter distribusi menyatu atau mencapai jumlah maksimum iterasi. (lihat Dempster, Laird dan Rubin [2])

Inisialisasi

Setiap kelas j, kelas M (atau cluster), dibentuk oleh vektor parameter (θ), disusun oleh mean (μ_j) dan oleh matriks kovariansi (P_j), yang mewakili fitur distribusi probabilitas Gaussian (Normal) digunakan untuk mengkarakterisasi entitas yang teramati dan tidak teramati dari kumpulan data x.

$$\theta(t) = \left(\mu_j(\mathrm{t}), P_j(\mathrm{t})\right), j = 1, \dots, M$$

Pada saat awal (t = 0) implementasi dapat menghasilkan secara acak nilai awal mean (μ_j) dan matriks kovarians (P_j). Algoritma EM bertujuan untuk memperkirakan vektor parameter (θ) distribusi sebenarnya.

E-Step

Langkah ini bertanggung jawab untuk memperkirakan probabilitas masing-masing elemen masing-masing cluster ($P(C_j|x_k)$). Setiap elemen disusun oleh vektor atribut (x_k). Tingkat relevansi poin dari masing-masing cluster diberikan oleh kemungkinan atribut masing-masing elemen dibandingkan dengan atribut elemen lain dari cluster C_j.

$$P(C_j|x) = \frac{|\sum_j(t)|^{-\frac{1}{2}} e^{n_j} P_j(t)}{\sum_{k=1}^{M} |\sum_j(t)|^{-\frac{1}{2}} e^{n_j} P_k(t)}$$

M-Step

Langkah ini bertanggung jawab untuk memperkirakan parameter distribusi probabilitas masing-masing kelas untuk langkah selanjutnya. Pertama dihitung mean (μ_j) kelas j yang diperoleh melalui rata-rata semua titik dalam fungsi tingkat relevansi masing-masing titik.

$$\mu_j(t+1) = \frac{\sum_{k=1}^{N} P(C_j|x_k)\, x_k}{\sum_{k=1}^{N} P(C_j|x_k)}$$

Untuk menghitung matriks kovariansi untuk iterasi berikutnya diterapkan Teorema Bayes, yang menyiratkan bahwa P(A|B) = P(B|A) * P(A) P(B), berdasarkan probabilitas kondisional kejadian kelas .

$$\sum_j (t+1) = \frac{\sum_{k=1}^{N} P(C_j|x_k)\,(x_k - \mu_j(t))(x_k - \mu_j(t))}{\sum_{k=1}^{N} P(C_j|x_k)}$$

Probabilitas terjadinya setiap kelas dihitung melalui mean probabilitas (Cj) dalam fungsi tingkat relevansi masing-masing titik dari kelas.

$$P_j(t+1) = \frac{1}{N} \sum_{k=1}^{N} P(C_j|x_k)$$

Atribut mewakili vektor parameter θ yang mengkarakterisasi distribusi probabilitas setiap kelas yang akan digunakan pada iterasi algoritma berikutnya.

BAB III
TUJUAN DAN MANFAAT PENELITIAN

Setiap penelitian memiliki tujuan dan manfaatnya masing-masing, demikian juga dengan penelitian ini yang dapat digambarkan sebagai berikut.

Tujuan khusus dari penelitian ini adalah

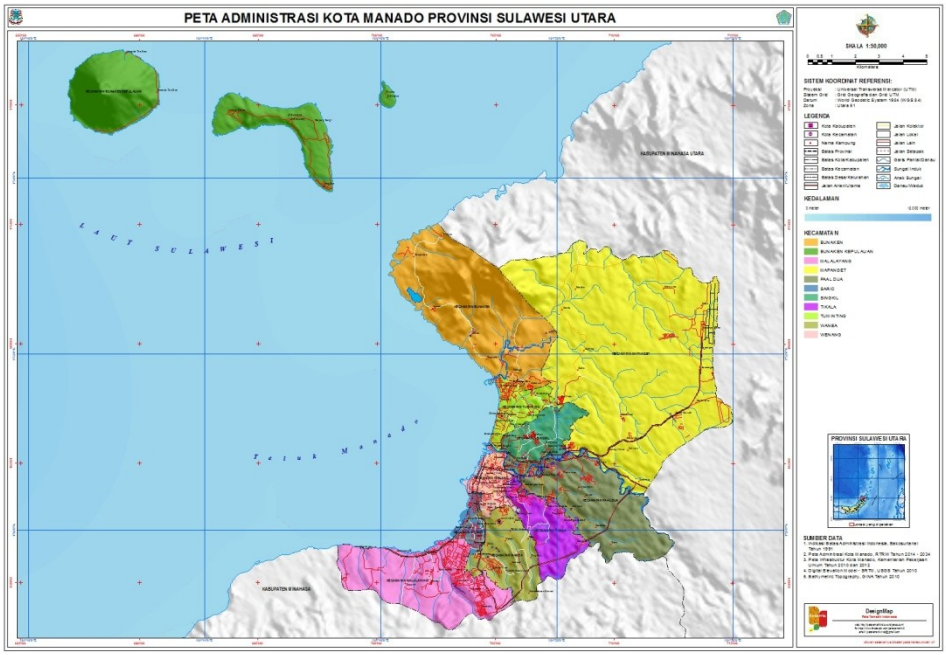
- memetakkan peta kekuatan politik diwilayah Manado menggunakan prinsip klaster.
- mengelompokkan wilayah-wilayah yang memiliki kesamaan dalam menentukan pilihan mereka berdasarkan wilayah mereka masing-masing.
- menentukan wilayah mana saja yang berada dalam definisi "floating voters" sehingga dianggap berpeluang untuk berpindah ke kelompok yang lain.

Manfaat dari penelitian ini adalah

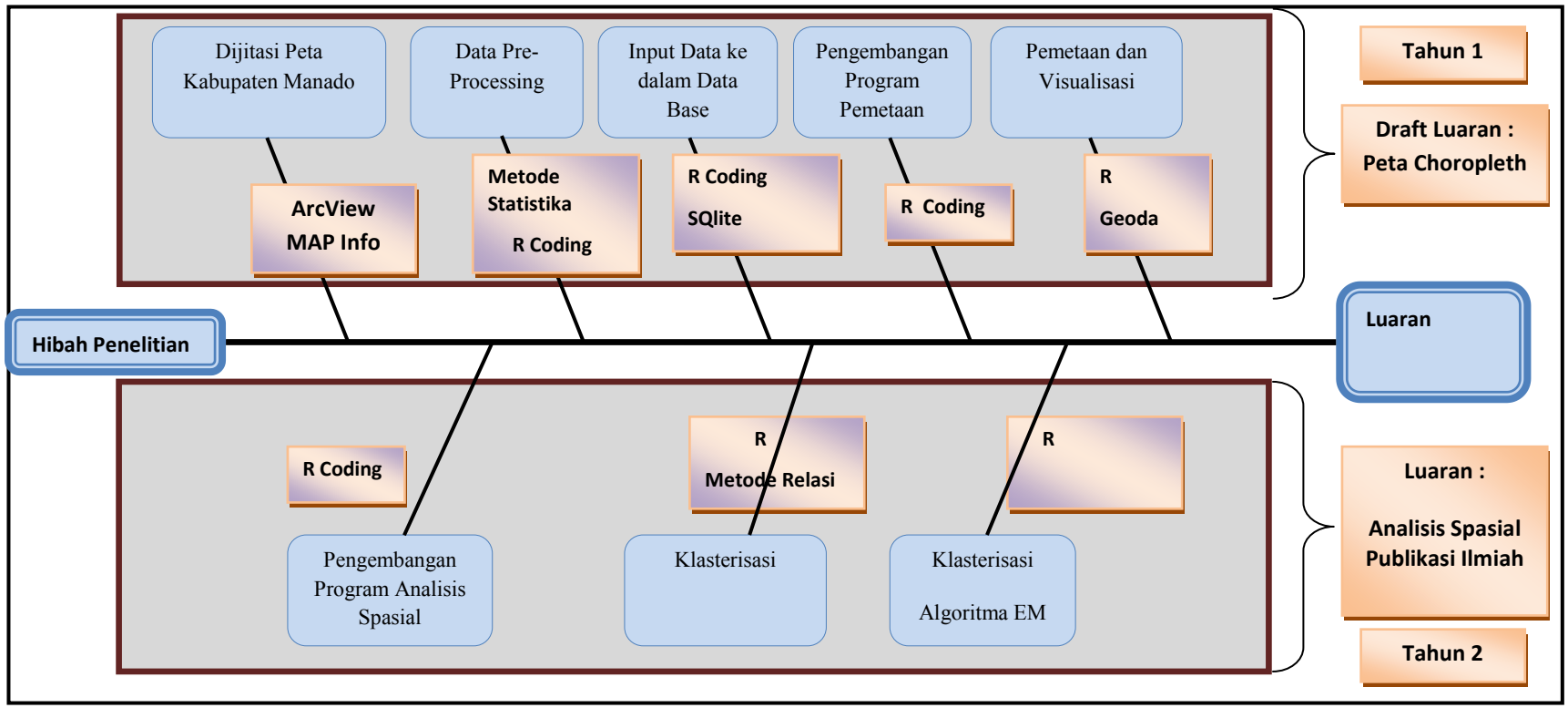
- mendapatkan peta kekuatan politik setiap peserta calon walikota Manado pada tahun 2015.
- mengetahui wilayah-wilayah mana saja yang potensial dan memiliki keinginan yang kuat dalam menyalurkan aspirasi mereka serta menentukan calon yang mereka sukai.
- mengetahui wilayah-wilayah yang termasuk dalam "uncertainty areas" (wilayah yang tidak pasti dalam menentukan pilihannya) sehingga pilihan mereka bisa saja berubah setiap saat tergantung dengan kondisi yang terjadi di lapangan.

