

PENERAPAN DATA MINING DALAM MENENTUKAN POTENSI KEBERHASILAN BAKAL CALON LEGISLATIF DI DAERAH PEMILIHAN JAWA BARAT MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBORS*

Alfin Duhawan Bagja^{*}, Gunawan Abdillah, Faiza Renaldi

Jurusan Informatika, Fakultas MIPA, Universitas Jenderal Achmad Yani

Jl. Terusan Jenderal Sudirman, Cimahi, Jawa Barat, 40513

*Email: alfinduhawanbagja@gmail.com

Abstrak

*Pemilihan umum adalah salah satu cara dalam sistem demokrasi untuk memilih wakil rakyat yang akan duduk di lembaga perwakilan rakyat. Popularitas bukanlah faktor utama yang dapat menjamin kesuksesan calon legislatif untuk lolos pada pemilihan umum. Dibandingkan dengan mengandalkan popularitas, menentukan strategi adalah hal yang tepat untuk dilakukan saat pemilihan umum. Salah satu strategi tersebut adalah dengan menentukan daerah pemilihan. Permasalahan yang terjadi dalam menentukan daerah pemilihan adalah terjadinya ketidaksinambungan wilayah antara calon legislatif dengan daerah pemilihannya. Maka perlu memperhatikan data latar belakang anggota legislatif sebelumnya yang telah berhasil di daerah pemilihannya. Data Mining didefinisikan sebagai proses penemuan pola dalam data, algoritma *K-Nearest Neighbors* adalah salah satu algoritma data mining yang melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Penelitian ini membangun sebuah sistem yang dapat menentukan daerah pemilihan yang berpotensi berhasil untuk bakal calon legislative di Jawa Barat. Hasil dari penelitian ini berupa sistem yang dapat menghasilkan rekomendasi daerah pemilihan yang berpotensi berhasil untuk bakal calon legislatif dengan tingkat akurasi data uji sebesar 85,62%.*

Kata kunci: pemilihan umum, calon legislatif, daerah pemilihan, data mining, k-nearest neighbors.

1. PENDAHULUAN

Pemilihan umum (Pemilu) adalah salah satu cara dalam sistem demokrasi untuk memilih wakil-wakil rakyat yang akan duduk di lembaga perwakilan rakyat, serta salah satu bentuk pemenuhan hak asasi warga negara di bidang politik. Popularitas dalam pemilihan umum sering di anggap sebagai modal penting untuk dapat lolos dan terpilih menjadi anggota legislatif, sehingga tidak sedikit para publik figur di bidang kesenian memilih meninggalkan karir mereka sebagai artis untuk berkarir di bidang politik. Berdasarkan hasil pemilu pada periode sebelumnya, banyak artis yang mencalonkan diri untuk bersaing pada pemilihan umum, hasilnya tidak sedikit dari mereka yang gagal untuk lolos menjadi anggota legislatif. Sebagai contoh pada pemilu periode 2014 April kemarin, beberapa nama ternama seperti aktor Sandy Nayoan, dan mantan juara dunia bulu tangkis Ricky Subagja. Bahkan artis pada periode sebelumnya lolos menjadi anggota parlemen pada periode 2014 tersingkir dari persaingan, seperti Ingrid Karsil, dan Nurul Arifin. Tidak hanya artis yang gagal pada pemilihan umum, beberapa politisi terkenal dan sudah lama berkarir di bidang politik seperti Ketua DPR periode 2009 Marzuki Alie dan Wakilnya Priyo Budi Santoso, Eva Kusuma Sundari, Nurul Arifin dan Sutan Bhatogana, turut gagal dalam pemilihan umum. Hal ini membuktikan bahwa popularitas bukanlah faktor utama yang dapat menjamin kesuksesan calon legislatif untuk lolos pada pemilihan umum. Dibandingkan dengan mengandalkan popularitas di kalangan masyarakat, membangun strategi lebih tepat untuk dapat dijadikan sebagai modal dalam bersaing pada pemilihan umum. Terdapat banyak strategi yang dapat diterapkan untuk bersaing pada pemilihan umum, salah satunya adalah menentukan daerah pemilihan.