BAB IV
METODOLOGI PENELITIAN

Penelitian ini dilakukan di Kota Manado yang memiliki 87 desa sebagai wilayah administrasi terkecil. Metode penelitian yang lengkap adalah seperti pada gambar tulang ikan berikut.



Gambar 1. Wilayah administrasi Kota Manado sebagai lokasi penelitian

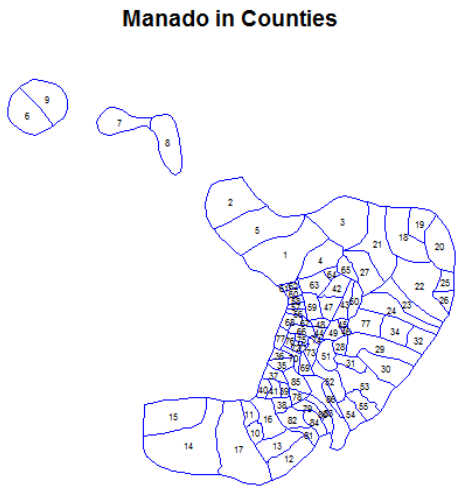


Gambar 2. Diagram Penelitian Fish Bone

BAB V.

HASIL DAN LUARAN YANG DICAPAI

Studi ini mencakup analisis geospasial dari dunia nyata: geo-demografis. Ukuran cluster dipilih berdasarkan model. Sebagai alat komputasi yang dipilih R. Penyajian data peta spasial dilakukan oleh ArcView.



DESA_NO >	DESA
001	MOLAS
002	TONGKEINA
003	PANDU
004	BAILANG
005	MERAS
006	MANADO TUA DUA
007	ALUNG BANUA
008	BUNAKEN
009	MANADO TUA SATU
010	BATUKOTA
011	BAHU
012	WINANGUN 1
013	WINANGUN 2
014	MALALAYANG SATU TIMUR
015	MALALAYANG DUA
016	KLEAK
017	MALALAYANG SATU BARAT
018	KIMA ATAS
019	MAPANGET BARAT
020	LAPANGAN
021	BENGKOL
022	PANIKI BAWAH
023	KAIRAGI DUA
024	KAIRAGI SATU
025	PANIKI 1
026	PANIKI 2
027	BUHA
028	DENDENGAN LUAR
029	PAAL DUA
030	RANOMUT
031	DENDENGAN DALAM
032	MALENDENG
033	KAIRAGI WERU
034	PERKAMIL
035	TITIWUNGEN SELATAN
036	TITIWUNGEN UTARA
037	SARIO UTARA
038	RANOTANA
039	SARIO KOTA BARU
040	SARIO TUMPAAN
041	SARIO
042	SINGKIL 2
043	KOMBOS BARAT
044	KARAME
045	TERNATE TANJUNG
046	TERNATE BARU
047	SINGKIL 1
048	WAWONASA
049	KETANG BARU
050	KOMBOS TIMUR
051	TIKALA ARES
052	BANJER
053	TIKALA BARU
054	TAAS
055	PAAL 4
056	SINDULANG SATU
057	SINDULANG DUA
058	BITUNGKARANGRIA
059	KAMPUNG ISLAM
060	MAASING
061	TUMUMPA 2
062	TUMUMPA 1
063	TUMINTING
064	MAHAWU
065	SUMOMPO
066	PINAESAAN
067	ISTIQLAL
068	CALACA
069	TELING BAWAH
070	BUMI BERINGIN
071	MAHAKERET BARAT
072	MAHAKERET TIMUR
073	TIKALA KUMARAKA
074	KOMO LUAR
075	LAWANG IRUNG
076	WENANG UTARA
077	WENANG SELATAN
078	WANEA
079	BUMI NYIUR
080	PAKOWA
081	KAROMBASAN SELATAN
082	KAROMBASAN UTARA
083	TELING ATAS
084	RANOTANA WERU
085	TANJUNG BATU
086	TINGKULU

Gambar 3. Batas Administratif Manado

Paket R yang digunakan adalah MCLUST-v5.1 yang dikembangkan oleh Chris Fraley dan Adrian Raftery, tersedia dalam repositori CRAN. MCLUST adalah perangkat lunak yang mencakup fitur berikut: pemodelan campuran normal (EM); Inisialisasi EM melalui pendekatan clustering hirarkis; memperkirakan jumlah kelompok berdasarkan Kriteria Informasi Bayesian (BIC); dan menampilkan, termasuk plot ketidakpastian dan proyeksi dimensi.

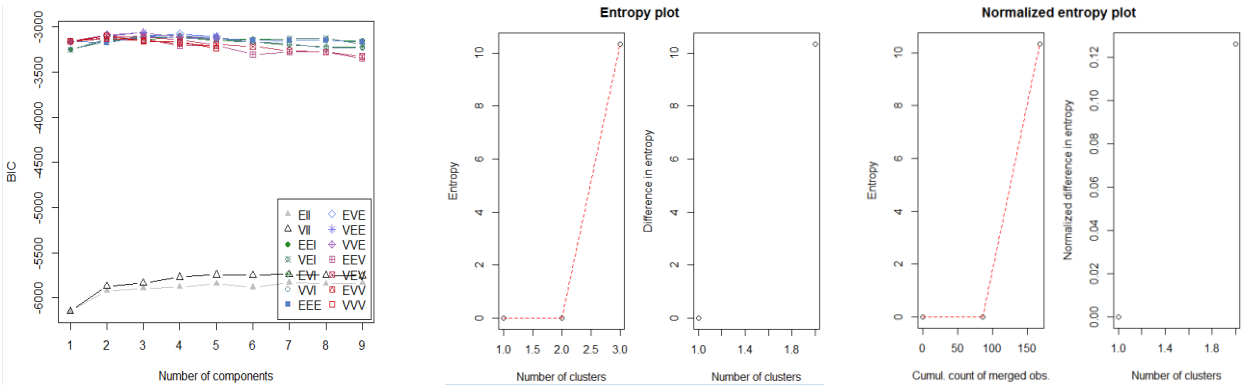
Contoh analisis geodemografi kami diamati dalam praktik, karena data geodemografi biasanya dikurangi menjadi beberapa batas administratif seperti kotamadya pada contoh (Gambar.3). Menurut Gambar 3, Kota Manado memiliki 86 desa dengan 2 pulau masing-masing dengan 2 desa yang terpisah dari daratan Sulawesi. Setiap desa yang memiliki pemilih dalam Pemilu akan dianalisis untuk melihat adanya posisi politik dan pandangan terhadap 3 calon Walikota, dimana masing-masing calon Walikota mewakili masing-masing pihak.

Gambar 4. Pemilih Daerah di Manado Pada Tiga Calon

Hal pertama yang harus dilakukan adalah memetakan setiap desa berdasarkan pemilih untuk ketiga kandidat tersebut seperti pada Gambar 4. Ini berarti bahwa kita dapat melihat desa mana yang memiliki pemilih tertinggi untuk masing-masing kandidat.

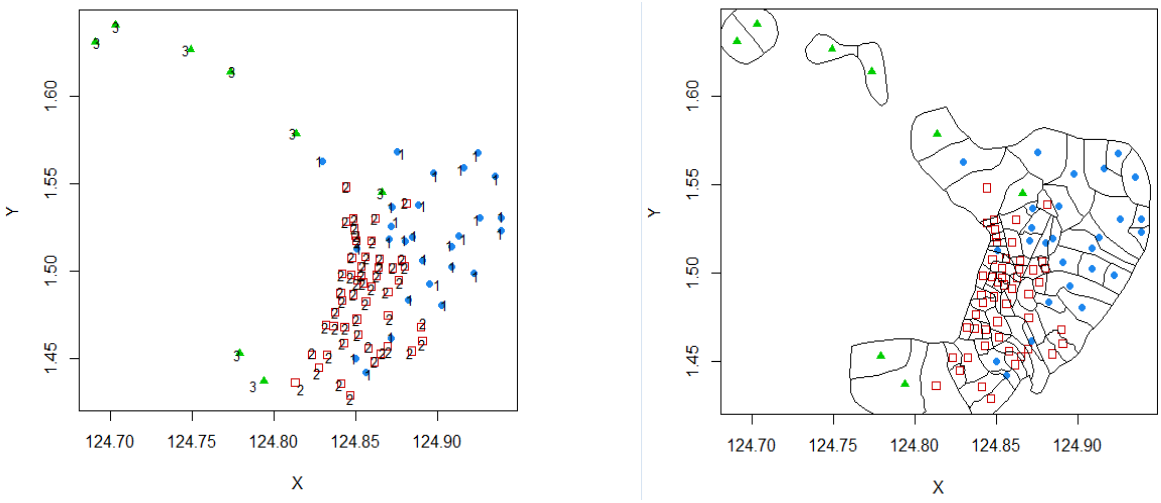
Secara umum dapat dikatakan bahwa kedua pulau yang tampak terpisah dari daratan Sulawesi (yaitu daratan Bunaken dan Bunaken) memiliki tingkat partisipasi yang rendah tidak hanya untuk satu kandidat, namun untuk semua kandidat. Hal ini menunjukkan dua hal, alasan pertama adalah bahwa populasi di dua pulau sangat rendah, dan alasan kedua adalah bahwa orang-orang di tempat itu tidak berniat memilih tiga kandidat karena mereka tidak menyukainya. Kecenderungan pemilih di Indonesia adalah mereka akan memilih dalam pemilihan umum jika kandidat sesuai dengan keinginan rakyat, jika tidak cocok maka masyarakat tidak akan memilih. Mungkin ini kasus pendudukan pulau Bunaken dan Bunaken serta beberapa daerah lainnya di Manado. Karena berdasarkan data yang ada ada beberapa daerah yang memiliki kepadatan penduduk rendah bila dibandingkan dengan daerah lain dengan kepadatan penduduk yang tinggi.

Dalam pemilihan model dan penentuan jumlah cluster, hanya matriks data yang disediakan, dan jumlah komponen pencampuran dan parameterisasi kovarian dipilih dengan menggunakan Bayesian Information Criterion (BIC). Ringkasan yang menunjukkan tiga model teratas dan sebidang jejak BIC untuk semua model yang dipertimbangkan kemudian diperoleh. Pada plot terakhir kita menyesuaikan kisaran sumbu y jadi untuk menghilangkan model tersebut dengan nilai BIC yang lebih rendah. Ada indikasi yang jelas dari campuran tiga komponen dengan kovarians memiliki bentuk yang berbeda namun volume dan orientasinya sama (EVE).



Gambar 5. Ukuran cluster optimal dan entropi berdasarkan BIC

Menurut Gambar 5, dengan menggunakan Model BIC VEE (ellipsoidal, bentuk dan orientasi yang sama), ada 3 cluster pada 10 entropi. Desa dikelompokkan secara klinis menjadi tiga kelompok. Berdasarkan tabel 1, modelnya digambarkan sebagai berikut λDkADᵀk dan parameter kovariannya adalah 1 + (d-1) + G [d (d-1) / 2] dan distribusinya adalah ellipsoidal.



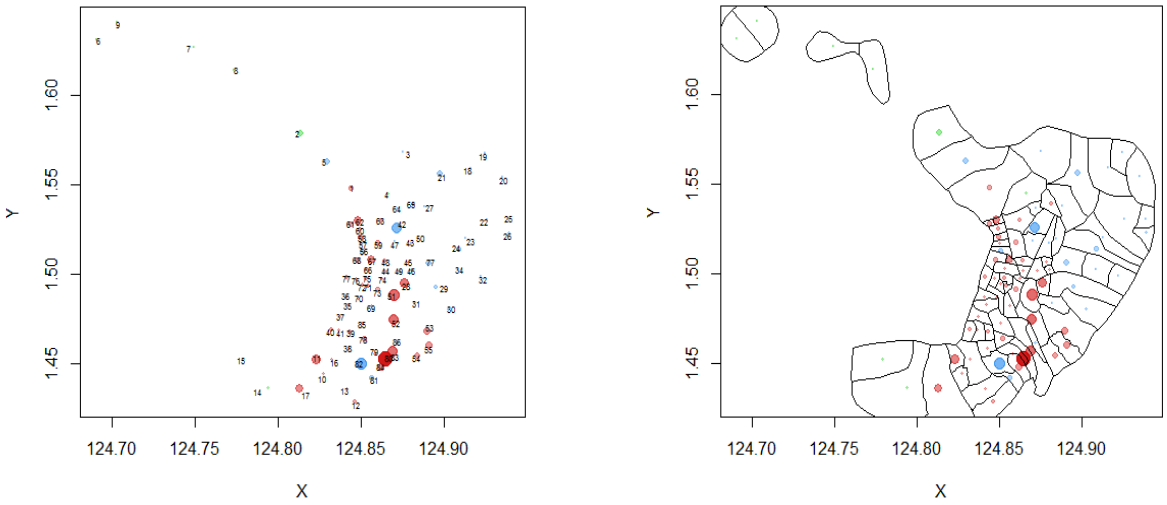
Gambar 6. Semua kandidat (3 kelompok desa optimal)

Berdasarkan model VEE, clusterisasi dibuat untuk mengklasifikasikan wilayah yang memiliki pandangan politik serupa di wilayah kota Manado. Ada 3 kelompok yang berbeda untuk mengklasifikasikan wilayah tersebut, kelompok pertama berjumlah 27 desa, kelompok kedua dari 21 desa dan sisanya berada di kelompok ketiga dari 8 desa. Setiap kelompok klaster memiliki probabilitas pencampuran yang berbeda, yaitu 0,2132314, 0,5562200 dan 0,2305486. Kelompok kedua dari 21 desa memiliki probabilitas pencampuran tertinggi, yang berarti desa-desa memang saling terkait satu sama lain

dan memiliki pandangan politik yang seragam. Hasil yang diperoleh seperti yang ditunjukkan pada Gambar 6.

Pertimbangkan satu hasil cluster dengan 8 anggota daerah. Ini unik untuk hasil cluster dimana ada 5 daerah dengan partisipasi rendah yang seharusnya berada dalam satu cluster, setelah algoritma EM digabungkan dengan 3 wilayah lainnya yang benar-benar memiliki tingkat partisipasi yang tinggi. Hal ini menunjukkan bahwa koordinat spasial berperan penting dalam menunjukkan kesamaan objek dalam proses clustering.

Ternyata tidak semua daerah memilih berada di satu cluster tertentu berarti satu desa mungkin berada di cluster lain. Inilah yang kita sebut wilayah yang tidak pasti, dimana wilayahnya bisa berada di antara dua wilayah.



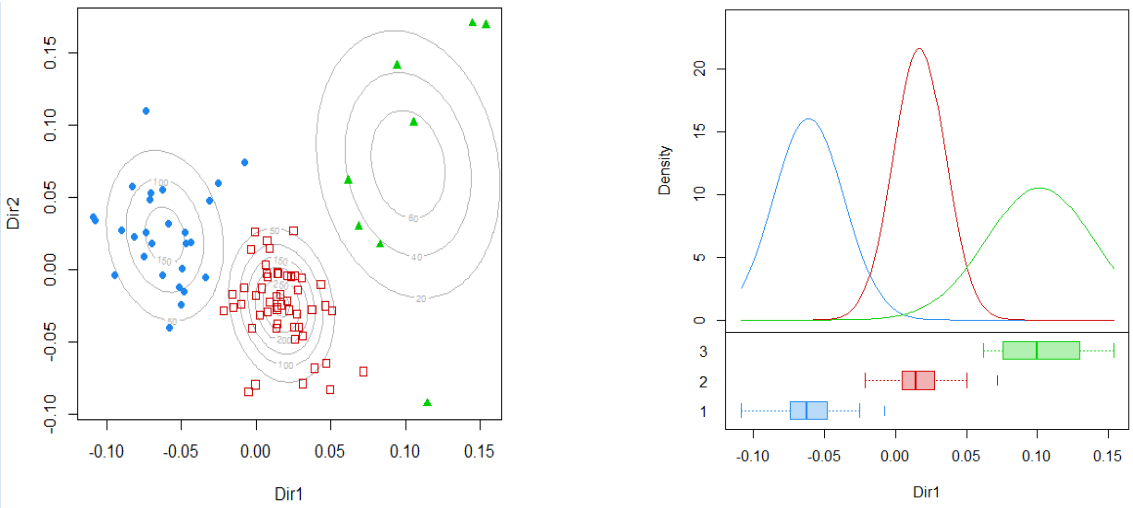
Gambar 7. Uncertainty areas

Secara Geo-politik kita dapat mengatakan bahwa wilayah ini terbagi menjadi dua dalam memilih calon walikota. Ini jelas digambarkan oleh daerah-daerah yang ditunjukkan oleh angka 80, 82, 83, 26, 42, 11, 17, 54, 55, 53, 62, 5 dan 61, masing-masing, seperti yang ditunjukkan pada Gambar 7. Wilayah yang tidak pasti ini dapat telah diperhatikan oleh walikota atau partai berikutnya untuk "mencuri" pemilih yang masih mengambang. Karena para pemilih ini dapat mengalihkan perhatian dan suara mereka ke kandidat lain, hal ini dimungkinkan oleh pengaruh spasial di mana wilayah tertentu mungkin terpengaruh oleh daerah lain yang berdekatan.