Dalam daerah pemilihan di Jawa Barat, terdapat 11 daerah pemilihan yang digolongkan menjadi Jawa Barat I sampai dengan Jawa Barat XI. Calon legislatif dapat menentukan daerah pemilihan mereka sendiri. Akan tetapi pada beberapa kesempatan dalam menentukan daerah pemilihan tersebut sering tidak optimal seperti terjadinya ketidaksinambungan wilayah antara calon legislatif dengan daerah pemilihan, selain itu latar belakang dari calon legislatif tersebut kurang cocok untuk berada pada daerah pemilihan yang dipilihnya, sehingga menyebabkan kurangnya pengetahuan seputar daerah pemilihan bagi calon legislatif tersebut. Hal ini berakibat calon

legislatif tersebut memiliki kemungkinan dapat kalah saing dengan calon legislatif lainnya. Latar belakang anggota legislatif dapat dijadikan sebagai acuan dalam menentukan daerah pemilihan di Jawa Barat. Informasi latar belakang yang dibutuhkan dalam menentukan daerah pemilihan dapat digali berdasarkan pada data anggota legislatif sebelumnya. Data mining didefinisikan sebagai proses penemuan pola dalam data (Ledyana, 2013). Dalam menemukan informasi kriteria pada data sebelumnya dapat menggunakan metode Data Mining. Pada penelitian terdahulu metode data mining sudah digunakan seperti pada penelitian, penentuan resiko kredit kepemilikan kendaraan bermotor (Ledyana, 2013), pengambilan keputusan pengajuan kredit (Pandie, 2012), klasifikasi bidang kerja kelulusan (Nursalim, et al., 2014).

K-Nearest Neighbors merupakan salah satu algoritma yang digunakan untuk mencari kecocokan data lama dengan data baru. Algoritma KNN sudah banyak digunakan pada penelitian-penelitian sebelumnya seperti klasifikasi data hasil produksi kelapa sawit pada PT. Minamas Kecamatan Parindu (Krisandi, et al., 2013). Dalam penelitian ini akan di gunakan metode data mining dalam menentukan keberhasilan bakal calon legislatif di daerah pemilihan Jawa Barat menggunakan algoritma KNN.

2. METODOLOGI

2.1 Pengumpulan Data

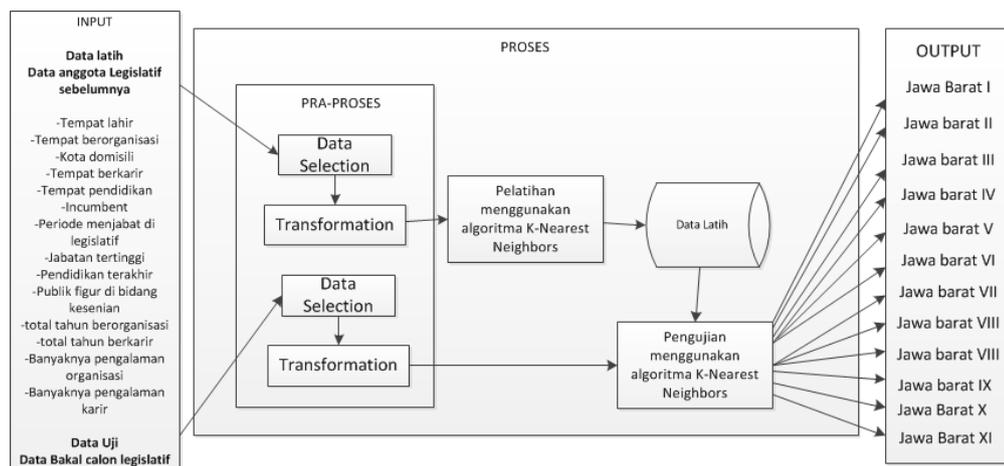
Pada tahapan ini dilakukan pengumpulan data yang dibutuhkan untuk penelitian ini. Data yang dikumpulkan adalah data legislatif sebelumnya, yang nanti akan dicari kesamaan untuk menentukan penempatan daerah pemilihan bakal calon legislatif pada pemilihan umum.