Pemilih yang berdiri di dua posisi kandidat memiliki kecenderungan untuk memindahkan keinginan mereka atau bahkan tidak memilih jika kandidat mereka kalah. Kemungkinan menjadi pemilih ayunan sangat besar karena kekecewaan yang mereka dapatkan.

Pemilihan umum di Indonesia yang memungkinkan pemilihan putaran kedua, dapat digunakan oleh kandidat terpilih untuk mengundang pemilih ayun untuk bergabung. Untuk kasus di Manado, pemilih yang berada di wilayah terapung sangat berisiko dan berpotensi diambil alih jauh dengan menambahkan beberapa janji kampanye untuk memajukan kesejahteraan rakyat dan wilayah mereka.

Penentuan daerah ketidakpastian selama masa pemilihan sangat penting dan menarik untuk didiskusikan sebagai bagian dari penelitian ini. Swing pemilih yang terdiri dari pemilih yang belum memutuskan dan pendukung lunak memiliki peran penting dalam menentukan kemenangan salah satu kandidat. Keberadaan suku, agama dan ras menentukan hasil pemilihan. Beberapa kasus melibatkan ketiga hal tersebut, karena pemilih akan memilih kandidat berdasarkan ras, agama dan ras.



Gambar 8. Dimensi Reduksi dengan kontur dan densitas

Dengan alasan memfasilitasi analisis klaster, pengurangan dimensi dilakukan tanpa mengurangi nilai informasi. Kita dapat melihat pada Gambar 8 sebagai hasil pengurangan dimensi yang menunjukkan peta permukaan dan densitasnya. Ada satu desa sebagai ilustrasi seperti yang ditunjukkan dalam peta ketidakpastian (Gambar 7), dan area lain yang juga terbukti dapat dilihat dengan jelas pada Gambar 8. Untuk alasan politis, kita dapat mengatakan daerah-daerah ini masih memiliki kemungkinan untuk mengubah pandangan politik mereka. untuk memilih jurusan mereka berikutnya.

Luaran (Output)

Output penelitian ini berupa model klastering probabilistik menggunakan algoritma EM dan dimodelkan lewat Peta Clustering.

BAB VI.

RENCANA TAHAPAN BERIKUTNYA

Untuk penelitian di tahun kedua, penelitian akan dilakukan berdasarkan:

Pertama, melakukan analisis spasial terhadap daerah kecamatan yang memiliki ketidakpastian dalam memilih. Ada 13 kelurahan dari 86 kelurahan di Manado yaitu: Pakowa, Teling Atas, Karombasan Utara, Paniki 2, Singkil 2, Bahu, Malalayang1 Barat, Taas, Paal 4, Tikala Baru, Tumumpa 1, Meras, Tumumpa 2.

Kedua, membuat model matematika variogram antara jarak setiap wilayah amatan dengan ragam yang dinyatakan sebagai semivariogram. Sebagai analisis cluster lebih lanjut akan dilakukan untuk area dengan probabilitas campuran untuk tiga cluster, dengan perhatian yang lebih kuat terhadap cluster 2 (sekitar 0.5562200), dengan membuat model dasar karakteristik pemilih.

Ketiga, untuk menginterpolasi wilayah yang tidak pasti dalam melakukan pemilihan walikota menggunakan metode Kriging berdasarkan model matematika semivariogram.

BAB VII

KESIMPULAN DAN SARAN

Algoritma EM dapat diterapkan untuk menentukan area mana yang memiliki pandangan politik serupa di Manado. Penerapan algoritma ini mengikuti model yang dipilih dimana melalui model kita dapat menentukan ukuran cluster adalah 3 cluster. Berdasarkan model VEE, clusterisasi dibuat untuk mengklasifikasikan wilayah yang memiliki pandangan politik serupa di wilayah kota Manado. Ada 3 kelompok yang berbeda untuk mengklasifikasikan wilayah tersebut, kelompok pertama berjumlah 27 desa, kelompok kedua dari 21 desa dan sisanya berada di kelompok ketiga dari 8 desa. Mengikuti ukuran cluster, juga bisa memperhitungkan probabilitas pencampuran akun. Setiap kelompok klaster memiliki probabilitas pencampuran yang berbeda, yaitu 0,2132314, 0,5562200 dan 0,2305486. Kelompok kedua dari 21 desa memiliki probabilitas pencampuran tertinggi, yang berarti desa-desa memang saling terkait satu sama lain dan memiliki pandangan politik yang seragam.

Ada beberapa daerah yang dianggap sebagai wilayah ketidakpastian, artinya setiap saat wilayah tersebut dapat berubah sewaktu-waktu. Ada 13 wilayah yang tidak pasti dan masing-masing 80, 82, 83, 26, 42, 11, 17, 54, 55, 53, 62, 5 dan 61. Daerah ini harus diperhitungkan oleh pemain politik sehingga bisa dimanfaatkan oleh sifat pemilih yang tidak pasti karena bisa dipengaruhi oleh orang lain.

UcapanTerima Kasih

Kami berterima kasih kepada Kementerian Riset dan Pendidikan Indonesia yang telah membantu membiayai dan menyetujui penelitian ini.

Daftar Pustaka

[1] G. J. Mclachlan and T. Krishnan., 1996, The EM Algorithm and Extensions. Wiley-Interscience, 1 edition, November.
[2] A. P. Dempster, N. M. Laird, and D. B. Rubin., 1977, Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39(1):138.
[3] C. Fraley and A. E. Raftery., 2007, Model-based Methods of Classification: Using the mclust Software in Chemometrics. Journal of Statistical Software. January, Volume 18, Issue 6. http://www.jstatsoft.org/
[4] L. Scrucca, M. Fop, T. B. Murphy and A. E. Raftery., 2016, mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. HHS Public Access. R J. 8(1): 289–317. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5096736/pdf/nihms793803.pdf
[5] McLachlan G, Peel D., 2000, Finite Mixture Models. Wiley; New York.
[6] Banfield J, Raftery AE., 1993, Model-based Gaussian and non-Gaussian clustering. *Biometrics*. 49: 803–821.
[7] Celeux G, Govaert G., 1995, Gaussian parsimonious clustering models. *Pattern Recognition*. 28:781–793.
[8] Hutapea, A., 2015, Dinamika Hukum Pemilihan Kepala Daerah di Indonesia, *Jurnal Rechts Vinding*, vol. 4.
[9] Arfa'i, 2014, Pengaruh Pemilihan Kepala Daerah Secara Langsung Dengan Penyelenggaraan Pemerintahan Daerah, Majalah Hukum Forum Akademika, vol. 25.

Lampiran. Artikel ilmiah (*draft*, status *submission)*

Gmail **Winsy Weku Wanea <winsyweku@gmail.com>**

Submission of Manuscript for Journal : International Journal of Ecological Economics and Statistics

2 pesan

Winsy Weku Wanea <winsyweku@gmail.com> 2 September 2017 16.04
Kepada: kks_ceser@yahoo.com

To
The Managing Editor-in-Chief
International Journal of Ecological Economics and Statistics

Subject: Submission of Manuscript for Journal : International Journal of Ecological Economics and Statistics

Reference:

1. **Paper Title: Identification of Manado's PILWAKO as the Candidate Mayor Territory Political Power in 2015 using EM algorithm with Model Based Selection**
2. **Subject Classification Numbers** (At least one of the followings)**:**
   A. Mathematics Subject Classification (MSC):
   (62P25, 86A32**)**
   B. Computing Classification System (CCS) :
   (D2.12)
3. **Journal Topic(s):**
4. **Author's name: Winsy Christo Deilan Weku**
5. **Affiliation(s), Mailing address of Author(s): Sam Ratulangi University, Indonesia, email:** winsyweku@gmail.com

Dear Editor,

With reference to above, please find my submission of paper for possible publications in ---- **International Journal of Ecological Economics and Statistics**.

I have read the "*Author Instructions*" of journal.

A. Please removed the designation (e.g. Director, Head, Professor, lecturer, research scholar etc.) and title (e.g. Dr. Ms. Mr. etc.) from author(s) name in your paper.

B. An active Email id of all Authors should be given in the paper.

I hereby affirm that the content of this manuscript are original. Furthermore it has been **neither** published elsewhere fully or partially or any language **nor** submitted for publication (fully or partially) elsewhere simultaneously.

Also I declare that *"It has not been published before, and it is not under consideration for publication in any other journals. It contains no matter that is scandalous, obscene, fraud, plagiarism, libelous, or otherwise contrary to law. I/we followed the Journal's adopted "Publication ethics and malpractice" statement which given in journal's website in **About** section and we will be answerable for the correctness (or plagiarism) and authenticity of article."*

I/we also affirm that the all authors have seen to the name of all indexing agencies on journal home web site which indexed the journal. I also agree that "It's depend on indexing agencies when, how and what manner they can index or not. Hence, I/we are well informed that on the basis of earlier indexing, journal/publisher can't predict the today or future indexing policy of third party (i.e. indexing agencies) as they have right to discontinue any journal at any time without prior information to the journal. As journal's role is just to provide the online access to the indexing agencies**".**

I affirm that the all authors have seen and agreed to the submitted version of the paper and their inclusion of name(s) as co-author(s).

Signature of Corresponding Author
Name: Winsy Weku
Affiliation: Sam Ratulangi University, Manado, Indonesia
Mailing address: winsyweku@gmail.com

winsyweku_ijees.docx
285K

International Journal of Ecological Economics and Statistics <kks_ceser@yahoo.com> 2 September 2017 23.32
Balas Ke: International Journal of Ecological Economics and Statistics <kks_ceser@yahoo.com>
Kepada: Winsy Weku Wanea <winsyweku@gmail.com>

International Journal of Ecological Economics & Statistics
www.ceserp.cpm/cp-jour/
www.ceser.in/ceserp/

Paper:
Identification of Manado's PILWAKO as the Candidate Mayor Territory Political Power in 2015 using EM algorithm with Model Based Selection

Reference: IJEES/MS010917

Dear Author,

This is to acknowledge that we have received your manuscript. Upon receiving the referees' reports, we shall inform you of the editorial decision with regard to your manuscript.

Kindly use the above given reference number in all your future inquiries about your manuscript. The reviewers taking minimum 10 week time after sending us consent.

To update the information about IJEES, please join the **IJEES Yahoo Group**
https://groups.yahoo.com/neo/groups/IJEES/info

Thanks for your submission for **International Journal of Ecological Economics & Statistics**.

With regards

Disclaimer/Regarding indexing issue:
We have provided the online access of all issues and papers to the indexing agencies (as given on journal web site). **It's depend on indexing agencies when, how and what manner they can index or not. Hence, we like to inform that on the basis of earlier indexing, we can't predict the today or future indexing policy of third party (i.e. indexing agencies) as they have right to discontinue any journal at any time without prior information to the journal. So, please neither sends any question nor expects any answer from us on the behalf of third party i.e. indexing agencies. Hence, we will not issue any certificate or letter for indexing issue.** Our role is just to provide the online access to them. So we do properly this and one can visit indexing agencies website to get the authentic information.

Dr. Kaushal K. Srivastava
Editor-in-Chief,
International Journal of Ecological Economics and Statistics™
[ISSN 0973-1385 (Print); 0973-7537 (Online)]
Email: eic.ijees@yahoo.com; kks_ceser@yahoo.com

CESER PUBLICATIONS
www.ceser.in/ceserp/
http://www.ceserp.com/cp-jour/

Our Journals:

International Journal of Applied Mathematics and Statistics™ [ISSN 0973-1377; 0973-7545]
International Journal of Mathematics and Computation™ [ISSN 0974-570X; 0974-5718]
International Journal of Mathematics and Statistics™ [ISSN 0974–7117; 0973-8347]
International Journal of Imaging and Robotics™ [ISSN 2231-525X]
International Journal of Tomography and Simulation™ [ISSN 2319-3336]
International Journal of Artificial Intelligence™[ISSN 0974-0635]
International Journal of Ecological Economics and Statistics™ [ISSN 0973-1385; 0973-7537]
International Journal of Statistics and Economics™ [ISSN 0975-556X]
International Journal of Ecology & Development™ [ISSN 0973-7308; 0972-9984]
International Journal of Engineering and Future Technology™ [ISSN 2455-6432]
International Journal of Library Science™ [ISSN 0975 – 7546]

****

Take negative always as a challenge and positive as an encouragement
****

From: Winsy Weku Wanea <winsyweku@gmail.com>
To: kks_ceser@yahoo.com
Sent: Saturday, 2 September 2017 1:34 PM
[Kutipan teks disembunyikan]
[Kutipan teks disembunyikan]

Identification of Manado's PILWAKO as the Candidate Mayor Territory Political Power in 2015 using EM algorithm with Model Based Selection

Winsy Weku[1] and **Altien Rindengan**[2]

[1]Department of Mathematics, Sam Ratulangi University, Indonesia
Email: winsyweku@gmail.com,
[2]Department of Mathematics, Sam Ratulangi University, Indonesia
Email: altien@unsrat.ac.id

ABSTRACT

PILWAKO is a special term and unique voting system residing in Indonesia followed by all levels of society that already have the conditions to follow the election. The problem is how to determine which areas can provide significant and measurable results to be analyzed, so that they can be used for next PILWAKO. In other words we are expected to determine a significant region or region that is still subject to change or an unstable tendency. A probabilistic clustering algorithm expectation maximization (EM) to clustering spatial dataset to identify the PILWAKO political area of the election of Manado Major in 2015. It implements parameterized Gaussian hierarchical clustering algorithms and the EM algorithm for parameterized Gaussian mixture models with the possible addition of a Poisson noise term. The result is that there are three clustered areas using EM algorithm. All three areas are well distributed. Among the clustered areas there are several areas that are uncertainty areas. This indicates that the region has a tendency to change in selecting candidates in the next Pilwako.

Key words : EM Algorithm , PILWAKO, Political Teritory

Mathematics Subject Classification: 62P25, 86A32

Computing Classification System: D2.12

1. INTRODUCTION

The PILWAKO (*Pemilihan Walikota*/Mayor Election) process in Indonesia is a long journey that should accommodate the political interests and desires of the people, central and regional interests, as well as national and international interests (see Hutapea). The implementation of PILWAKO is conducted simultaneously intended to avoid fraudulent practices and candidates must follow the public. Unlike previous executions undertaken through an appointment mechanism by the authorities, PILWAKO is conducted to accommodate the community's desire to choose a more transparent and better known leader. This is intended to avoid rejection by the wider community who are not familiar with the hierarchical system.

Direct elections are the process of election of Provincial Head Governor namely election of Governor and Deputy Governor, Regency level that is election of Regent and Vice Regent and City Level that is election of Mayor and Vice Mayor which executed simultaneously or simultaneously concurrently directly and democratically by people, namely to exercise the sovereignty of the people. The government and the

House of Representatives have agreed that the pilwako/pilkada schedule will be simultaneously divided into three waves, namely in 2015, 2017, and 2018 and so on until the Regional Head Election simultaneously in 2027. In general, pilwako/pilkada simultaneously present as a means of strengthening in Indonesia. Far from the least there are three things that must be answered from the arrival of pilkada simultaneously; First, to create efficient and effective elections; Second, to strengthen the degree of representation between the community and the regional head; Third, creating effective and efficient regional governance.

In this research will be undertaken research on classify areas in Manado city election as mayoral candidate in 2015 using Gaussian Mixture Model (EM algorithm) to identify clustering area. Each region is expected to have a probability to converge in one particular cluster, to be estimated by maximizing the probability vector of the distribution parameter.

The specific purpose of this study is to determine which areas contribute to the selection of the mayor of Manado so that it will make it easier for the government and researchers to mapping the political power of each couple in the future.

2. GENERATION OF THE DATA

In this section will be presented some theoretically related to this research, such as clusterization and EM algorithm.

2.1. Clustering

Clustering is one way to describe the similarity of objects in large data groups that have certain characteristics. Clustering methods analyze and explore a dataset to associate objects in groups, such that the objects in each groups have common characteristics. These characteristics may be expressed in different ways: for example, one may describe the objects in a cluster as the population generated by a joint distribution, or as the set of objects that minimize the distances from the centroid of the group.

In some fields of science, especially statistics and optimization or operations research and other applied fields, the main problem is how to optimize by maximizing or minimizing a function along with its variables. As these optimization problems may assume several different types, each one with its own characteristics, many techniques have been developed to solve them. This techniques are very important in data mining and knowledge discovery area as it can be used as basis for most complex and powerful methods.

One of these techniques is the Maximum Likelihood and its main goal is to adjust a statistical model with a specific data set, estimating its unknown parameters so the function that can describe all the parameters in the dataset. In other words, the method will adjust some variables of a statistical model from a dataset or a known distribution, so the model can “describe” each data sample and estimate others.

It was realized that clustering can be based on probability models to cover the missing values. This provides insights into when the data should conform to the model and has led to the development of new clustering methods such as Expectation Maximization (EM) that is based on the principle of Maximum Likelihood of unobserved variables in finite mixture models.

In this paper we propose the use probabilistic based classification techniques for two datasets to identify the clustering area. We conducted extensive comparisons of the mentioned indices in conjunction with the Expectation-Maximization clustering algorithm on a number of widely used data sets, and make a simple analysis of the experimental results.

The EM algorithm can be used to match and group a data set based on the similarity of the observed objects. However, since we use spherical Gaussians, a variance scalar is used in place of the covariance matrix. The prior probability for each Gaussian is the fraction of points in the cluster defined by that Gaussian. These parameters can be initialized by randomly selecting means of the Gaussians, or by using the output of EM algorithm for initial centers. The algorithm converges on a locally optimal solution by iteratively updating values for means and variances.