2.2 Analisis Kebutuhan

Pada tahapan ini, menganalisa kebutuhan penelitian ini, berdasarkan data yang sudah diperoleh, dan metode yang akan digunakan. Setelah itu melakukan studi litelatur seputar metode yang digunakan.

2.3 Gambaran Sistem

Penggambaran perancangan sistem untuk menentukan potensi keberhasilan bakal calon legislatif pada daerah pemilihan Jawa Barat menggunakan algoritma k-nearest neighbors dijelaskan pada Gambar 1



Gambar 1. Perancangan Sistem

2.3.1 Data Mining

Data Mining adalah suatu konsep yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data Mining* juga didefinisikan sebagai proses penemuan pola dalam data, atau sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Ginting, et al., 2014).

Data Mining merupakan salah satu bagian proses *Knowledge Discovery in Database* (KDD) yang bertugas mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. Proses dari KDD sebagai berikut :

1. *Data Selection* : pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.
2. *Cleansing* : membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data seperti kesalahan cetak.
3. *Transformation* : mengubah data menjadi format untuk di proses dalam data mining
4. *Data mining* : mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.
5. *Evaluation* : Pola informasi yang di hasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

2.3.2 Algoritma K-Nearest Neighbors

Algoritma K-NN tergolong dalam algoritma *supervised* yaitu proses pembentukan algoritma diperoleh melalui proses pembelajaran (*learning*) pada data lama yang sudah terklasifikasi dan hasil pembelajaran tersebut dipakai untuk mengklasifikasikan data baru dengan output yang belum diketahui. Dalam algoritma K-NN sebuah data baru diklasifikasikan berdasarkan jarak data baru tersebut dengan tingkat kemiripan data baru terdekat terhadap data pola (Ndaumanu, et al., 2014).

Terdapat banyak persamaan pada algoritma K-Nearest Neighbors dalam menentukan jarak kedekatan antara data lama dengan data baru, salah satunya adalah dengan *euclidean distance*. Pada penelitian ini digunakan persamaan 2.1 berikut untuk menemukan nilai *similarity* antara data latih dengan data uji :

$$\text{Similarity}(p, q) = \frac{\sum_{i=1}^n f(p_i, q_i) \times w_i}{w} \dots\dots\dots(2.1)$$

Keterangan :

p = Data uji

q = Data latih

n = Jumlah atribut dalam tiap kasus, terdapat 14 atribut

i = Atribut individu antara 1 sampai dengan 14

f = Fungsi *similarity* atribut i antara kasus p dan kasus q

w = Bobot yang diberikan pada atribut ke-i.

2.4 Perancangan dan Pembuatan Perangkat Lunak

Pada tahapan ini dilakukan perancangan dan pembuatan perangkat lunak sistem yang dapat menentukan keberhasilan bakal calon legislatif pada pemilihan umum di daerah pemilihan Jawa Barat.

2.5 Pengujian dan Evaluasi

Pada tahapan ini dilakukan pengujian sistem dan evaluasi, pengujian dilakukan dengan cara menginputkan data bakal calon legislatif yang nanti akan di proses untuk menemukan nilai *similarity* dengan data anggota legislatif sebelumnya.

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan

Dalam penelitian ini terdapat 14 data masukkan pada sistem ini yaitu, tempat lahir, tempat berorganisasi, tempat pendidikan, kota domisili, tempat berkarir, incumbent (anggota legislatif yang masih menjabat lalu kembali mencalonkan diri pada pemilihan umum), total tahun berkarir di legislatif, jabatan tertinggi, pendidikan terakhir, publik figur di bidang kesenian, total tahun berorganisasi, total tahun berkarir, banyaknya pengalaman organisasi, dan banyaknya perjalanan karir. Tabel 1 dibawah ini menunjukkan salah satu nilai atribut dari variabel incumbent.