2.2. Gaussian Finite Mixture Models

Let $x = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ be a sample of n independent identically distributed observations. The distribution of every observation is specified by a probability density function through a finite mixture model of G components, which takes the following form (see Scrucca and Fop)

$$f(x_i;\Psi)=\sum_{k=1}^{G} \pi_k f_k(x_i;\theta_k) \qquad (1)$$

where $\boldsymbol{\Psi} = \{\pi_1, \ldots, \pi_{G-1}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_G\}$ are the parameters of the mixture model, $f_k(\boldsymbol{x}_i; \boldsymbol{\theta}_k)$ is the kth component density for observation $\boldsymbol{x}_i$ with parameter vector $\boldsymbol{\theta}_k$, $(\pi_1, \ldots, \pi_{G-1})$ are the mixing weights or probabilities (such that $\pi_k > 0$, $\sum_{k=1}^{G} \pi k = 1$), and G is the number of mixture components.

Assuming that G is fixed, the mixture model parameters $\boldsymbol{\Psi}$ are usually unknown and must be estimated. The log-likelihood function corresponding to equation (1) is given by $\ell(\Psi;x_1,\ldots,x_n)=\sum_{i=1}^{n} log(f(x_i;\Psi))$. Direct maximisation of the log-likelihood function is complicated, so the maximum likelihood estimator (MLE) of a finite mixture model is usually obtained via the EM algorithm (see Dempster, Laird and Rubin, MacLahlan and Peel.

In the model-based approach to clustering, each component of a finite mixture density is usually associated with a group or cluster. Most applications assume that all component densities arise from the same parametric distribution family, although this need not be the case in general. A popular model is the Gaussian Mixture Model (GMM), which assumes a (multivariate) Gaussian distribution for each component, i.e. $f_k(\boldsymbol{x}; \boldsymbol{\theta}_k) \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. Thus, clusters are ellipsoidal, centered at the mean vector $\boldsymbol{\mu}_k$, and with other geometric features, such as volume, shape and orientation, determined by the covariance matrix $\boldsymbol{\Sigma}_k$. Parsimonious parameterisations of the covariances matrices can be obtained by means of an eigen-decomposition of the form $\sum k=\lambda_k D_k A_k D^T_k$, where λ_k is a scalar controlling the volume of the

ellipsoid, $\boldsymbol{A}_k$ is a diagonal matrix specifying the shape of the density contours with det($\boldsymbol{A}_k$) = 1, and $\boldsymbol{D}_k$ is an orthogonal matrix which determines the orientation of the corresponding ellipsoid (see Banfield and Raftery, Celeux and Govaert). In one dimension, there are just two models: E for equal variance and V for varying variance. In the multivariate setting, the volume, shape, and orientation of the covariances can be constrained to be equal or variable across groups.

2.3. Best Model

The model options available in the R package mclust are summarized in Table 1. In one dimension, there are just two models: E for equal variance and V for varying variance. In more than one dimension, the model identifiers encode geometric characteristics of the model. For example, EVI denotes a model in which the volumes of all clusters are equal (E), the shapes of the clusters may vary (V), and the orientation is the identity (I). Clusters in this model have diagonal covariances with orientation parallel to the coordinate axes. Parameters associated with characteristics designated by E or V are determined from the data. A 'best' model can be estimated by fitting models with differing parameterizations and/or.

Table 1: Parameterizations of the multivariate Gaussian mixture model available in mclust.

Identifier	Model	#Covariance parameters	Distribution
EII	λI	1	Spherical
VII	$\lambda_k I$	G	Spherical
EEI	λA	D	Diagonal
VEI	$\lambda_k A$	G + (d-1)	Diagonal
EVI	λA_k	1 + G(d-1)	Diagonal
VVI	$\lambda_k A_k$	Gd	Diagonal
EEE	λDAD^T	d(d+1)/2	Ellipsoidal
EEV	$\lambda D_k A D_k^T$	1+(d-1)+G[d(d-1)/2]	Ellipsoidal
VEV	$\lambda_k D_k A D_k^T$	G+(d-1)+G[d(d-1)/2]	Ellipsoidal
VVV	$\lambda_k D_k A_k D_k^T$	G[d(d+1)/2]	Ellipsoidal

Model identifiers use three letters to encode code geometric characteristics: volume, shape, and orientation. E means equal and V means varying across components or clusters; I refers to the identity matrix in specifying shape or orientation and is a special case of E. In the column labeled '# Covariance parameters', d denotes the dimension of the data, and G denotes the number of mixture components. The total number of parameters for each model can be obtained by adding Gd parameters for the means and G − 1 parameters for the mixing proportions.

2.4. EM Algorithm

The EM algorithm is an unsupervised clustering method, that is, doesn't require a training phase, based on mixture models. It follows an iterative approach, sub-optimal, which tries to find the parameters of the probability distribution that has the maximum likelihood of its attributes.

In general lines, the algorithm's input are the data set (x), the total number of clusters (M), the accepted error to converge (e) and the maximum number of iterations. For each iteration, first is executed the E-Step (E-xpectation), that estimates the probability of each point belongs to each cluster, followed by the M-step (M-aximization), that re-estimate the parameter vector of the probability distribution of each class. The algorithm finishes when the distribution parameters converges or reach the maximum number of iterations. (see Dempster, Laird and Rubin)

The EM algorithm extends this basic approach to clustering in two important ways:

- Instead of assigning examples to clusters to maximize the differences in means for continuous variables, the EM clustering algorithm computes probabilities of cluster memberships based on one or more probability distributions. The goal of the clustering algorithm then is to maximize the overall probability or likelihood of the data, given the (final) clusters.
- Unlike the classic implementation of k-means clustering, the general EM algorithm can be applied to both continuous and categorical variables (note that the classic k-means algorithm can also be modified to accommodate categorical variables).

Initialization

Each class *j*, of *M* classes (or clusters), is constituted by a parameter vector *(θ),* composed by the mean (μ_j) and by the covariance matrix (P_j), which represents the features of the Gaussian probability distribution (Normal) used to characterize the observed and unobserved entities of the data set *x*.

$$\theta(t) = \left(\mu_j(\mathrm{t}), P_j(\mathrm{t})\right), j = 1, \dots, M$$

On the initial instant (t = 0) the implementation can generate randomly the initial values of mean (μ_j) and of covariance matrix (P_j). The EM algorithm aims to approximate the parameter vector (θ) of the real distribution.

E-Step

This step is responsible to estimate the probability of each element belong to each cluster ($P(C_j|x_k)$). Each element is composed by an attribute vector (x_k). The relevance degree of the points of each cluster is given by the likelihood of each element attribute in comparison with the attributes of the other elements of cluster C_j.

$$P(C_j|x) = \frac{\left|\sum_j(t)\right|^{-\frac{1}{2}} e^{n_j} P_j(t)}{\sum_{k=1}^{M}\left|\sum_j(t)\right|^{-\frac{1}{2}} e^{n_j} P_k(t)}$$

M-Step

This step is responsible to estimate the parameters of the probability distribution of each class for the next step. First is computed the mean (μ_j) of class j obtained through the mean of all points in function of the relevance degree of each point.

$$\mu_j(t+1) = \frac{\sum_{k=1}^{N} P(C_j|x_k)\, x_k}{\sum_{k=1}^{N} P(C_j|x_k)}$$

To compute the covariance matrix for the next iteration is applied the Bayes Theorem, which implies that P(A|B)=P(B|A)*P(A)P(B), based on the conditional probabilities of the class occurrence.

$$\sum_j (t+1) = \frac{\sum_{k=1}^{N} P(C_j|x_k)\,(x_k - \mu_j(t))(x_k - \mu_j(t))}{\sum_{k=1}^{N} P(C_j|x_k)}$$

The probability of occurrence of each class is computed through the mean of probabilities (C_j) in function of the relevance degree of each point from the class.

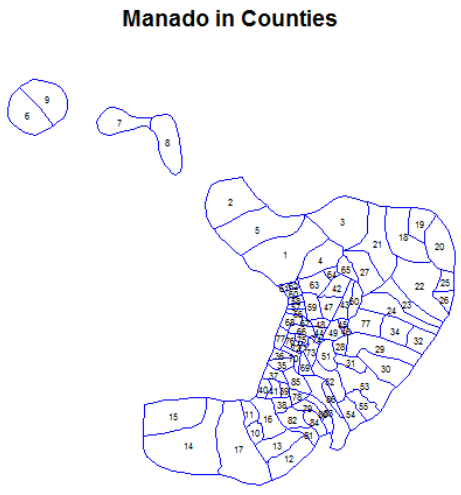
$$P_j(t+1) = \frac{1}{N}\sum_{k=1}^{N} P(C_j|x_k)$$

The attributes represents the parameter vector θ that characterize the probability distribution of each class that will be used in the next algorithm iteration.

3. RESULTS

This study includes geospatial analysis examples from real world: geo-demographic. Cluster size was chose based on model. As a computational tool was chose R. Spatial map data presentation were done by ArcView. The R package that will be used is the MCLUST-v5.1 developed by Chris Fraley and Adrian Raftery, available in CRAN repository.

Our example of geodemographic analysis was observed in practice, because the geodemographic data usually reduced to some administrative boundaries such as the municipalities in the example (Fig.1).



DESA_NO >	DESA
001	MOLAS
002	TONGKEINA
003	PANDU
004	BAILANG
005	MERAS
006	MANADO TUA DUA
007	ALUNG BANUA
008	BUNAKEN
009	MANADO TUA SATU
010	BATUKOTA
011	BAHU
012	WINANGUN 1
013	WINANGUN 2
014	MALALAYANG SATU TIMUR
015	MALALAYANG DUA
016	KLEAK
017	MALALAYANG SATU BARAT
018	KIMA ATAS
019	MAPANGET BARAT
020	LAPANGAN
021	BENGKOL
022	PANIKI BAWAH
023	KAIRAGI DUA
024	KAIRAGI SATU
025	PANIKI 1
026	PANIKI 2
027	BUHA
028	DENDENGAN LUAR
029	PAAL DUA
030	RANOMUT
031	DENDENGAN DALAM
032	MALENDENG
033	KAIRAGI WERU
034	PERKAMIL
035	TITIWUNGEN SELATAN
036	TITIWUNGEN UTARA
037	SARIO UTARA
038	RANOTANA
039	SARIO KOTA BARU
040	SARIO TUMPAAN
041	SARIO
042	SINGKIL 2
043	KOMBOS BARAT
044	KARAME
045	TERNATE TANJUNG
046	TERNATE BARU
047	SINGKIL 1
048	WAWONASA
049	KETANG BARU
050	KOMBOS TIMUR
051	TIKALA ARES
052	BANJER
053	TIKALA BARU
054	TAAS
055	PAAL 4
056	SINDULANG SATU
057	SINDULANG DUA
058	BITUNGKARANGRIA
059	KAMPUNG ISLAM
060	MAASING
061	TUMUMPA 2
062	TUMUMPA 1
063	TUMINTING
064	MAHAWU
065	SUMOMPO
066	PINAESAAN
067	ISTIQLAL
068	CALACA
069	TELING BAWAH
070	BUMI BERINGIN
071	MAHAKERET BARAT
072	MAHAKERET TIMUR
073	TIKALA KUMARAKA
074	KOMO LUAR
075	LAWANG IRUNG
076	WENANG UTARA
077	WENANG SELATAN
078	WANEA
079	BUMI NYIUR
080	PAKOWA
081	KAROMBASAN SELATAN
082	KAROMBASAN UTARA
083	TELING ATAS
084	RANOTANA WERU
085	TANJUNG BATU
086	TINGKULU

Figure 1. Administrative Boundaries of Manado

According to Figure 1, Manado City has 86 villages with 2 islands each with 2 villages separated from the mainland of Sulawesi. Every village that has voters in the General Election will be analyzed to see the existence of political positions and views on 3 Mayor candidates, where each Mayor candidate represents their respective parties.

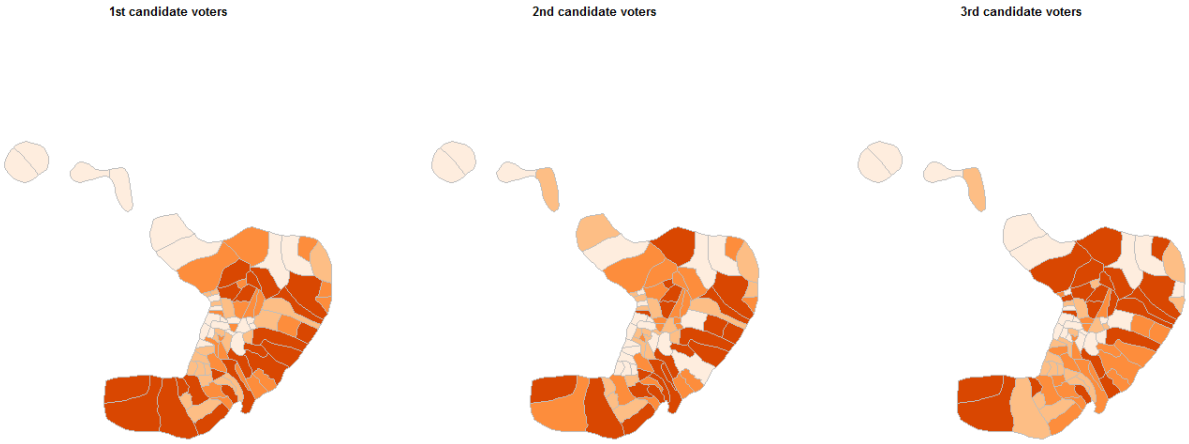


Figure 2. Regional State Voters In Manado On The Three Candidates

The first thing to do is to map each village on the basis of voters for all three candidates as in Figure 2. This means that we can see which villages have the highest electorate for each candidate.

In general it can be said that the two islands that appear to be separated from the mainland of Sulawesi (i.e. Bunaken and Bunaken mainland) have low participation rates not only for one candidate, but for all candidates. This indicates two things, the first reason is that the population on the two islands is very low, and the second reason is that the people in that place do not intend to choose the three candidates because they do not like them. The tendency for voters in Indonesia is that they will come to vote in the general election if the candidates are in accordance with the wishes of the people, otherwise if they do not match then the community will not come to vote. Perhaps this is the case for the occupation of the Bunaken and Bunaken islands as well as several other areas in Manado. Because based on existing data there are some areas having low population density when compared to other areas of high population density.

In the selection of models and the determination of the number of clusters, only the data matrix is provided, and the number of mixing components and the covariance parameterisation are selected using the Bayesian Information Criterion (BIC). A summary showing the top-three models and a plot of the BIC traces for all the models considered is then obtained. In the last plot we adjusted the range of y-axis so to remove those models with lower BIC values. There is a clear indication of a three-component mixture with covariances having different shapes but the same volume and orientation (EVE).

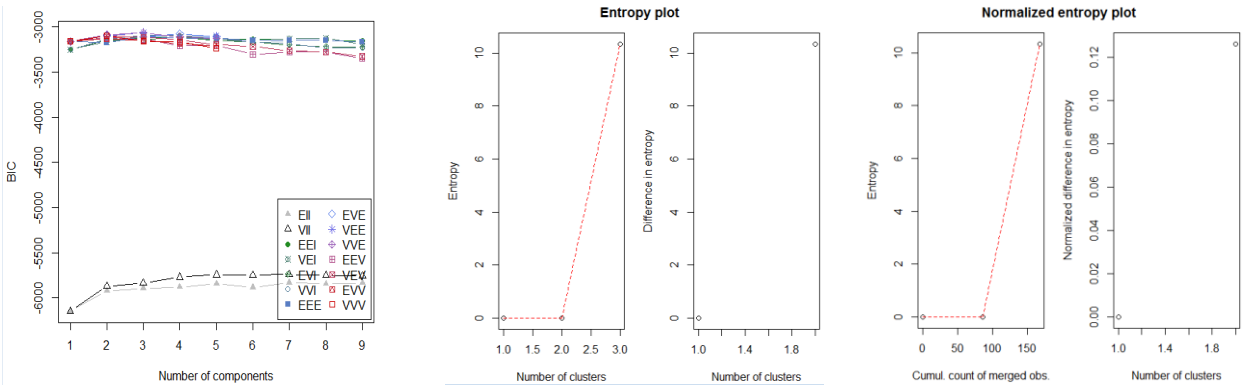


Figure 3. Optimal cluster size and entropy based on BIC

According to Figure 3, using the BIC VEE Model (ellipsoidal, equal shape and orientation), there are 3 clusters at 10th entropy. The villages are classified clinically into three groups. Based on tabel 1, the model is describe as follows $\boldsymbol{\lambda D_k A D^T_k}$ and the covariance parameter is **1 + (d-1) + G[d(d-1)/2]** and its distribution is ellipsoidal.

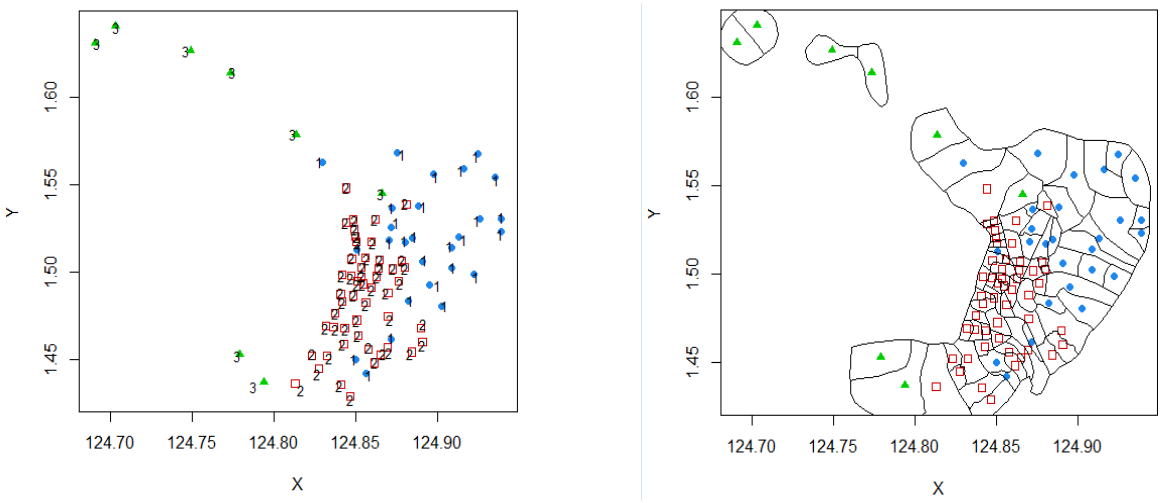


Figure 4. All candidate (3 clusters of villages are optimum)

Based on the VEE model, clusterization is made to classify any region that has similar political views in the city area of Manado. There are 3 different groups to classify the region, the first group amounted to 27 villages, the second group of 21 villages and the rest are in the third group of 8 villages. Each cluster group has different mixing probabilities, ie 0.2132314, 0.5562200 and 0.2305486, respectively. The second cluster group of 21 villages has the highest mixing probabilities, which means the villages are indeed closely related to each other and have a uniform political outlook. The results obtained as shown in Figure 4.