Tabel 1. nilai atribut variabel incumbent

Variabel	Atribut
Incumbent	Ya dan Tidak

Penentuan bobot pada penelitian ini ditentukan berdasarkan kepentingan dari setiap variabel. Bobot dari setiap variabel dijelaskan pada tabel 2 dibawah ini

Tabel 2. Bobot

variabel	bobot
tempat lahir	0,8
tempat berorganisasi	1
tempat pendidikan	0,9
kota domisili	0,8
tempat berkarir	1
incumbent	0,6
total tahun berkarir di legislatif	0,5
jabatan tertinggi	0,7
pendidikan terakhir	0,8
publik figur di bidang kesenian	0,7
total tahun berorganisasi	0,5
total tahun berkarir	0,5
banyaknya pengalaman organisasi	0,6
banyaknya perjalanan karir	0,6

Terdapat 3 Data latih dari anggota legislatif periode sebelumnya, seperti pada tabel 3 dibawah

Tabel 3. Data latih

Nama	Tempat Lahir	Tempat Berorganisasi	Tempat pendidikan	Domisili kota	Tempat berkarir	Incumbent	Total tahun berkarir di legislatif	Jabatan tertinggi	Pendidikan terakhir	Publik figur di bidang kesenian	Total tahun berorganisasi	Total tahun berkarir	Banyaknya pengalaman organisasi	Banyaknya perjalanan karir
Data Latih 1 (Jawa Barat III)	Kab. Bandung	Kab. Bandung	Jakarta	Jakarta	Kab. Bandung	Tidak	Belum Punya	Ketua	S3	Bukan	7 tahun	> = 10 tahun	4 Organisasi	6 karir
Data Latih 2 (Jawa Barat II)	Non Jabar Nasional	Kota Bandung	Jakarta	Jakarta	Kota Bandung	Ya	>=10 tahun	Dewan	S2	Bukan	>= 10 tahun	>= 10 tahun	5 Organisasi	6 karir
Data latih 3 (Jawa Barat IV)	Kab. Bandung	Kab. Bandung	Kab. Bandung	Kab. Bandung	Kab. Bandung	Tidak	Belum Punya	Ketua	SMA	Ya	4 tahun	5 tahun	3 Organisasi	2 karir

Contoh Data uji seperti pada tabel 4 dibawah ini

Tabel 4. Data Uji

Nama	Handoko
Tempat Lahir	Kota Bandung
Tempat Berorganisasi	Kota Bandung
Tempat Pendidikan	Jakarta
Domisili Kota	Jakarta
Tempat berkarir	Kota Bandung
Incumbent	Tidak
Total tahun berkarir di legislatif	Belum Punya
Jabatan tertinggi	Dewan
Pendidikan terakhir	S1
Publik figur di bidang kesenian	Bukan
Total tahun berorganisasi	4 tahun
Total tahun berkarir	5 tahun
Banyaknya pengalaman organisasi	7 organisasi
Banyaknya perjalanan karir	2 karir

Proses perhitungan menggunakan algoritma K-Nearest Neighbors, dengan menemukan nilai similarity yang paling tinggi antara data bakal calon legislatif dengan data anggota legislatif periode sebelumnya. Sebelum masuk kedalam perhitungan, hal yang pertama dilakukan adalah dengan menentukan kedekatan antara nilai atribut, tabel 5 dibawah ini menjelaskan tentang kedekatan nilai atribut variabel incumbent. Pada Variabel incumbent terdapat 2 atribut, jadi untuk mendapatkan nilai kedekatan atribut tersebut maka $1 : 2 = 0,5$

Tabel 5. kedekatan nilai atribut variabel Incumbent

Incumbent	Ya	Tidak
Ya	1	0,5
Tidak	0,5	1

Setelah mendapatkan nilai kedekatan dari masing masing atribut variabel maka selanjutnya adalah melakukan perhitungan menggunakan algoritma K-Nearest neighbors. Rumus perhitungan terdapat pada persamaan (1). Hasil perhitungan dapat dilihat pada tabel dibawah ini