Consider one cluster result with 8 member regions. It is unique to the cluster results where there are 5 low-participation areas that should be in one cluster, after the EM algorithm is combined with the other 3

regions that actually have a high level of participation. This shows that spatial coordinates play an important role in showing the similarity of objects in the clustering process.

Apparently not all regions choose to be in one particular cluster means that one village may be in another cluster. This is what we call an uncertain territory, where the region can be between two regions.

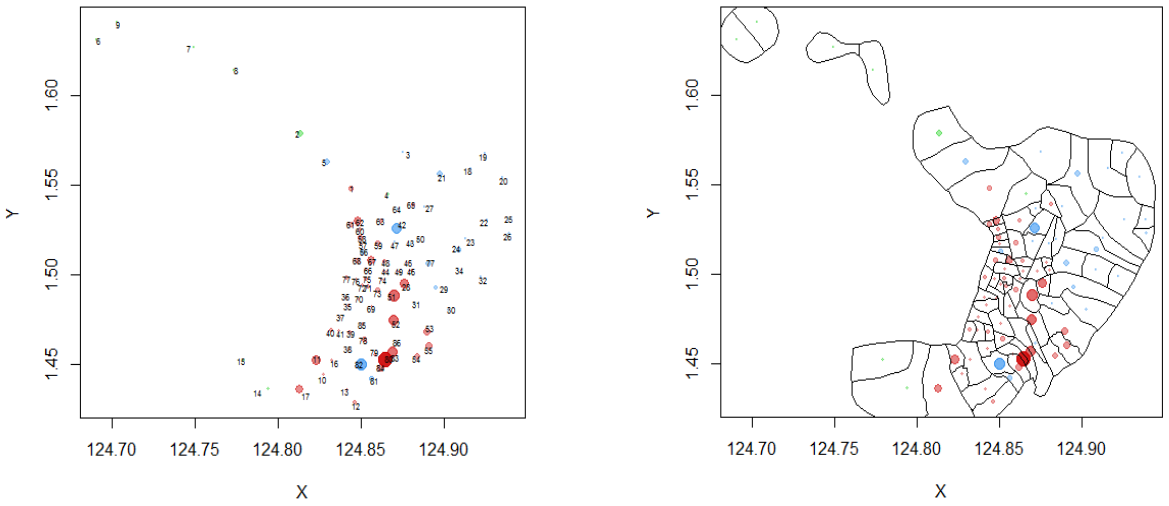


Figure 5. Uncertainty areas

Geo-politically we may say that the region is divided into two in choosing a candidate for mayor. This is clearly illustrated by the regions represented by the numbers 80, 82, 83, 26, 42, 11, 17, 54, 55, 53, 62, 5 and 61, respectively, as shown in Figure 5. This uncertain territory could have been noticed by the next mayor or party to "steal" the still-floating voters. Because these voters may divert their attention and voice to other candidates, this is made possible by spatial influences in which a particular region may be affected by other adjacent areas.

Voters who stand in two candidate positions have a tendency to move their desires or not even vote if their candidate loses. The possibility of becoming a swing voters is huge because of the disappointment they get.

General elections in Indonesia that allow for the election of the second round, can be used by the elected candidates forward to invite swing voters to join. For the case in Manado, voters who are in a floating region are very risky and potentially to be taken over much less by adding some campaign promises to advance the welfare of their people and territories.

Determination of uncertainty areas during the election period is very important and interesting to discuss as part of this research. Swing voters consisting of undecided voters and soft supporters have an important role in determining the victory of one of the candidates. The existence of tribe, religion and race determine the election result. Some cases involving these three things, because voters will choose candidates based on race, religion and race.

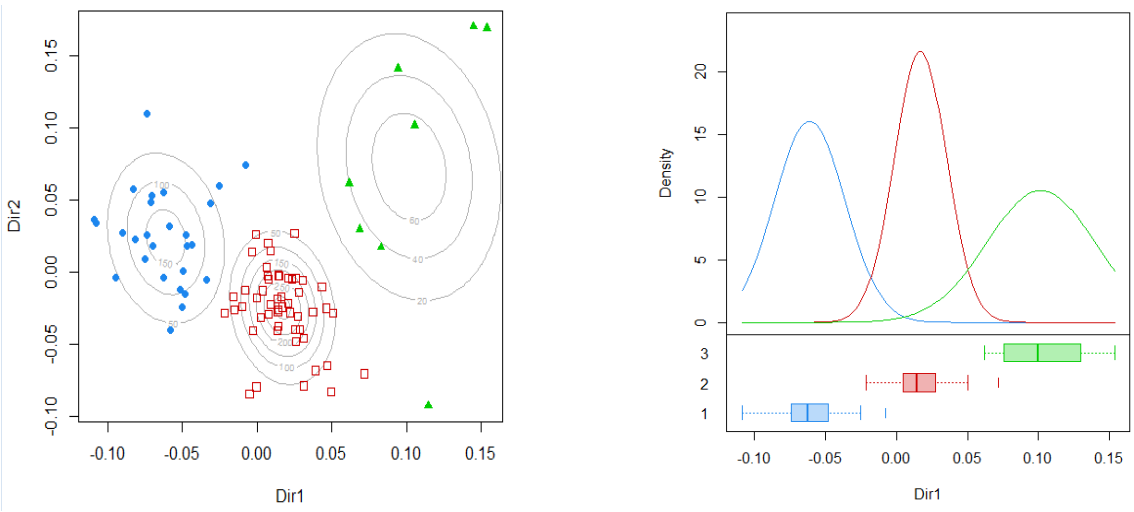


Figure 6. Dimension Reduction with countur and density

By reason of facilitating cluster analysis, dimension reduction is done without reducing the value of information. We can see in Figure 6 as the result of dimensional reduction showing the surface map and its density. There is one village as illustration as shown in the uncertainty map (Figure 5), and other areas that are also evident can be clearly seen in Figure 6. For politically reasoning, we may say these areas still have the possibility to change their political view to choose their next major.

4. CONCLUSION

The EM algorithm can be applied to determine which areas have similar political views in Manado. The application of this algorithm follows the selected model where through the model we can determine the cluster size is 3 clusters. Based on the VEE model, clusterization is made to classify any region that has similar political views in the city area of Manado. There are 3 different groups to classify the region, the first group amounted to 27 villages, the second group of 21 villages and the rest are in the third group of 8 villages. Following the size of the cluster, it can also be taken into account mixing probabilities. Each cluster group has different mixing probabilities, ie 0.2132314, 0.5562200 and 0.2305486, respectively. The second cluster group of 21 villages has the highest mixing probabilities, which means the villages are indeed closely related to each other and have a uniform political outlook.

There are some areas that are considered to be areas of uncertainty, meaning that at any time the territory may change at any time. There are 13 uncertain territories and they are 80, 82, 83, 26, 42, 11, 17, 54, 55, 53, 62, 5 and 61, respectively. These areas must be taken into account by political players so that they can be taken advantage of the uncertain nature of the voters because they can be influenced by others.

Acknowledgement

We are grateful to the Indonesian Ministry of Research and Higher Education who has helped finance and approve the research through Penelitian Produk Terapan (PPT).

5. REFERENCES

G. J. McLachlan and T. Krishnan., 1996, The EM Algorithm and Extensions. Wiley-Interscience, 1 edition, November.

A. P. Dempster, N. M. Laird, and D. B. Rubin., 1977, Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39(1):138.

C. Fraley and A. E. Raftery., 2007, Model-based Methods of Classification: Using the mclust Software in Chemometrics. Journal of Statistical Software. January, Volume 18, Issue 6. http://www.jstatsoft.org/

L. Scrucca, M. Fop, T. B. Murphy and A. E. Raftery., 2016, mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. HHS Public Access. R J. 8(1): 289–317. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5096736/pdf/nihms793803.pdf

McLachlan G, Peel D., 2000, Finite Mixture Models. Wiley, New York.

Banfield J, Raftery AE., 1993, Model-based Gaussian and non-Gaussian clustering. *Biometrics*. 49: 803–821.

Celeux G, Govaert G., 1995, Gaussian parsimonious clustering models. *Pattern Recognition*. 28:781–793.

Hutapea, A., 2015, Dinamika Hukum Pemilihan Kepala Daerah di Indonesia, *Jurnal Rechts Vinding*, vol. 4.