Tabel 6. Hasil Perhitungan

Kasus	Hasil Similarity
Kedekatan Data Latih 1 dengan Data Uji	$((0*0.9)+(0*1)+(1*0.9)+(1*0.8)+(0*1)+(1*0.6)+(1*0.5)+(1*0.7)+(0.6*0.8)+(1*0.7)+(0.73*0.5)+(0.55*0.5)+(0.73*0.6)+(0.82*0.6))/10 = \mathbf{0.625}$
Kedekatan Data Latih 2 dengan Data Uji	$((0*0.9)+(1*1)+(1*0.9)+(1*0.8)+(1*1)+(0.5*0.6)+(0.09*0.5)+(1*0.7)+(0.8*0.8)+(1*0.7)+(0.46*0.5)+(0.55*0.5)+(0.82*0.6)+(0.82*0.6))/10 = \mathbf{0.7574}$
Kedekatan Data Latih 3 dengan Data Uji	$((0*0.9)+(0*1)+(0*0.9)+(0*0.8)+(0*1)+(1*0.6)+(1*0.5)+(1*0.7)+(0.6*0.8)+(0.5*0.7)+(1*0.5)+(1*0.5)+(0.64*0.6)+(1*0.6))/10 = \mathbf{0.4614}$

Berdasarkan tabel 6, Nilai Similarity tertinggi terdapat pada Data latih 2 dengan data uji dengan nilai similarity 0.7574, sehingga rekomendasi daerah pemilihan untuk data uji Handoko adalah Jawa Barat II.

3.2 Hasil

Hasil dari penelitian ini adalah sebuah sistem yang dapat menentukan potensi keberhasilan bakal calon legislatif di daerah pemilihan Jawa Barat menggunakan algoritma K-Nearest Neighbors. Keluaran dari sistem ini berupa rekomendasi daerah pemilihan bagi bakal calon legislatif berdasarkan nilai *similarity* dengan anggota legislatif periode sebelumnya.

4. KESIMPULAN DAN SARAN

Penelitian ini telah menghasilkan sebuah sistem yang dapat menentukan potensi keberhasilan bakal calon legislatif di daerah pemilihan Jawa Barat dengan menggunakan algoritma K-Nearest Neighbors. Sistem dapat menentukan daerah pemilihan yang berpotensi berhasil bagi bakal calon legislatif berdasarkan 14 data masukkan yaitu, tempat lahir, tempat berorganisasi, tempat pendidikan, kota domisili, tempat berkarir, incumbent (anggota legislatif yang masih menjabat lalu kembali mencalonkan diri pada pemilihan umum), total tahun berkarir di legislatif, jabatan tertinggi, pendidikan terakhir, publik figur di bidang kesenian, total tahun berorganisasi, total tahun berkarir, banyaknya pengalaman organisasi, dan banyaknya perjalanan karir. Adapun hasil dari pengujian sistem menggunakan metode *blackbox* dengan persentase 85,62%.

Saran untuk sistem ini diharapkan adanya penambahan data anggota legislatif periode sebelumnya untuk meningkatkan keakuratan dari penentuan daerah pemilihan ini.

DAFTAR PUSTAKA

- Ginting, S. L. B., Zarman, W. & Darmawan, A., 2014. Teknik Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighborhood. *Jurnal Teknik Komputer Unikom*, 3(2), pp. 29-34.
- Krisandi, N., Helmi & Prihandono, B., 2013. Algoritma K-Nearest Neighbors Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. Minamas Kecamatan Parindu. *BuletinIlmiah Math. Stat. dan Terapannya (Bimaster)*, 2(1), pp. 33-38.
- Leidyana, H., 2013. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *Jurnal Penelitian Ilmu Komputer, System embedded & Logic*, 1(1), pp. 65-76.
- Ndaumanu, R. I., Kusriani & Arief, R. M., 2014. Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor. *Jatisi*, 1(1), pp. 1-15.
- Nursalim, Suprapedi & Himawan H., 2014. Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Teknologi Informasi*, 10(1), pp. 31-43.
- Pandie, E. S. Y., 2012. Implementasi Algoritma Data Mining K-Nearest Neighbour (K-NN) Dalam Pengambilan Keputusan Pengajuan Kredit. *Seminar Nasional Sains dan teknik 2012 (SAINSTEK 2012)*. pp. 31-34.